MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0260
Measure Title: Assessment of Health-related Quality of Life in Dialysis Patients
Measure Steward: Witten and Associates, LLC
Brief Description of Measure: Percentage of eligible dialysis patients who complete a health-related quality of life assessment with or without assistance using the KDQOL-36 (36-question survey that assesses patients´ functioning and well-being) at least once during a calendar year.
Developer Rationale: Percentage of eligible dialysis patients who complete a health-related quality of life assessment with or without assistance using the KDQOL-36 (36-question survey that assesses patients´ functioning and well-being) at least once during a calendar year.

Numerator Statement: Number of eligible (not excluded) individuals with ESRD (ICD-10 N18.6) on dialysis who complete a KDQOL-36 with or without assistance at least once per calendar year
Denominator Statement: Number of individuals with ESRD (ICD-10 N18.6) on peritoneal dialysis, in-center hemodialysis, and home hemodialysis treated by the dialysis facility during the calendar year minus those dialysis patients who meet exclusion criteria in S.10.
Denominator Exclusions: Patients with ESRD (ICD-10 N18.6) on dialysis who are <18 years old; who are unable to complete the survey due to mental status that could invalidate the results; who are non-English speaking/reading and no native language translation or interpreter is available; or who have been on dialysis for <3 months. A patient who declines to complete one survey but completes one survey during the calendar year is counted as having a completed survey.

Measure Type: Process
Data Source: Patient Reported Data/Survey [*NQF Staff Note: the KDQOL survey instrument itself is not the data source for this measure; for small independent dialysis facilities, data on survey completion rates is collected by entities such as KDQOL Complete, while large dialysis facilities have their own databases from which completion data can be derived*.]
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Nov 15, 2007 **Most Recent Endorsement Date:** Nov 15, 2007

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary:

Evidence during the previous review was provided for the survey, not the actual measure.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- The evidence for this measure is based on KDOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification; GUIDELINE 12. Based on 1989-2001 datya, the evidence presented supports the recommendation that: "Patients with GFR <60 mL/min/1.73 m2 should undergo regular assessment for impairment of functioning and well-being: 1)to establish a baseline and monitor changes in functioning and well-being over time, and 2)to assess the effect of interventions on functioning and well-being."
- The guidelines developed by the Work Group are based on a systematic review of the literature using an approach based on the procedure outlined by the Agency for Healthcare Research and Quality (formerly the Agency for Health Care Policy and Research) with modifications appropriate to the goals. An Evidence Review Team was appointed by the NKF to collaborate with the Work Group to conduct a systematic review of the literature on which to base the guidelines.
- The guidelines referenced 28 articles cited in the guideline from prospective and retrospective studies, 2 descriptive studies, 2 cross-sectional studies, 4 comparison studies, 3 cohort studies, 7 clinical trials (1 Phase III, 2 randomized controlled clinical trials), 1 epidemiological, 2 case control studies, 3 reviews and 1 report.
- Additional evidence was provided from PubMed searches, national renal meeting abstracts and posters, and Fresenius Kidney Care.

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm: [Box 1] Measure **does not** assess performance on a health outcome or patient-reported outcome →[Box 3] Measure **does** assess a process that is based on a systematic review and grading of the body of empirical evidence where the specific focus of the evidence supports what is being measured → [Box 4] A summary of the quantity, quality, and consistency (QQC) of evidence from a systematic review **is provided** in the submission form →[Box 5] The systematic review concludes that there is high/moderate certainty that the net benefit of this process is substantial → Highest eligible rating is HIGH

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Using KDQOL-Complete data, the developer tested whether there are statistically significant differences in the performance measure between facilities with at least 10 patients.

Year	N	Mean	Median	Range	Interquartile

					Range
2013	1,240	85.4	91.8	0-100	78.3-100
2014	1,222	85.5	92.1	0-100	78.3-98.1
2015	1,261	85.6	91.8	16.7-100	78.3-100

Disparities

- Using full KDQOL-Complete data, the developer only included the first KDQOL-36 survey for patients who completed the survey more than once in a calendar year. Data was compared for "All", "Completed", "Refused", and "Excluded" for 2013, 2014, 2015.
- 2013

	N	Mean Age	Male	White	Black	Asian	Native Am	Pacific Island	Missing	Diabetes
All	79,872	62.4 (14.9)	55.5%	45.4%	26%	4.5%	1.6%	1.8%	20.8%	52.7%
Completed	62,620	61.9 (14.7)	55.7%	46.3%	25.9%	4.3%	1.8%	1.8%	20.2%	52.8%
Refused	11,790	61.9 (15.1)	49.9%	43.7%	25.3%	3.7%	1.3%	2.4%	23.5%	51.6%
Excluded	5,462	69.2 (14.5)	59.1%	38.6%	28.5%	8.0%	2.6%	1.3%	21.0%	54.6%

- 2014

	N	Mean Age	Male	White	Black	Asian	Native Am	Pacific Island	Missing	Diabetes
All	84,167	62.4 (14.9)	56.3%	45.4%	25.8%	5.5%	1.5%	1.9%	19.9%	52.6%
Completed	66,386	62.0 (14.7)	56.2%	46.6%	25.8%	5.4%	1.5%	1.8%	18.9%	52.9%
Refused	12,015	61.7 (15.1	51.2%	42.5%	25.5%	4.4%	1.5%	1.7%	24. 3%	50.5%
Excluded	5,756	69.2 (14.6)	59.1%	37.8%	27.4%	9.4%	0.6%	2.5%	22.3%	54.1%

- 2015

	N	Mean Age	Male	White	Black	Asian	Native Am	Pacific Island	Missing	Diabetes
All	87,892	62.4 (14.7)	56.4%	45.1%	26.6%	5.4%	1.5%	1.9%	19.5%	52.5%
Completed	69,746	62.0 (14.5)	56.4%	45.9%	27.0%	5.1%	1.5%	1.9%	18.7%	52.4%
Refused	12,475	61.7 (15.0)	59.2%	43.3%	24.7%	5.2%	1.7%	1.9%	23.1%	51.6%
Excluded	5,671	69.5 (14.6)	50.4%	38.8%	26.5%	9.9%	0.7%	2.2%	21.9%	55.3%

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

○ *Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **The KDQoL is a composite measure of patient outcomes - -

So this is a process measure. Testiong does apparently show that the frequency of completion correlates with QoL scores on a facility-level...

Thus I would rate the evidence as HIGH for this measure

**While measuring is likely better than not measuring, there is no evidence presented that links the process to health outcomes - we are asked to make that leap.

Rating: Low

**While measuring is likely better than not measuring, there is no evidence presented that links the process to health outcomes - we are asked to make that leap.

Rating: Low

**The evidence for this measure is based on KDOQI Clinical Practice Guidelines for Chronic Kidney Disease. Based on 1989-2001 data, the evidence presented supports the recommendation that: "Patients with GFR <60 mL/min/1.73 m2 should undergo regular assessment for impairment of functioning and well-being. The guidelines developed by the Work Group are based on a systematic review of the literature using an approach based on the procedure outlined by the Agency for Healthcare Research and Quality (formerly the Agency for Health Care Policy and Research) with modifications appropriate to the goals.

**This measure is good for taking a snapshot of the overall well being of a person on dialysis. In reviewing the KDQOL 36 there are only a handful of questions that dialysis providers can realistically improve. For example questions 2, 3 and 19-28 can directly be tied to dialysis treatment performed.

**Measure focuses on a process of care: percentage of eligible patients who complete the KDQOL 36 annually (reported at the facility level); Evidence is tangential to the measure as specified - measure could be considered a distal process step relative to the evidence;

the process being measured (completion of KDQOL 36) is an initial step in the proposed conceptual framework for improving outcomes.

**This is a process measure, and is based on systematic review of literature.

**There are significant issues with the KDQOL.

1. The original validation work was done many years ago when dialysis technology was different and side effects were different. As a result some of the questions are no longer applicable.

2. The original validation work was done on 200 patients but that validation has not held up on larger more recent data. http://www.davitaclinicalresearch.com/wp-content/pdfs/ASN_2010/SA-PO2616-ASN2010_KDQOL_Valid_8Nov10DVW.pdf

3. There are floor and celling effects and threats to external and internal validity as a result

**Measure of KDQOL survey completion rate -- percentage completion and refusal of eligible patients who took survey in, particularly, last 3 years (national data). Results consistent over last 3 years.

Patient outcomes reflect patient self-perception of dealing with kidney disease as well as general physical and mental status. This proposal does not address patient outcomes (results of survey) per se.

**no clear relation to intervention.

**This is a process measure. Evidence is based on KDOQI from data from 1989-2002. Since that time CMS is requiring pain and depression screening that has more recent data to support those measures. Since this was the only measurement tool available it did allow for measurements of the burden of illness and s/s's.

The evidence to support the measure is moderate as better data supports depression and pain screening.

**Process Measure. Evidence does apply directly to the mesure. Measure does assess a process that is based on a systematic review and grading of the body of empirical evidence where the specific focus of the evidence supports what is being measured. There is a relationships between the measured outcome and at least one healthcare action--intervention--identified and supported by the stated rationale. Net benefit of process is substantial

1b. Performance Gap

Comments: **Taken as a composite measure, there is no evidence that QoL measure completeness has changed for dialysis patients

from 2013-2015. There is modest evidence of a performance "gap" - - i.e., evidence that on a facility level, the is variation in measures that might improve.
Preliminary rating: Moderate
**Top table: Are these percentages? I see no test for statistical significance. I suppose the mean being quite a bit lower than the median suggests there were a number of poor performers, so there may be a performance gap.
Bottom table (disparities): I honestly cannot make sense of this data
Rating: Insufficient or maybe low
**Top table: Are these percentages? I see no test for statistical significance. I suppose the mean being quite a bit lower than the median suggests there were a number of poor performers, so there may be a performance gap.
Bottom table (disparities): I honestly cannot make sense of this data
Rating: Insufficient or maybe low
**The developer tested whether there are statistically significant differences in the performance measure between facilities with at least 10 patients.
There was disparities noted among groups based upon race
**Patients on dialysis continue to suffer from poor quality of life and improvements can be made. The loss of ones kidney function and the need to rely on a machine to live is devastating and can be hard to adjust to physically, emotionally and relationships often suffer. The dietary restrictions, time constraints and lifestyle changes are extremely difficult to adapt to. The key to improving quality of life is to ensure patients are receiving the best treatment possible with no side effects such as cramping, crashing and nausea. If patient feels good after treatment they have more control over their life and the ability to improve it. If patient feels washed out or need to sleep after treatment they have little chance of ever improving their quality of life.
**Gap in 2015: median 91.8%; 10%=61.2% and 5%=48.9
**There is a performance gap, though it is not huge. There is some variability. There is minimal racial and gender disparity.
**It is not clear if there is a performance gap. The definition of excluded and refused require more granularity to compare between providers.
**Gap might related to the 15% (of eligible patients) who refuse survey... issue might be how to increase survey responses >85% and reduce 15% refusal rate. Question, for another proposal, might be what goal of completion should be for units. Is it relevant to even set a goal?
**this is a process measure which is fairly well 'gamed" by dialysis providers. Not clear that implementation of measure has actually improved patient s" lives
**Measure will help assess function and well-being. Gives information to providers on the impact (Burden) of kidney disease and signs and symptoms related to their disease. Since it is a process measure there is no measure of gap in care as the results as defined as above, below or average. It is the providers responsibility through the C4C's to develop a plan of care based on the below average results.
Non-English speaking patients are excluded which may lead to disparities id translation services in certain languages is not available. Low numbers of non-white populations in the disparities table may signal a population that is less likely to complete the results.
**Disparities--race was added to the survey in 2013.

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. **Data source(s):** Registry or database of completed surveys **Specifications:** • The developer attests the specifications have changed in the following ways since the last submission: o The exclusion of "<3 months at the facility" was revised to "<3 months on dialysis." ▪ The ESRD Conditions for Coverage at 42 CFR 494.90 require the dialysis interdisciplinary team to perform a reassessment during the 4th month of dialysis and to use results of that reassessment for the patient plan of care meeting that is held 15 days after the team completes that reassessment. Excluding patients during the first 3 months of dialysis will align the survey with the first reassessment. o The exclusion of "cognitive impairment, dementia, psychosis" was revised to "unable to complete due to

mental status" to incorporate those diagnoses as well as others that make completion impossible or unreliable.
    - Patients who refuse to complete the survey were removed from the exclusions from the denominator. Facilities need to track and make efforts to to increase the number of patients completing the survey.
    - The target populations was broadened to include more than just seniors since dialysis patients are populations at risk, dual eligible beneficiaries, individuals with multiple chronic illnesses, veterans.
- The numerator of this measure is: *Number of eligible (not excluded) individuals with ESRD (ICD-10 N18.6) on dialysis who complete a KDQOL-36 with or without assistance at least once per calendar year*
- The denominator is: *Number of individuals with ESRD (ICD-10 N18.6) on peritoneal dialysis, in-center hemodialysis, and home hemodialysis treated by the dialysis facility during the calendar year minus those dialysis patients who meet exclusion criteria in S.10 and noted above.*

Questions for the Committee :

- *How do facilities track completion of the survey? What data is captured in KDQOL-Complete?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:
Testing during the previous review was provided for the survey, not the actual measure.
Describe any updates to testing: The developers have now provided results of reliability testing for the performance measure as specified.

SUMMARY OF TESTING
Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- Empirical testing of computed performance scores for reportable clinics was conducted using a beta-binomial model.

Results of reliability testing:

- The internal reliability of the measure resulted in Cronbach's alpha of 0.926 for 2013, 0.925 for 2014, and 0.923 for 2015 from KDQOL-Complete data.
- In terms of understanding reliability in detecting signal to noise, a reliability score of 0.70 or greater is considered acceptable for drawing conclusions about groups. This data analysis of critical data elements, demonstrates this measure construct to be reliable and to detect meaningful differences among facilities and their patient population

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: [Box 1] Submitted specifications **are** precise, unambiguous, and complete so that they can be consistently implemented → [Box 2] Empirical reliability testing **was** conducted using statistical tests with the measure as specified → [Box 4] Reliability testing **was** conducted with computed performance measure scores for each measured entity → [Box 5] The method described **was** appropriate for assessing reliability → [Box 6] There **is** high certainty that the performance scores are reliable → Highest eligible rating is HIGH

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:
Testing during the previous review was provided for the survey, not the actual measure.
Describe any updates to testing: The developers have now provided results of validity testing for the performance measure as specified.

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- The developer assessed the measure's association with actual quality of life scores of patients.
- This was done using linear mixed models with the patient-level quality of life scores for each scale as the dependent variable and facility completion rate as the main independent variable.
- The models were adjusted for patient-level characteristics (age, sex, race, and diabetes). The models accounted for facility clustering using, assuming a compound symmetry covariance structure.

Validity testing results:

- The developer found significant and positive association between facility completion rates and QoL scores for all five scales. Results displayed are the estimated effect of comparing a facility with a 10% higher completion rate (i.e., 90% vs. 80%). The developer notes that while the estimates may appear small, one should keep in mind that this estimate applies to all patients within the facility.

QoL Scale	Estimate	p-value
Symptoms	0.21564	<0.0001
Effects	0.36450	<0.0001
Burden	0.19037	0.0223
MCS	0.08023	0.0060

PCS	0.05872	0.0389

- The developer found that facilities with higher completion rates were associated with statistically significantly higher patient-level QoL scores within the facility, indicating that it is plausible that facilities with higher rates would be obtaining completed questionnaires from sicker patients, since it has been documented that individuals completing the QoL scores tend to be younger and healthier.

Questions for the Committee:

- *Do you expect a correlation between completion rates and the Quality of Life score?*
- *Does the finding "it has been documented that individuals completing the QoL scores tend to be younger and healthier." Have an impact on the validity of the measure result?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer conducted the following analyses to assess that the exclusion criteria were being applied correctly and the results are valid:
  - For the age exclusion, what percent of the time were individuals actually <18 (calculated as date declined minus date of birth)?
  - What are the exclusions by other available characteristics?
  - Are the exclusions needed to make the comparisons fair?
  - What are the distributions of exclusions across facilities?
- In order to assess if the exclusions are needed, the developer assessed the association between completion vs. refusal and completion vs. excluded, which are both presented below.
- The developer found that individuals reported to have been excluded based on age <18 years were calculated to be <18 years approximately 90% of the time.
- The findings of completion rates across facilities showed that males were significantly less likely to complete the KDQOL-36 survey (rather than refuse) compared to female patients, while older patients, minorities, and those whose race is unreported (missing) were significantly more likely to complete the KDQOL-36 survey (rather than refuse) compared to white race.
- The findings of completion rates across facilities showed that males were significantly more likely to complete the KDQOL-36 survey (rather than be excluded) compared to female patients when compared to those who were excluded, while older patients, minorities, and those whose race is unreported (missing) were significantly less likely to complete the KDQOL-36 survey (rather than be excluded) when compared to white race.
- With regard to the frequency of different exclusions, in facilities with 10 or more patients, staff excluded those with cognitive impairment/dementia/active psychosis in the greatest numbers (mean 24.8%; median 20%). Staff exclusion for age <18 was very low (mean 0.2%; median 0). Staff seldom excluded those who had been at the facility <3 months (mean 3.7%; median 0). Staff exclusion for language showed the greatest variation across facilities (mean 4.1%; median 8.5%).
- Staff were more likely to exclude Asian patients for language. There is no KDQOL translation for all Asian languages.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Linear regression was employed to test for significant differences across all facilities using the facility ID as a dummy variable and assessing the Type III SS for overall variation, adjusted for year. Models were run both adjusted and unadjusted for patient characteristics.
- The developer provides a 'box and whisker' plot to display results of this analysis, as well as summary statistics.
- The developer's interpretation of the results is that there is "enough significant variation across the facilities to assess for meaningful differences in scores."

Question for the Committee:

○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not needed

2b7. Missing Data

- The developer states that every patient in KDQOL-Complete was successfully categorized into "Completed," "Refused," or "Excluded". However, since the developer cannot know whether facilities have patients whose data is never entered into KDQOL-Complete, it cannot assess whether data are missing.

Guidance from the Validity Algorithm: [Box 1] Measure specifications **are** consistent with the evidence provided in support of the measure → [Box 2] All potential threats to validity that are relevant to the measure **were** empirically assessed → [Box 3] Validity testing **was** conducted using the measure as specified and appropriate statistical test → [Box 6] Validity testing **was** conducted with computed performance scores for each measured entity → [Box 7] The method **was** described and appropriate for assessing validity → [Box 8] Highest eligible rating is HIGH

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **Specification is consistent with evidence

2a2. Reliability Testing

Comments:
**Reliability testing seems adequate
Preliminary rating: High
**Not familiar enough with methodology to be able to evaluate.
I am concerned that data is collected in three different systems and thus reliability between these system should have been tested.
**Data analysis of critical data elements, demonstrates this measure construct to be reliable and to detect meaningful differences among facilities and their patient population
**The KDQOL has proven to be reliable tool.
**Reliability in 2015 reported as "0.093" - presumably a typo and actually 0.93 given reliability in prior years;
Of potential concern is reliability in smaller facilities and whether estimates are stable;
Clarify internal reliability testing of the measure
Distribution of facility reliability scores does denote clinics below .70-0.80 - ~ 10% < 0.80 and ~ <5% < 0.70
**Reliability testing was performed and appear to demonstrate sufficient reliability.
**we have extensively tested reliability at the population level with our databases and found limited reliability at a population level. There does seem to be reliability for an individual patient over time.
**Testing measured by % implementation of survey at units >10 patients. Reliability appears good. Survey is implemented per CMS requirement.
**meets reliability tests
**The KDQOL is a tool that is widely used and has proven to be to be reliable and able to detect differences among various populations and across facilities
**Reliability testing with an adequate number of entities and patients to generalize for widespread implementation and with an appropraite method. The results do demonstrating sufficient reliability measured by Cronbach's alpha greater than .90.

2b2. Validity Testing

**Apparently tried to show that facilities with higher completion rates had higher QoL scores. While interesting, that does not seem

to be the correct way to validate this measure.
**The developers have now provided results of validity testing for the performance measure as specified.
The models were adjusted for patient-level characteristics (age, sex, race, and diabetes). The models accounted for facility clustering using, assuming a compound symmetry covariance structure.
Facilities with higher completion rates were associated with statistically significantly higher patient-level QoL scores within the facility, indicating that it is plausible that facilities with higher rates would be obtaining completed questionnaires from sicker patients, since it has been documented that individuals completing the QoL scores tend to be younger and healthier.
**Yes. The tool has been validated.
**Validity testing examined association between facility completion rates and QoL scores for five scales; unclear how patients who refused or were excluded were included in the testing (see page 28), discuss/clarify interpretation of the reported estimates as they appear small
**Developer demonstrates some association with QoL scores.
**1. The original validation work was done many years ago when dialysis technology was different and side effects were different. As a result some of the questions are no longer applicable.
2. The original validation work was done on 200 patients but that validation has not held up on larger more recent data. http://www.davitaclinicalresearch.com/wp-content/pdfs/ASN_2010/SA-PO2616-ASN2010_KDQOL_Valid_8Nov10DVW.pdf
3. There are floor and celling effects and threats to external and internal validity as a result
**The statement on validity in the document (facilities with higher completion rates...) doesn't make sense. It's not clear why a higher completion rate would result in higher QOL scores if more patients are surveyed in a unit. If younger and/or healthier patients are completing more of surveys, then, yes, scores might be skewed. If however, scoring is related to age (as well as diabetic status and gender) then that would negate any effect of younger/healthier patients on population screened at each unit. There appears to be sufficient validity to warrant continuation of use of KDQOL. Yes, scores, as used do reflect QOL.
**meets validity tests
**This is a composite measure with significant validity and consistence across patient populations and facilities
**Validity testing was done using linear mixed models with the patient level quality of life scores for each scale as the dependence variable and facility completion rate as the main independent variable. The models were adjusted for patient level characteristics (age, sex race and diabetes). The models accounted for facility clustering using, assuming a compound synmetry covariance structure. Empirical validity testing of the measure score. The developer found that facilities with higher complete rates were associated with statistically significantly higher patient -level QOL scores within the facilities--indicating that that it is plausible that facilities with higher rates would be obtaining completed questionnaires from sicker patients, since it has been documented that individual completed the QOL scores tend to be younger and healthier

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Validity testing involved correlation of HRQoL completion with with HRQoL completion rates at the facility level. I am not clear how this demonstrates validity. In my simple mind, the results of this testing do not demonstrate sufficient validity of the measure. On the other hand, I still believe (as the metric previously claimed) that there is face validity to this measure.
**Exclusions are reasonable - - not sure about evidence for these exclusions
Staff exclusion for language surely showed significant variation across facilities.... other exclusions, not as clear.
Absence of adequate denominator completeness limits the validity
Preliminary rating for validity: Low
**Checkout these two statement, appearing in this order on the submission - the only difference is the word less/more giving exact opposite meanings!

- The findings of completion rates across facilities showed that males were significantly less likely to complete the KDQOL-36 survey (rather than refuse) compared to female patients, while older patients, minorities, and those whose race is unreported (missing) were significantly more likely to complete the KDQOL-36 survey (rather than refuse) compared to white race.
- The findings of completion rates across facilities showed that males were significantly more likely to complete the KDQOL-36 survey (rather than be excluded) compared to female patients when compared to those who were excluded, while older patients, minorities, and those whose race is unreported (missing) were significantly less likely to complete the KDQOL-36 survey (rather than be excluded) when compared to white race.

Since validity was not established, I fail to see threats to the validity. And submitting this contradictory information makes this ineligible for endorsement, in my view. Rating for validity: Insufficient
**Language differences lead to exclusion which may be a factor in bias

**It is concerning that the survey is mostly filled out by younger and healthier people. Females have a higher rate of filling out than males. Confused as to why this survey has not been translated to assist Asian patients to be given the opportunity to participate.
**Data presented on the impact of exclusions
Variation in exclusions across facilities - especially for language
? Impact of revising denominator and exclusions
Exclusions have been revised since prior versions: 1) those who refuse to complete the survey are now included in the denominator - is this appropriate? ; 2) exclusions include unable to complete "due to mental status" - is this so broad as to allow for inappropriate exclusions? Is this based on assessment at the time of attempted administration or medical record (as outlined in specifications)?; 3)more clarity for language exclusions may be needed - interpretation and attempts to implement may be highly variable between facilities
**Younger patients and femeles appear more likely to complete teh KDQOL.the
**1. Existing threats to validity as mentioned above for KDQOL
2. Not clear now to handle non English speakers consistently across providers. This will create regional differences as certain areas have less English speakers (we have units that are predominately Chinese, Vietnamese, etc) This creates and SDS issue by ethnicity. The analysis provided by the developer does not address this. Better would be for each ethnicity number completed/fielded
3. Not clear how to handle patient refusals. Should a dialysis unit be penalized if a patient refuses?
**Exclusions are appropriate. Error on calculating age apparently accounts for 10% error in that exclusion. Other exclusions (e.g., significant physical disability as with quadriplegia with mental abilities intact) might be an additional reason for exclusion... a clinician's judgment.
**less response among young and old.
**Non-English speaking patients are excluded in particular those of Asian decent. More recent data shows a decline in patients completing the KDQOL survey due to survey fatigue with the addition of depression and pain screening and the ICH-CHAPSNo
**Exclusions--threats to validity. All potential threats that are relevant to the measure were empirically assessed. Validity testing was conducted with computed performance scores for each measured entity. The method was described and appropriate for assessing validity.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the 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.

- All data elements are Generated or collected by and used by healthcare personnel during the provision of care.
- There are various ways to gather the information:
  - The RAND website posts the survey instrument, instructions, and a free Excel template that staff can download to the facility computer for scoring and retention.
  - Large dialysis organizations have developed their own scoring and reporting programs.
  - KDQOL-Complete is a subscription-based data management and analysis service offers online scoring of the KDQOL-36 in multiple languages.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments: None

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- Dialysis facilities are required to offer the health-related quality of life survey and use the results of the survey plan care for patients under 42 CFR 494.90(a)(6) in the ESRD Conditions for Coverage (CfC). State Agency ESRD surveyors review facilities requesting recertification and cite facilities that are not in compliance with the ESRD CfC.
- Developer is in contact with CROWNWeb in order to incorporate the measure.
- The Kidney Disease Quality of Life (KDQOL) survey is administered by IMPAQ International, LLC, on behalf of the Centers for Medicare & Medicaid Services (CMS), to assess self-reported quality of life among Medicare end-stage renal disease (ESRD) beneficiaries included in the Comprehensive ESRD Care (CEC) Model.
- No specific plans for public reporting are included.

Improvement results:

- In 2013, 62,620 patients out of 74,410 eligible (not excluded) patients (84.2%) completed the survey. In 2014 66,386 patients out of 78,401 eligible (not excluded) patients (84.7%) completed the survey. In 2015, 69,746 patients out of 82,221 eligible (not excluded) patients (84.8%) completed the survey.
- If completion rates are calculated within each facility and averaged over all facilities, completion rates were 85.4% in 2013, 85.5% in 2014, and 85.6% in 2015.

Unexpected findings (positive or negative) during implementation: The developer reports no challenges or unexpected findings in implementation.

Potential harms: The developer reports no unintended consequences were noted during testing.

Feedback : No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

- *Would this measure be more useful and meaningful to patients and other stakeholders if the QoL results were reported as a PRO-PM?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement

4c. Unintended Consequences

**Is required to be offered - -

It would indeed be more valuable if the results of the measure rather than the frequency of completion of the measure were the focus of a QoL metric.

Preliminary rating: Moderate

**There is accountability to CMS that testing be done, but that existed before the measure and is independent of the measure. I doubt public reporting would matter since it must be done for compliance reasons. So I don't see much use or usability for this measure

**Dialysis facilities are required to offer the health-related quality of life survey and use the results of the survey plan care for patients under 42 CFR 494.90(a)(6) in the ESRD Conditions for Coverage (CfC). State Agency ESRD surveyors review facilities requesting recertification and cite facilities that are not in compliance with the ESRD CfC.

Developer is in contact with CROWNWeb in order to incorporate the measure.

The Kidney Disease Quality of Life (KDQOL) survey is administered by IMPAQ International, LLC, on behalf of the Centers for Medicare & Medicaid Services (CMS), to assess self-reported quality of life among Medicare end-stage renal disease (ESRD) beneficiaries included in the Comprehensive ESRD Care (CEC) Model.

However no specific plans for public reporting are included.

There are no unintended consequences.

**It is important to measure Quality of Life and report publicly. The scope of this KDQOL 36 is very broad to be a performance measure fro dialysis facility and should remain a process measure. Individual questions on KDQOL that pertain to the patients treatment or education of a treatment option that will better serve the patient should be the focus.

**Current use includes QI with benchmarking for facilities enrolled in KDQOL-Complete; Not currently publically reported

**This measure is not currently publicly reported or in an accountability program. However, this is planned.

**As a process metric of surveys fielded, this measure has little public use. And since there are no established action plans to improve quality of life (especially when fielded cross sectionally at a given point in time in a patients dialysis treatment) it will be hard for the public to use this data to make purchasing decisions.

**Important measure and the only one that provides feedback from patient on his/her QOL. Completion data are usually reported in unit QAPI documents/discussions. Goal would be to increase # of patients taking surveys and decrease refusals. Does minimum completion goal need to be set eventually? What should that level be? Ultimate goal (not discussed in this proposal) would be to use results consistently to help dialysis patients increase their QOL. This is what SW should be and usually is doing.

**Not publicly reported.

**Process measure is reported as number of patients competing, refusing or not eligible for the survey.

Unintended consequence is the number of surveys now required for patients to complete. Number of patients completing surveys is declining.

**Measure is not being publicly reported. Accountability applications is being used for to help improve patient's quality of life with interventions. Clinical processes typically include mutliple steps: assess-ID problem/potential problem-choose/plan interviention with patient input-provide intervention-evaluate impact on health status.

Criterion 5: Related and Competing Measures
Related or competing measures • None identifed

Pre-meeting public and member comments

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5677: Assessment of Health-Related Quality of Life (QoL) in Dialysis Patients (Witten and Associates, LLC)

KCP recognizes the importance of assessing the health-related quality of life for individuals with ESRD. Nevertheless we have an overarching concern about the measure, as well as specific concerns about the new specifications, evidence, performance gap, and validity.

OVERARCHING ISSUE. Annual administration of the KDQOL is already required by Federal regulation, the

Conditions for Coverage. KCP questions how endorsement of a measure for a process that is already mandated and surveyed will further improve patient care.

SPECIFICATIONS. We support the changes to the exclusions that align them with the Conditions for Coverage, but KCP opposes eliminating the exclusion for patient refusal. First, the Conditions for Coverage permit patient refusal as long as it is documented. We believe approving a measure that directly conflicts with Federal regulation is problematic. Second, not accepting patient decisionmaking ignores patient autonomy; providers should not be forced to face intruding on patient decisionmaking vs. facing a penalty for poorer performance on this measure. We further note there is no performance gap when the specifications include patient refusal.

EVIDENCE. As noted, KCP recognizes the importance of assessing health-related quality of life, but questions the lack of direct evidence for the measure. The developer cites KDOQI and the Institute of Medicine on the importance of functional assessment, however no peer-reviewed, empirical evidence is provided that the specifications (i.e., annual completion rate) are associated with higher quality.

PERFORMANCE GAP. Based on the updated specifications, the performance range in 2015 was 16.7%-100%, with a median of 91.8% using "KDQOL-Complete" (K-C) data. Although the performance rate at the patient-level with the updated exclusion criteria (i.e., refusals = fail) is 84.8% (2015), 84.7% (2014), and 84.2% (2013), the performance rate with refusals as an exclusion (old specifications) is 100% in 2013, 2014, 2015. KCP also further examined the data and notes the refusal exclusion appears stable over this period. We posit the change in specifications creates a gap where otherwise none exists, as well as puts the measure in conflict with the Conditions for Coverage.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5678: KCP 0260 comments, cont.

VALIDITY. KCP has two concerns about the measure's validity: the validity testing and the lack of risk adjustment.

The developer performed validity testing on a sample that included all patients—i.e., those who refused, those who completed the survey, and those who met the exclusion criteria. It assessed association of completion with patients' KDQOL scores (linear fixed models with the score for each of the five scales as the dependent variable and facility completion rate as the main independent variable). The models adjusted for patient-level characteristics of age, gender, race, and diabetes. Based on this, it appears the measure was not tested as specified. First, all patients were used, even those who qualify for exclusions. Second, associations were examined, but the models were adjusted for patient-level characteristics even though the measure itself is not adjusted. Performance on the measure cannot be asserted as being associated with better quality (the five KDQOL scales) if the measure as specified is not used.

The developer also notes, "This finding [association between completion and scores] is important because it is plausible that facilities with higher rates would be obtaining completed questionnaires from sicker patients, since it has been documented that individuals completing the QoL scores tend to be younger and healthier." Again, the developer draws this conclusion from analyzing a different data set and a risk-adjusted model. The measure is not whether an all-population, risk-adjusted measure of completion validates against the scale results: Testing and demonstration of validity must be of the measure as specified.

Finally, KCP has expressed concern about NQF 0260 in other contexts (e.g., use in CMS' Comprehensive ESRD

Care Initiative) because of the lack of risk adjustment for case mix. In fact, the developer's data demonstrate that case mix impacts a facility's score. Specifically, the developer presents data on the distribution of patient characteristics and the facility-level survey completion rate; the analysis uses refusals and completions, so comports with the proposed specifications. Facilities with more males will score, on average, 0.45% lower (per 10% difference) compared to facilities that have fewer males. Conversely, facilities with higher proportions of Asians—likely to exist in certain geographic areas—will score higher. We believe the lack of adjustment for the measure presents a significant threat to validity, particularly given a median performance of 91.8% with the updated specifications.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0260
Measure Title: Assessment of Health-related Quality of Life in Dialysis Patients
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: NA Click here to enter composite measure #/ title
Date of Submission: 4/21/20164/21/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☒ Process: Assessment of Health-related Quality of Life in Dialysis Patients

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

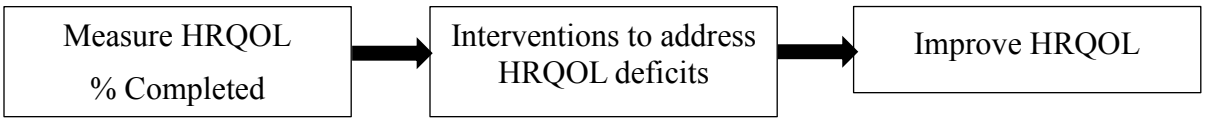
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):
KDOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification; GUIDELINE 12. ASSOCIATION OF LEVEL OF GFR WITH INDICES OF FUNCTIONING AND WELL-BEING (2002) - http://www2.kidney.org/professionals/KDOQI/guidelines_ckd/p6_comp_g12.htm

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

GUIDELINE 12. ASSOCIATION OF LEVEL OF GFR WITH INDICES OF FUNCTIONING AND WELL-BEING

Impairments in domains of functioning and well-being develop during the course of chronic kidney disease and are associated with adverse outcomes. Impaired functioning and well-being may be related to sociodemographic factors, conditions causing chronic kidney disease, complications of kidney disease, or possibly directly due to reduced GFR.

- *Patients with GFR <60 mL/min/1.73 m^2 should undergo regular assessment for impairment of functioning and well-being:*
  - *To establish a baseline and monitor changes in functioning and well-being over time*
  - *To assess the effect of interventions on functioning and well-being.*

"CLINICAL APPLICATIONS

The conferees at the Institute of Medicine (IOM) Workshop "Assessing Health and Quality of Life Outcomes in Dialysis" recommended that ESRD providers:

- *Assess functioning and well-being in kidney disease using standardized survey instruments that are valid, reliable, responsive to changes, easily interpretable, and easy to use, such as the Dartmouth COOP Charts, the Duke Health Profile/Duke Severity of Illness (DUKE/DUSOI), Medical Outcomes Study 36-Item Short Form (SF-36), or the Kidney Disease Quality of Life (KDQOL).*
- *Assess patient functioning and well-being early in chronic kidney disease to establish a baseline, to maintain or improve health status, and to manage the disease continuum by linking clinical and health outcomes with functional status outcomes.*[454]

Data reported in the reviewed studies suggest that decreased kidney function affects patients' functioning and well-being through several dimensions. Deficits in functioning are reported by patients even at early stages of chronic kidney disease, and persist even after transplantation. The implications of these findings are:

- *Clinicians should assess functional status and well-being as soon as possible after referral in order to obtain baseline data and enable early intervention to improve functioning and well-being.*
- *Clinicians should regularly reassess functioning and well-being to ascertain the patient's current status and the effectiveness of interventions to improve functioning and well-being. Reassessment is needed when a patient reports increased frequency or severity of symptoms, has a new complication of kidney disease, has an access for dialysis placed, starts dialysis, changes modality, or participates in a clinical or rehabilitation intervention (eg, counseling, peer support, education, physical therapy or independent exercise, or vocational rehabilitation).*

These recommendations are based on the opinions expressed by the authors of most of the studies reviewed for this guideline, as well as those of recognized experts in functioning and health status outcomes measurement who attended the IOM Workshop.

National Kidney Foundation, K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification. Am J Kidney Dis 39:S1-S000, 2002 (suppl 1), p 161-169.

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

- *"Indices of functioning and well-being are impaired in chronic kidney disease (Strength of Evidence – (R) Review of reviews & selected original articles*
- *Impairment in indices of functioning and well-being are associated with worse outcomes in chronic kidney disease – (R) Review of reviews & selected original articles*
- *Impairment in functioning and well-being are associated with sociodemographic characteristics - Review of reviews & selected original articles*
- *Impairment in functioning and well-being may be due to conditions that cause chronic kidney disease (such as diabetes or hypertension) or complications of decreased GFR (such as anemia, malnutrition, bone disease or neuropathy) – (R) Review of reviews & selected original articles*
- *Indices of functioning and well-being are related to the level of GFR; below a GFR of approximately 60 ml/min/1.73 m2, there is a higher prevalence of impairments in indices of functioning and well-being – (S) analysis of individual*

patient data from a single large, generalizable study of high methodological quality; (C) compilation of original articles (evidence tables)"

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)
*"**S:** Analysis of individual patient data from a single large, generalizable study of high methodological quality*
***C:** Compilation of original articles*
***R:** Review of reviews & selected original articles*
***O:** Opinion"*
(From NKF Guideline 12)
1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?
☒ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION
1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):
1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.
1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:
1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)
1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**
Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE
1a.6.1. Citation (*including date*) and **URL** (*if available online*):
1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**
Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE
If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?
"Impairments in domains of functioning and well-being develop during the course of chronic kidney disease and are associated with adverse outcomes. Impaired functioning and well-being may be related to sociodemographic factors, conditions causing chronic kidney disease, complications of kidney disease, or possibly directly due to reduced GFR.

- *Patients with GFR <60 mL/min/1.73 m^2 should undergo regular assessment for impairment of functioning and well-being:*
  - *To establish a baseline and monitor changes in functioning and well-being over time*
  - *To assess the effect of interventions on functioning and well-being."*

National Kidney Foundation, K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification. Am J Kidney Dis 39:S1-S000, 2002 (suppl 1)

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:
In text citations are in table in 1a.7.5

- ***"Indices of functioning and well-being are impaired in chronic kidney disease - (R) Review of reviews & selected original articles***

Dialysis patients report significantly more bodily pain, lower vitality, poorer general health, greater physical, mental, and social dysfunction, and greater limitations in their ability to work and participate in activities due to their health and emotions than the US reference norm. At least 25% are depressed.[455] Dialysis patients' exercise capacity is significantly worse than that of healthy controls.[456] Kidney failure negatively affects sense of control and health outlook in those on dialysis.[457] About 39% of those who worked full or part-time 6 months before dialysis do not continue working when they start dialysis.[4] Elderly people on dialysis engage in few previously enjoyed activities outside their homes and many leave home only for dialysis because of weakness.[458]

- ***Impairment in indices of functioning and well-being are associated with worse outcomes in chronic kidney disease – (R) Review of reviews & selected original articles***

Impaired functioning and well-being in dialysis patients is linked to increased risk of death and hospitalization while improvement in scores has been associated with better outcomes. Patients with SF-36 Physical Component Summary (PCS) scores <34.6 had a 2.03 relative risk of dying and a 1.67 relative risk of being hospitalized. Each 5-point improvement in PCS scores was associated with 10% longer survival and 6% fewer hospital days. On the SF-36, a Mental Health scale score ≤52 and a Mental Component Summary (MCS) score ≤42 indicate depression. Each 5-point improvement in the MCS score is associated with 2% fewer hospital days.[455]

- ***Impairment in functioning and well-being are associated with sociodemographic characteristics - Review of reviews & selected original articles***

Low income and low education were associated with greater impairments in functioning and well-being in patients with chronic kidney disease.[459]

- ***Impairment in functioning and well-being may be due to conditions that cause chronic kidney disease (such as diabetes or hypertension) or complications of decreased GFR (such as anemia, malnutrition, bone disease or neuropathy) – (R) Review of reviews & selected original articles***

Hypertension, diabetes with angina, prior cardiac infarction,[460] osteoporosis, bone fractures,[461] and malnutrition[462] have been shown to impair functioning and well-being in those with no known kidney disease. Among veterans with diabetes, neuropathy and kidney disease have been associated with the greatest decrease in functioning and well-being.[463] Anemia has been linked to poor functioning and well-being in patients with severely decreased GFR and dialysis patients, and improving anemia with erythropoietin has been linked to improvement in functioning and well-being.[284,464-468]

- ***Indices of functioning and well-being are related to the level of GFR; below a GFR of approximately 60 ml/min/1.73 m2, there is a higher prevalence of impairments in indices of functioning and well-being – (S) analysis of individual patient data from a single large, generalizable study of high methodological quality; (C) compilation of original articles (evidence tables)***

Data from cross-sectional studies and baseline data from longitudinal studies were reviewed to assess the relationship between level of kidney function and level of functioning and well-being. Populations studied include those with decreased kidney function, including those with functioning transplants, and dialysis patients when compared with healthy subjects or kidney transplant recipients. While much of the data on functioning and well-being related to outcomes have been obtained in dialysis patients, there is convincing evidence that abnormalities in functioning and well-being begin earlier in chronic kidney disease and may well be related to declining GFR.

SYMPTOMS: *Reduced kidney function is associated with increasing symptoms such as tiring easily, weakness, low energy, cramps, bruising, bad tasting mouth, hiccoughs, and poor odor perception. This is true in patients with native kidney disease and those with kidney transplants. Diabetic dialysis and transplant patients are more likely to report poor health than dialysis or transplant patients who do not have diabetes.*

PHYSICAL FUNCTIONING: *Decreased GFR in NHANES III subjects is associated with impaired walking and lifting ability. In transplant recipients, reduced kidney function is also associated with poorer physical function scores. In one study of patients with decreased GFR, impairment in physical function was not significantly related to the level of kidney function, but physical impairment was 8 times worse than in the general population. Dialysis patients report greater physical dysfunction than transplant recipients and diabetic dialysis and transplant patients are more likely to report physical dysfunction than those patients who do not have diabetes.*

DEPRESSION: *Reduced kidney function is associated with poorer psychosocial functioning, higher anxiety, higher distress, decreased sense of well-being, higher depression, and negative health perception. Depressed patients are more likely to report poor life satisfaction, irrespective of kidney function. Dialysis patients report significantly lower "happiness with*

personal life" and lower psychosocial functioning than transplant recipients. In elderly Mexican Americans, kidney disease has been found to be predictive of depressive symptoms.
EMPLOYMENT AND USUAL ACTIVITIES: *Reduced kidney function is associated with lower employment. In those with chronic kidney disease and GFR <50, the presence of physical dysfunction is significantly related to unemployment, but the association to kidney function is not significant since physical dysfunction is not uniformly present. Full-time employment is higher for those with decreased GFR (mean serum creatinine 5.4 mg/dL, 69%) compared with those with kidney failure (mean serum creatinine 13.7 mg/dL, 12%). More dialysis patients report their health limits work and other activities than those with functioning transplants. Dialysis and transplant patients with diabetes are more likely to report difficulty working than dialysis and transplant patients without diabetes.*
SOCIAL FUNCTIONING: *Reduced kidney function is associated with reduced social activity, social functioning, and social interaction. Dialysis patients report fewer neighborhood acquaintances, social contacts, and worse social well-being than healthy individuals while transplant recipients report higher social function and social interaction than those on dialysis. Diabetics on dialysis or with transplants are more likely to report problems with social interaction than nondiabetic patients. Level of perceived social support in chronic kidney disease is not associated with the level of kidney function."*

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.
See 1A.4.4
REVIEW OF EVIDENCE (For the NKF Guideline)
"The guidelines developed by the Work Group are based on a systematic review of the literature using an approach based on the procedure outlined by the Agency for Healthcare Research and Quality (formerly the Agency for Health Care Policy and Research) with modifications appropriate to the goals. An Evidence Review Team was appointed by the NKF to collaborate with the Work Group to conduct a systematic review of the literature on which to base the guidelines."

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: 1989-2001

QUANTITY AND QUALITY OF BODY OF EVIDENCE
1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)
For KDOQI Guideline 12, Association of Level of GFR with Indices of Functioning and Well-being, the Evidence Review Team screened 1,795 abstracts, retrieved 107 articles, added 4. The table below lists 28 articles cited in the guideline from prospective and retrospective studies, 2 descriptive studies, 2 cross-sectional studies, 4 comparison studies, 3 cohort studies, 7 clinical trials (1 Phase III, 2 randomized controlled clinical trials), 1 epidemiological, 2 case control studies, 3 reviews and 1 report.

JOURNAL ARTICLE CITATION	Design
80. Sesso R, Yoshihiro MM: Time of diagnosis of chronic renal failure and assessment of quality of life in haemodialysis patients. Nephrol Dial Transplant 12:2111-2116, 1997	Comparison
452. Kutner NG: Assessing end-stage renal disease patients' functioning and well-being: measurement approaches and implications for clinical practice. Am J Kidney Dis 24:321-333, 1994	Review
453. Tarlov AR, Ware JE Jr, Greenfield S, et al.: The Medical Outcomes Study. An application of methods for monitoring the results of medical care. JAMA 262:925-930, 1989	Cross-sectional Study
454. Rettig RA, Sadler JH, Meyer KB, et al.: Assessing health and quality of life outcomes in dialysis: A report on an Institute of Medicine workshop. Am J Kidney Dis 30:140-155, 1997	Report
455. DeOreo PB: Hemodialysis patient-assessed functional health status predicts continued survival, hospitalization, and dialysis-attendance compliance. Am J Kidney Dis 30:204-212, 1997	Prospective Study
456. Johansen KL: Physical functioning and exercise capacity in patients on dialysis. Adv Renal Repl Ther 6:141-148, 1999	Review
457. Kutner NG, Cardenas DD, Bower JD: Rehabilitation, aging and chronic renal disease.	Comparison

JOURNAL ARTICLE CITATION	Design
Am J Phys Med Rehabil 71:97-101, 1992	
458. Ifudu O, Mayers J, Matthew J, et al.: Dismal rehabilitation in geriatric inner-city hemodialysis patients. JAMA 271:29-33, 1994	Cohort Study
459. Harris LE, Luft FC, Rudy DW, et al.: Clinical correlates of functional status in patients with chronic renal insufficiency. Am J Kidney Dis 21:161-166, 1993	Randomized Controlled Clinical Trial
460. Bardage C, Isacson DG: Hypertension and health-related quality of life. An epidemiological study in Sweden. J Clin Epidemiol 54:172-181, 2001	Epidemiolog-ical
461. Hall SE, Criddle RA, Comito TL, et al.: A case-control study of quality of life and functional impairment in women with long-standing vertebral osteoporotic fracture. Osteoporos Int 9:508-515, 1999	Case Control Study
462. Schneider SM, Pouget I, Staccini P et al.: Quality of life in long-term home enteral nutrition patients. Clin Nutr 19:23-28, 2000	Retrospective study
463. Ahroni JH, Boyko EJ: Responsiveness of the SF-36 among veterans with diabetes mellitus. J Diabetes Complications 14:31-39, 2000	Cohort Study
464. Beusterien KM, Nissenson AR, Port FK, et al.: The effects of recombinant human erythropoietin on functional health and well-being in chronic dialysis patients. J Am Soc Nephrol 7:763-773, 1996	Phase III Clinical Trial Comparative
465. Revicki DA, Brown RE, Feeny DH, et al.: Health-related quality of life associated with recombinant human erythropoietin therapy for predialysis chronic renal disease patients. Am J Kidney Dis 25:548-554, 1995	Randomized Clinical Trial
466. Eschbach JW, Abdulhadi MH, Browne JK, et al.: Recombinant human erythropoietin in anemic patients with end-stage renal disease. Results of a phase III multicenter clinical trial. Ann Intern Med 111:992-1000, 1989	Controlled Clinical Trial
467. Evans RW, Rader B, Manninen DL: The quality of life of hemodialysis recipients treated with recombinant human erythropoietin. Cooperative Multicenter EPO Clinical Trial Group. JAMA 263:825-330, 1990	Clinical Trial
468. Moreno F, Aracil F, Perez R, et al.: Controlled study on the improvement of quality of life in elderly hemodialysis patients after correcting end-stage renal disease-related anemia. Am J Kidney Dis 27:548-556, 1996	Controlled Clinical Trial
469. Rocco MV, Gassman JJ, Wang SR, et al.: Cross-sectional study of quality of life and symptoms in chronic renal disease patients: The Modification of Diet in Renal Disease Study. Am J Kidney Dis 29:888-896, 1997	Cross-sectional Study
470. Korevaar JC, Jansen MA, Merkus MP, et al.: Quality of life in predialysis end-stage renal disease patients at the initiation of dialysis therapy. The NECOSAD Study Group. Perit Dial Int 20:69-75, 2000	Prospective Cohort Study
471. Shidler NR, Peterson RA, Kimmel PL: Quality of life and psychosocial relationships in patients with chronic renal insufficiency. Am J Kidney Dis 32:557-566, 1998	Descriptive Study
472. Klang B, Bjorvell H, Clyne N: Quality of life in predialytic uremic patients. Qual Life Res 5:109-116, 1996	Case-Control Study
473. Fujisawa M, Ichikawa Y, Yoshiya K, et al.: Assessment of health-related quality of life in renal transplant and hemodialysis patients using the SF-36 health survey. Urology 56:201-206, 2000	Comparison
474. Griep MI, Van der Niepen P, Sennesael JJ, et al.: Odour perception in chronic renal disease. Nephrol Dial Transplant 12:2093-2098, 1997	Control Trial
475. Sacks CR, Peterson RA, Kimmel PL: Perception of illness and depression in chronic renal disease. Am J Kidney Dis 15:31-39, 1990	Descriptive

JOURNAL ARTICLE CITATION	Design
476. Manninen DL, Evans RW, Dugan MK: Work disability, functional limitations, and the health status of kidney transplantation recipients posttransplant. Clin Transplant 5:193-203, 1991	Cohort
477. Churchill DN, Torrance GW, Taylor DW, et al.: Measurement of quality of life in end-stage renal disease: the time trade-off approach. Clin Invest Med 10:14-20, 1987	Comparison
478. Black SA, Goodwin JS, Markides KS: The association between chronic diseases and depressive symptomatology in older Mexican Americans. J Gerontol Series A Biol Sci Med Sci 53:M188-M194, 1998	Descriptive
479. Cagney KA, Wu AW, Fink NE, et al.: Formal literature review of quality of life instruments used in end stage renal disease. Am J Kidney Dis 36:327-336, 2000	Review

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

"LIMITATIONS AND EXCEPTIONS: *Most study samples were not randomly selected. Medication usage was not reported even if medications (eg, anti-depressants) could affect outcomes. Seven of 12 studies did not provide full information on patient demographics. Three studies reported differences between groups of very unequal sizes and one reported percentages but did not report whether observed differences were statistically significant.*

Historically, there has been no "gold standard" definition for quality of life or functioning and well-being. Researchers have studied multiple variables using standardized and non-standardized instruments. Thus, results are not comparable to one another.[479] With lack of instrument comparability, findings appear to be conflicting. Many studies have examined the relationships between functioning and well-being and treatment modalities after the onset of kidney failure. Few studies of persons with decreased GFR have examined the relationship between level of GFR and functioning and well-being. Three of the studies of individuals with decreased GFR had such severely restrictive inclusion criteria for level of kidney function that functioning and well-being deficits were already present. Of the 12 studies reported, 3 reported no measure of kidney function and 2 reported only serum creatinine, a less reliable measure of kidney function than GFR or creatinine clearance. Most of the studies reported only mean values for kidney function. Only the MDRD Study and NHANES III examined functioning and well-being at a wide range of levels of kidney function. Precise statements about how early deficits in domains of functioning and well-being occur as kidney function deteriorates require this essential data. Finally, since anemia has been shown to limit functioning and well-being, inadequate anemia management in studies conducted prior to the widespread use of erythropoietin could have affected outcomes. Therefore, recent functioning and well-being outcomes may not be comparable to outcomes reported in studies prior to 1989 even if the same instruments were used.'

Table 6. Approach to the Evidence Review **(from NKF Guideline 12)**
Develop and refine topics;
Determine approach to topics: *Established concepts - summary of published reviews and selected original articles;* *New concepts - systematic review of original articles and analysis of primary data, if available.*
Retrieval of evidence (literature review);
Analysis of primary data from the Third National Health and Nutrition Examination Survey (NHANES III) and other sources;
Evaluation of evidence (types and quality);
Synthesis of evidence (tables);
Translation of evidence into clinical practice guidelines;
Identification of guidelines suitable for translation into clinical performance measures;

Public review and revisions;
Approval by Board of Directors of the NKF

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

From the guideline: *"Assess patient functioning and well-being early in chronic kidney disease to establish a baseline, to maintain or improve health status, and to manage the disease continuum by linking clinical and health outcomes with functional status outcomes.*[454]*"*

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

No adverse effects were reported.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

PubMed search, national renal meeting abstracts and posters, and Fresenius Kidney Care.

1a.8.2. Provide the citation and summary for each piece of evidence.

Johnstone S, Dombro L, Garza GS, et al. *Declines in hemodialysis patients' physical and mental component scores before death* [Abstract] J Am Soc Nephrol 25, 2014: 58A-59A

4,227 Fresenius patients completed KDQOL surveys prior to death. Both PCS & MCS scores declined in patients before death. Decline in PCS score appears more linear, decline in MCS score is more pronounced approximately 6 months before death.

Johnstone S, Li N, Maddux FW, Weissman-Hunt AR, et al. *Reducing hemodialysis therapy non-adherence: A social worker initiated program* [Abstract]. J Am Soc Nephrol 25, 2014: 310A-311A

348 patients completing 6 of 8 session intervention from 195 FMCNA (Fresenius) centers from 2/1/13-10/30/13, with 6-mo baseline & 6-mo follow-up period. Mean age 54 years, 47% males; 61% white, 35% black; 56% DM, 17% CAD, and 36% CHF. Baseline vs follow-up missed treatment rate 146 vs. 127/100 pt mos (p=0.008). Significant improvement (p<0.05) pre to post in CESD-10, Financial/Insurance Stressor, Family/ Relationship Stressor, Health Symptoms Stressor, Loss/Grief Stressor, Difficulty Falling Asleep, Difficulty Staying Asleep, Difficulty Awakening, Interrupted Sleep, Restless Legs, PCS (35.7 vs. 37.7), MCS (46.3 vs. 49.2), Burden (60 vs. 65.8), Symptoms (69.6 vs. 76.7), and Kidney Disease Effects (54 vs. 62.6).

Ricardo AC, Hacker E, Lora CM, et al. *Validation of the Kidney Disease Quality of Life Short Form 36 (KDQOL-36) US Spanish and English versions in a cohort of Hispanics with chronic kidney disease.* Ethn Dis. 2013 Spring;23(2):202-9.

420 Hispanic (150 English, 270 Spanish-speakers), and 409 non-Hispanic White individuals, matched by age (mean 57 years), sex (60% male), kidney function (mean estimated GFR 36ml/min/1.73m2), and DM (70%). Reliability of each KDQOL-36 subscale [PCS and MCS, Symptoms/Problems, Burden of Kidney Disease & Effects of Kidney Disease] was very good (Cronbach's alpha >0.8). Construct validity supported by negative correlation between MCS & Beck Depression Inventory in all three subgroups (r=-0.56 to -0.61, P<.0001). Inverse correlation between Symptoms/Problems & the Patient Symptom Form (r= -0.70 to -0.77, P<.0001). Significant, positive correlation between PCS & a physical activity survey (r=+0.29 to +0.38, P< or =.003); & between the PCS & MCS scores & the Kansas City Questionnaire (r= +0.31 to +0.64, P<.0001). Reliability & validity were similar across all racial/ethnic groups analyzed separately. Findings support the use of KDQOL-36 as a measure of HRQOL in this cohort of US Hispanics with CKD.

Breckenridge K, Bekker HL, Gibbons E, et al. How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. Nephrol Dial Transplant. 2015 Oct;30(10):1605-14.

Survey sent to 45 European renal registries prior to an ERA-EDTA-QUEST-funded consensus meeting on how to routinely collect PROMs/PREMs in European renal registries. Of 45 registries, 23 responded. "A systematic review identified 157 potentially relevant articles of which 9 met the inclusion criteria and were analysed for barriers and facilitators to routine PROM/PREM collection. Thirteen themes were identified and mapped to a three-stage framework around establishing the need, setting up and maintaining the routine collection of PROMs/PREMs." Participants voted on PROMs/PREMs. "Of all the instruments discussed, the KDQOL™-36 seemed to be preferred by delegates as it offers both generic and disease specific outcomes…There was clear consensus that the aim should be to include all patients on RRT in any PROMs/PREMs programme, and that data should be collected on at least an annual basis."

Dunn D. DaVita seeing success with STI approach. Nephrol News Issues. 2016 Apr;30(4):31.

DaVita provided in-person training, resource materials, and tracking tools to 1,800 social workers over 11 months. Social workers provided interventions to over 2,000 patients through the Empowering Patients Program. Results to date show fewer missed treatments, reduced depression (CES-D), and overall improvement in quality of life (KDQOL-36) according to DaVita Kidney Care internal data.

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0260-MeasSubmEvidence-2016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

Assessing functioning and well-being (health-related quality of life) will help health care providers monitor the impact of kidney disease on the daily lives of persons on dialysis.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Using KDQOL-Complete data, we are able to test whether there are statistically significant differences in the performance measure between facilities with at least 10 patients.

2013

N=1,240; Mean=85.4; Median=91.8; Range=0-100; Interquartile Range=78.3-100; 100% Max=100; 99%=100; 95%=100; 90%=100; 75% Q3=100; 50% Median=91.8; 25% Q1=78.3; 10%=59.4; 5%=46.9; 1%=28.1; 0% Min=0

2014

N=1,222; Mean=85.5; Median=92.1; Range=0-100; Interquartile Range=78.3-98.1; 100% Max=100; 99%=100; 95%=100; 90%=100; 75% Q3=98.1; 50% Median=92.1; 25% Q1=78.3; 10%=61.4; 5%=50.0; 1%=29.1; 0% Min=0

2015

N=1,261; Mean=85.6; Median=91.8; Range=16.7-100; Interquartile Range=78.3-100; 100% Max=100; 99%=100; 95%=100; 90%=100; 75% Q3=97.9; 50% Median=91.8; 25% Q1=78.3; 10%=61.2; 5%=48.9; 1%=29.3; 0% Min=16.7

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.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

We used the full KDQOL-Complete dataset, counting only the first KDQOL-36 survey for patients who completed the survey more than once in a calendar year. We compared the data for All, Completed, Refused, and Excluded for 2013, 2014, 2015.

2013:

All-79,872; Mean Age 62.4+ (14.9); Male-55.5%; White-45.4%; Black-26%; Asian-4.5%; Native Am-1.6%; Pacific Island-1.8%; Missing-20.8%; Diabetes 52.7%

Completed-62,620; Mean Age 61.9 (14.7); Male-55.7%; White-46.3%; Black-25.9%; Asian-4.3%; Native Am-1.8%; Pacific Island-1.8%; Missing-20.2%; Diabetes 52.8%

Refused-11,790; Mean Age 61.9 (15.1); Male-49.9%; White-43.7%; Black-25.3%; Asian-3.7%; Native Am-1.3%; Pacific Island-2.4%; Missing-23.5%; Diabetes 51.6%

Excluded-5,462; Mean Age 69.2 (14.5); Male-59.1%; White-38.6%; Black-28.5%; Asian-8.0%; Native Am-2.6%; Pacific Island-1.3%; Missing-21.0%; Diabetes 54.6%

2014

All-84,167; Mean Age 62.4 (14.9); Male-56.3%; White-45.4%; Black-25.8%; Asian-5.5%; Native Am-1.5%; Pacific Island-1.9%; Missing-19.9%; Diabetes 52.6%

Completed-66,386; Mean Age 62.0 (14.7); Male-56.2%; White-46.6%; Black-25.8%; Asian-5.4%; Native Am-1.5%; Pacific Island-1.8%; Missing-18.9%; Diabetes 52.9%

Refused-12,015; Mean Age 61.7 (15.1); Male-51.2%; White-42.5%; Black-25.5%; Asian-4.4%; Native Am-1.5%; Pacific Island-1.7%; Missing-24.3%; Diabetes 50.5%

Excluded-5,756; Mean Age 69.2 (14.6); Male-59.1%; White-37.8%; Black-27.4%; Asian-9.4%; Native Am-0.6%; Pacific Island-2.5%; Missing-22.3%; Diabetes 54.1%

2015

All-87,892; Mean Age 62.4 (14.7); Male-56.4%; White-45.1%; Black-26.6%; Asian-5.4%; Native Am-1.5%; Pacific Island-1.9%; Missing-19.5%; Diabetes 52.5%

Completed-69,746; Mean Age 62.0 (14.5); Male-56.4%; White-45.9%; Black-27.0%; Asian-5.1%; Native Am-1.5%; Pacific Island-1.9%; Missing-18.7%; Diabetes 52.4%

Refused-12,475; Mean Age 61.7 (15.0); Male-59.2%; White-43.3%; Black-24.7%; Asian-5.2%; Native Am-1.7%; Pacific Island-1.9%; Missing-23.1%; Diabetes 51.6%

Excluded-5,671; Mean Age 69.5 (14.6); Male-50.4%; White-38.8%; Black-26.5%; Asian-9.9%; Native Am-0.7%; Pacific Island-2.2%; Missing-21.9%; Diabetes 55.3%

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

The United States Renal System reported that in 2013 over 468,000 patients were on dialysis. African Americans, Hispanics, Pacific Islanders, Native Americans and older Americans are at increased risk. Kidney disease is a leading cause of hospitalization and death.

Low HRQOL scores predict higher relative risk of death and hospitalization. Healthy People 2020 will monitor progress toward improving HRQOL as one of its 4 foundation health measures.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

United States Renal Data System. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015.
Lopes AA, Bragg-Gresham JL, Satayathum S, McCullough K, Pifer T, Goodkin DA, Mapes DL, Young EW, Wolfe RA, Held PJ, Port FK; Worldwide Dialysis Outcomes and Practice Patterns Study Committee. (2003). Health-related quality of life and associated outcomes among hemodialysis patients of different ethnicities in the United States: the Dialysis Outcomes and Practice Patterns Study (DOPPS). Am J Kidney Dis 2003 Mar; 41(3):605-15.
Mapes DL, Lopes AA, Satayathum S, McCullough KP, Goodkin DA, Locatelli F, Fukuhara S, Young EW, Kurokawa K, Saito A, Bommer J, Wolfe RA, Held PJ, Port FK. (2003). Health-related quality of life as a predictor of mortality and hospitalization: the Dialysis Outcomes and Practice Patterns Study (DOPPS). Kidney Int. 2003 Jul; 64(1):339-49.
Walters BA, Hays RD, Spritzer KL, Fridman M, Carter WB. (2002). Health-related quality of life, depressive symptoms, anemia, and malnutrition at hemodialysis initiation. American Journal of Kidney Disease, 40 (6), 1185-94.
HealthyPeople.gov

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

The ESRD Conditions for Coverage published in 2008 requires administration of a "standardized mental and physical assessment tool...at regular intervals, or more frequently on an as-needed basis" at 42 CFR 494.90(a)(6). Further, Dialysis Therapies: A National Dialogue published by Mehrotra et al. in CJASN in 2014 states, "The National Institute of Diabetes, Digestive, and Kidney Diseases–supported Kidney Research National Dialogue asked the scientific community to formulate and prioritize research objectives that would improve our understanding of kidney function and disease. Kidney Research National Dialogue participants identified the need to improve outcomes in ESRD by decreasing mortality and morbidity and enhancing quality of life as high priority areas in kidney research."

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

Health and Functional Status : Functional Status

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.)*

www.rand.org/health/surveys_tools/kdqol.html

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)*

No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
- We revised the exclusion of "<3 months at the facility" to "<3 months on dialysis." The ESRD Conditions for Coverage at 42 CFR 494.90 require the dialysis interdisciplinary team to perform a reassessment during the 4th month of dialysis and to use results of that reassessment for the patient plan of care meeting that is held 15 days after the team completes that reassessment. Excluding patients during the first 3 months of dialysis will align the survey with the first reassessment.
- We revised the exclusion of "cognitive impairment, dementia, psychosis" to "unable to complete due to mental status" to incorporate those diagnoses as well as others that make completion impossible or unreliable.
- We removed patients who refuse to complete the survey from the exclusions from the denominator. Facilities need to track and make efforts to to increase the number of patients completing the survey.
- We broadened the target populations to include more than just seniors since dialysis patients are populations at risk, dual eligible beneficiaries, individuals with multiple chronic illnesses, veterans.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of eligible (not excluded) individuals with ESRD (ICD-10 N18.6) on dialysis who complete a KDQOL-36 with or without assistance at least once per calendar year

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Annually

S.6. **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, 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.
Number of eligible (not excluded) individuals with ESRD (ICD-10 N18.6) on peritoneal dialysis, in-center hemodialysis, and home hemodialysis who complete a KDQOL-36 survey with or without assistance during the calendar year. A patient who declines to complete one survey but completes one survey during the calendar year is counted as having a completed a survey that year. A patient who completes more than one survey in a calendar year is counted only once.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of individuals with ESRD (ICD-10 N18.6) on peritoneal dialysis, in-center hemodialysis, and home hemodialysis treated by the dialysis facility during the calendar year minus those dialysis patients who meet exclusion criteria in S.10.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Total number of individuals with ESRD (ICD-10 N18.6) on all types of dialysis at the dialysis facility minus patients who meet exclusion criteria in S.10.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with ESRD (ICD-10 N18.6) on dialysis who are <18 years old; who are unable to complete the survey due to mental status that could invalidate the results; who are non-English speaking/reading and no native language translation or interpreter is available; or who have been on dialysis for <3 months. A patient who declines to complete one survey but completes one survey during the calendar year is counted as having a completed survey.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)
1 - Age <18 calculated by date of exclusion minus date of birth in the medical record
2 - Unable to complete due to mental status (revised exclusion) from the medical record
3 - Non-English speaking/reading (no language translation or interpreter available) from medical record and RAND translations for KDQOL-36 and interpreter resources like www.LanguageLine.com or other service
4 - <3 months on dialysis (revised exclusion) calculated by date of exclusion minus date of first dialysis on Form CMS 2728 in medical record

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
NA

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
NA

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
NA

S.16. **Type of score:**
Categorical, e.g., yes/no
If other:

S.17. **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)*
Passing score defines better quality

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Number of completed surveys divided by the number of eligible dialysis patients (all treatment types) treated at the dialysis facility during the calendar year. Exclusion criteria are described in S.10.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
NA

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
At each dialysis facility, dialysis staff (usually the social worker) offers the survey to those dialysis patients who are not excluded.

Dialysis staff describe the purpose of the survey, how to complete it alerting the patient to pay close attention to the time period related to each question. Staff collect the survey, score it, and report results to the patient and dialysis team.

The survey downloaded from the RAND site states, "This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do you usual activities."

The survey downloaded from the KDQOL-Complete site states, "The KDQOL-36 survey lets you rate your quality of life with kidney disease. Hundreds of studies have found that how you view your physical and mental function is vital. People who had a poor view of their lives were more likely to need hospital care and less likely to live a long time. You are the only one who can tell us how you feel about your life. In fact, how you rate your quality of life is one of the best ways to know how you are doing. The Dialysis Outcomes and Practice Patterns Study (DOPPS) looks at people who are on dialysis around the world. The DOPPS found a strong link between how people feel, their quality of life, and how well they do on dialysis. We ask you to take this survey so you can share things that may affect how well you feel while you receive dialysis treatment. At the end of the survey, we will provide a report that will tell you information about:
* Your scores on each of 5 subtests
* How your scores compare to others like you with regard to age, sex, and diabetic status
* Things you can do to improve your scores
Over time, tracking your scores will help you learn how taking care of yourself affects how you feel.
Help us to help you feel your best with kidney failure."

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Anecdotally, dialysis facility staff report that completion rates are higher when the patient understands the survey purpose and how it will be used, when survey is completed at the dialysis facility rather than taken home to complete and return, when staff provide help as needed, when results are shared with patients in a timely manner, and when the patient is involved in goal setting to address problems related to the patient´s health-related quality of life.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Patient Reported Data/Survey

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Kidney Disease Quality of Life (KDQOL-36) survey

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Dialysis Facility
If other:

S.28. **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.)*
NA

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0260-Measure-Testing-Attachment-2016.docx

Measure Number (*if previously endorsed*)**:** 0260
Measure Title: Assessment of Health-related Quality of Life in Dialysis Patients
Date of Submission: 4/21/20164/21/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *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.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.
2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.
2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]
2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.
2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient reported data/surveyPatient reported data/survey	☒ other: Patient reported data/surveyPatient reported data/survey

Please note: Patient reported survey data is entered into KDQOL-Complete by subscribers that range from individual dialysis facilities to facilities owned or managed by small and medium sized dialysis organizations located throughout the U.S. and its territories.

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*).

This measure has been in full implementation since 2008. The data presented here were collected from KDQOL-Complete, which receives patient data from virtually all of the independent dialysis clinics in the United States via a HIPAA secure data portal, so each year data is available for reliability and validity testing on a large population. The "refused" column includes the number patients each year who declined to complete a survey. The "excluded" column consists of patients age <18, those with cognitive impairment/dementia/active psychosis, those who have been treated at a facility <3 months, or those for whom no translation or interpreter was available for their language. The table below also reports facility counts with ≥10 patients, because some facilities in KDQOL-Complete—primarily rural or

pediatric ones—have very low numbers of patients who are ≥18 years old, thus one or two refusals could skew their data significantly.

Measure	All	Completed	Refused	Excluded	Facilities with ≥ 10 pts.
2013					
Patients	79,872	62,620	11,790	5,462	-
Facilities	1,493	-	-	-	1,240
2014					
Patients	84,157	66,386	12,015	5,756	-
Facilities	1,377	-	-	-	1,222
2015					
Patients	87,892	69,746	12,475	5,671	-
Facilities	1,383	-	-	-	1,261

Felicia Speed, LMSW, Corporate Director of Social Work Services at Fresenius Kidney Care provided in the data table (below) in an email on April 21, 2016:

Measure	All	Completed	Refused	Excluded
2013	198,572	160,007	16,435	22,130
2014	187,021	149,817	20,060	17,144
2015	168,772	132,710	21,847	14,215

1.3. What are the dates of the data used in testing? 2013-20152013-2015

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Dialysis facilitiesDialysis facilities	☒ other: Dialysis facilitiesDialysis facilities

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*)

To test the measure of survey completion, we used the full data set of KDQOL-Complete subscribing dialysis facilities throughout the U.S. (See the table above in question 1.2 for the number of facilities 2013, 2014, and 2015) and its territories that treated at least 10 patients. The Fresenius Kidney Care data provided in 1.2 (above) was not used in the remaining analyses.

1.6. How many and which patients 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*)

To test the measure of survey completion, we used the full KDQOL-Complete data set, counting only the *first* KDQOL-36 survey for patients who completed the survey more than once in a calendar year. (See the table above in question 1.2 for the number of facilities 2013, 2014, and 2015)

Measure	Mean (SD) or Percent											
	2013				2014				2015			
	All N= 79,872	Com N= 62,620	Ref N= 11,790	Excl N= 5,462	All N= 84,167	Com N= 66,386	Ref N= 12,015	Excl N= 5,756	All N= 87,892	Com N= 69,746	Ref N= 12,475	Excl N= 5,671
Age (years)	62.4 (14.9)	61.9 (14.7)	61.9 (15.1)	69.2 (14.5)	62.4 (14.9)	62.0 (14.7)	61.7 (15.1)	69.2 (14.6)	62.4 (14.7)	62.0 (14.5)	61.7 (15.0)	69.5 (14.6)
Male (%)	55.5	55.7	49.9	59.1	56.3	56.2	51.2	59.1	56.4	56.4	59.2	50.4
Race: (%)												
White	45.4	46.3	43.7	38.6	45.4	46.6	42.5	37.8	45.1	45.9	43.3	38.8
Black	26.0	25.9	25.3	28.5	25.8	25.8	25.5	27.4	26.6	27.0	24.7	26.5
Asian	4.5	4.3	3.7	8.0	5.5	5.4	4.4	9.4	5.4	5.1	5.2	9.9
Native Am.	1.6	1.8	1.3	2.6	1.5	1.5	1.5	0.6	1.5	1.5	1.7	0.7
Pac Island.	1.8	1.8	2.4	1.3	1.9	1.8	1.7	2.5	1.9	1.9	1.9	2.2
Missing	20.8	20.2	23.5	21.0	19.9	18.9	24.3	22.3	19.5	18.7	23.1	21.9
Diabetes (%)	52.7	52.8	51.6	54.6	52.6	52.9	50.5	54.1	52.5	52.4	51.6	55.3

Legend: Com = Survey Completed; Ref = Patient Refused Survey; Excl = Excluded

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.

While all KDDQOL-Complete subscribing facilities were used to estimate the parameters to calculate reliability, overall reliability scores are presented as the mean of facility reliability for those with 10 or more patients to avoid unstable estimates. While we think this approach is best, we also calculated the average for *all* facilities and it did not differ from those we chose to report. For these analyses, patients who did not complete the KDQOL-36 survey because of the exclusion criteria were not considered and were removed from the analysis.

All patients were employed in the validity analyses. This includes patients who completed the KDQOL-36 survey, those who refused to complete the survey, *and* those who were removed based on exclusion criteria in the measure definition.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

The KDQOL-Complete database collects data on required fields including age, gender, diabetes status, and dialysis modality. In 2013, race/ethnicity was added as a required field, so data on new patients added to the database since 2013 include this demographic characteristic. Some additional fields are optional. Please see the table in 1.6 above.

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*)

Reliability is a function of provider-to-provider variation and samples size. Empirical testing of computed performance scores for reportable clinics was conducted using a beta-binomial model. Reliability ranges from 0.0 (no consistency) to 1.00 (perfect consistency). The extent to which the reliability falls below 1.00 is the extent to which errors of measurement are present. Reliability of 0.70 or greater is considered acceptable for drawing conclusions about groups.

- The BETABIN macro was used on each measure (SAS).
- Use the macro to get α and β.
- Provider-to-provider variance: $\sigma_2 = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)_2$
- Plug this variance value into the reliability equation: $\sigma_2 / (\sigma_2 + (p(1 – p)/n))$
    - p = rate
    - n = number of eligible patients
- Determine reliability rate for each clinic.
- Average the reliability rate over all clinics.
- Reliability = 0.926 (2013), 0.925 (2014), and 0.093 (2015)

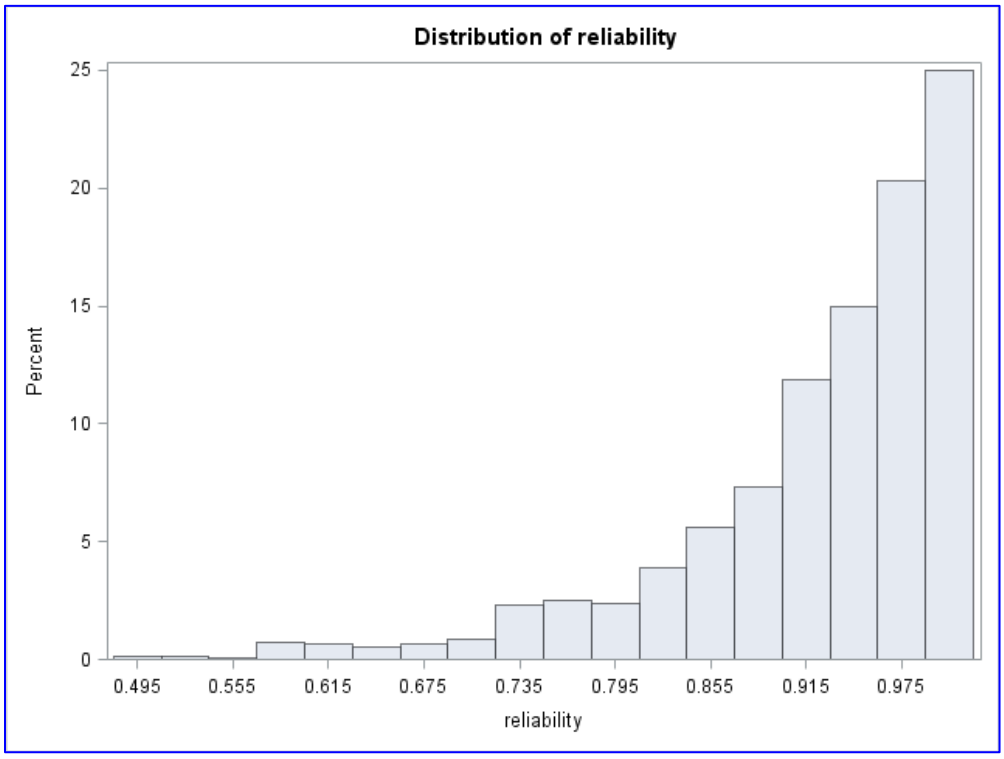
Dialysis facilities submit patient-specific demographics and KDQOL-36 survey data for their facility's census for measure calculation via KDQOL-Complete, a HIPAA secure data portal.

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 internal reliability of the measure was excellent, with Cronbach's alpha of 0.926 for 2013, 0.925 for 2014, and 0.923 for 2015 from KDQOL-Complete data.

The distribution of facility reliability scores for the most recent year of data was as follows:

2015: N=1,261 Facilities



Basic Statistical Measures			
Location		Variability	
Mean	0.922861	Std Deviation	0.08924
Median	0.951826	Variance	0.00796
Mode	1.000000	Range	0.51417
		Interquartile Range	0.09973

Quantiles(Definition 5)	
Quantile	Estimate
100% Max	1.000000
99%	1.000000
95%	1.000000
90%	1.000000
75% Q3	0.989883
50% Median	0.951826
25% Q1	0.890155
10%	0.802026
5%	0.738065
1%	0.596184
0% Min	0.485829

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?*)

In terms of understanding reliability in detecting signal to noise, a reliability score of 0.70 or greater is considered acceptable for drawing conclusions about groups. This data analysis of critical data elements, demonstrates this measure construct to be reliable and to detect meaningful differences among facilities and their patient population.

Reliability coefficient value	Interpretation
.90 and up	Excellent
.80 - .89	Good
.70 - .79	Adequate
below .70	May have limited applicability

2b2. VALIDITY TESTING

2b2.1. 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 <u>performance measure score</u> 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*)

2b2.2. 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)*

In order to assess the validity of this process measure, we assessed its association with the actual quality of life scores of patients. This was done using linear mixed models with the patient-level quality of life scores for each scale as the dependent variable and facility completion rate as the main independent variable. The models were adjusted for patient-level characteristics (age, sex, race, and diabetes). The models accounted for facility clustering using, assuming a compound symmetry covariance structure.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

We found a significant and positive association between facility completion rates and QoL scores for all five scales. Results displayed are the estimated effect of comparing a facility with a 10% higher completion rate (i.e., 90% vs. 80%). While the estimates may appear small, one should keep in mind that this estimate applies to all patients within the facility.

Associations between facility completion rates and patient-level QoL scores (per 10% higher compl)

QoL Scale	Estimate	p-value
Symptoms	0.21564	<0.0001
Effects	0.36450	<0.0001
Burden	0.19037	0.0223
MCS	0.08023	0.0060
PCS	0.05872	0.0389

2b2.4. 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?*)

Although an outcome quality measure would be preferred, we are encouraged by the results of this process measure. We found that facilities with higher completion rates were associated with statistically significantly higher patient-level QoL scores within the facility. This finding is important because it is plausible that facilities with higher rates would be obtaining completed questionnaires from sicker patients, since it has been documented that individuals completing the QoL scores tend to be younger and healthier.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Because data is entered by facility staff over a HIPAA-compliant Internet connection, it was impossible to audit medical records to confirm accurate data entry. Therefore, we conducted the following analyses to assess that the exclusion criteria were being applied correctly and the results are valid:

1. For the age exclusion, what percent of the time were individuals actually <18 (calculated as date declined minus date of birth)?
2. What are the exclusions by other available characteristics?
3. Are the exclusions needed to make the comparisons fair?
4. What are the distributions of exclusions across facilities?

2b3.2. 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*)

1. Age Exclusion: Overall 89 individuals were excluded because they were <18 years of age. Of these, we verified (by calculating age at time of survey – date of birth) that 80 of the patients were, in fact <18 years of age. By years, these values were as follows:
    - 2013 = 27/29 (93.1%)
    - 2014 = 22/24 (91.7%)
    - 2015 = 31/35 (86.1%)

2. Patient characteristics by exclusion criteria (2015 shown), other than age shown above:

Measure	Reason for Exclusion			
	Age <18 years (N=38)	**Cognitive Impairment (N=4,082)**	**Language Barrier (N=707)**	**In facility <3 months (N=844)**
Male (%)	42.1	50.3	45.7	55.3
Race: (%)				
White	44.7	41.5	20.5	40.8
Black	29.0	29.8	6.9	26.7
Asian	2.6	5.7	38.8	6.4
Native American	0	0.6	0.4	1.7
Pacific Islander	2.6	1.2	9.2	0.7
Missing	21.1	20.9	24.2	23.8
Diabetes (%)	5.3	57.4	51.6	50.1

3. In order to assess if the exclusions are needed, we assessed the association between completion vs. refusal and completion vs. excluded, which are both presented below. Associations are shown for both the patient level data (logistic regression) and at the facility level (linear regression).

Patient-Level model for the Odds of Completion vs. Refusal (excluded patients completely removed)

Parameter	OR (Completed vs. Refused)	p-value
Age (per 10 years)	1.01	**0.002**
Male (vs. Female)	0.89	**< 0.0001**
Race: (vs. White)		
Black	0.98	0.11
Asian	1.04	0.22
Missing	0.80	**<0.0001**
Year (per 1 year)	1.03	**<0.0001**

** NOTE: this shows there has been improvement in the odds of completion over the years (3% more likely to complete each year)

Facility-Level model for the association between distribution of patient characteristics at the facility and the facility level completion rate (using completed and refused patient as the denominator)

Parameter	Estimate (difference in facility completion percentage points)	p-value
Facility Mean Age (per 10 yr)	+0.0	0.95
% Male (per 10% more)	-0.45	**0.012**
% Race (per 10% more):		
White	.	Ref
Black	+0.18	0.09
Asian	+0.56	**0.02**

*Adjusted for year

Patient-Level model for the Odds of Completion vs. Excluded (patients who refused are completely removed)

Parameter	OR (Completed vs. Excluded)	p-value
Age (per 10 yrs)	0.97	***< 0.0001***
Male (vs. Female)	1.19	***< 0.0001***
Race: (vs. White)		
Black	0.70	***< 0.0001***
Asian	0.44	***< 0.0001***
Missing	0.74	***<0.0001***
Year (per 1 year)	1.03	***0.0007***

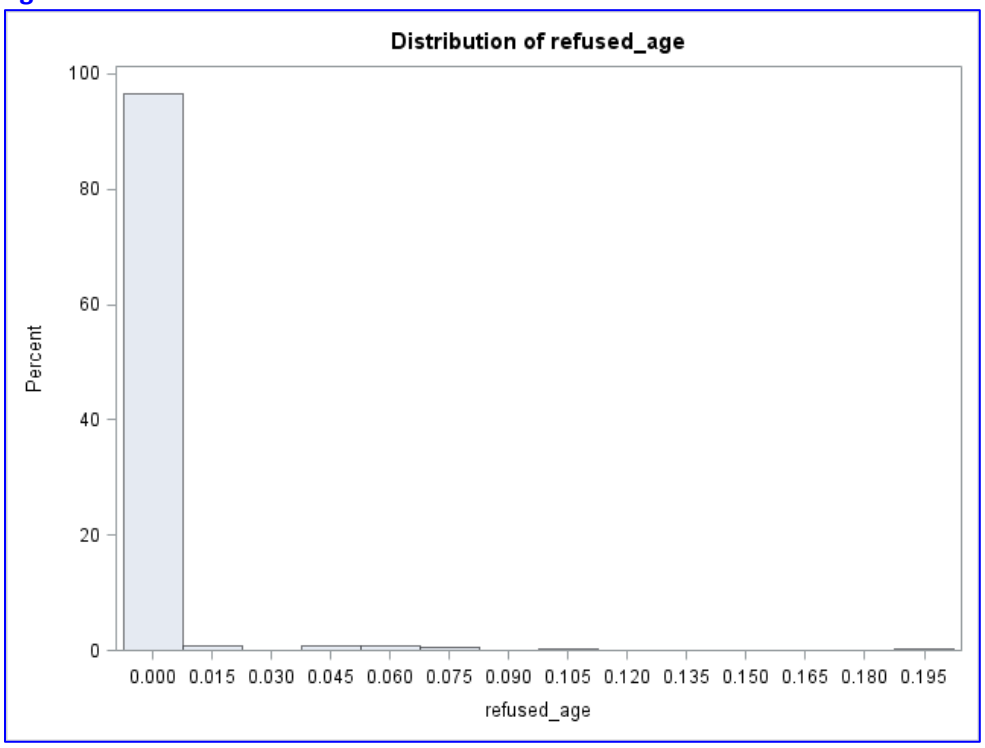
Facility-Level model for the association between distribution of patient characteristics at the facility and the facility level completion rate (using completed and excluded patient as the denominator)

Parameter	Estimate (difference in facility completion percentage points)	p-value
Facility Mean Age (per 10 yr)	-3.2	***<.0001***
% Male (per 10% more)	+0.23	***0.0448***
% Race (per 10% more):		
White	.	Ref
Black	-0.48	***<.0001***
Asian	-0.53	***0.0006***
Missing	-0.28	***<.0001***

* Adjusted for year

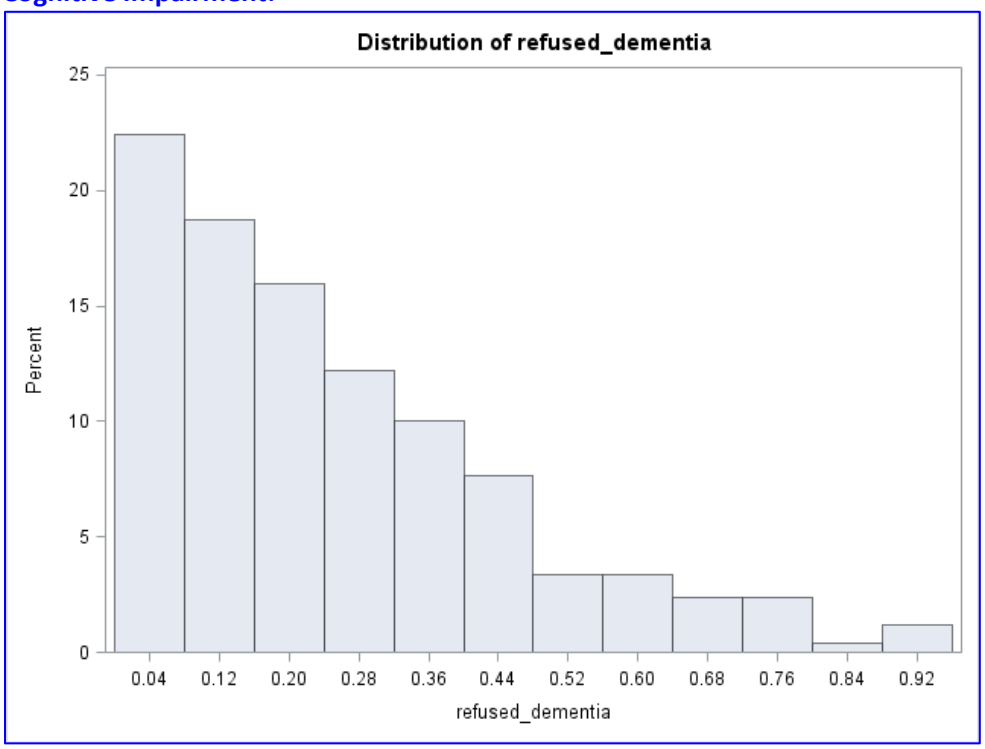
4. Distribution of exclusions across facilities (for facilities with 10 or more patients and among only individuals with an exclusion reason, N=508 facilities in 2015):

Age < 18:



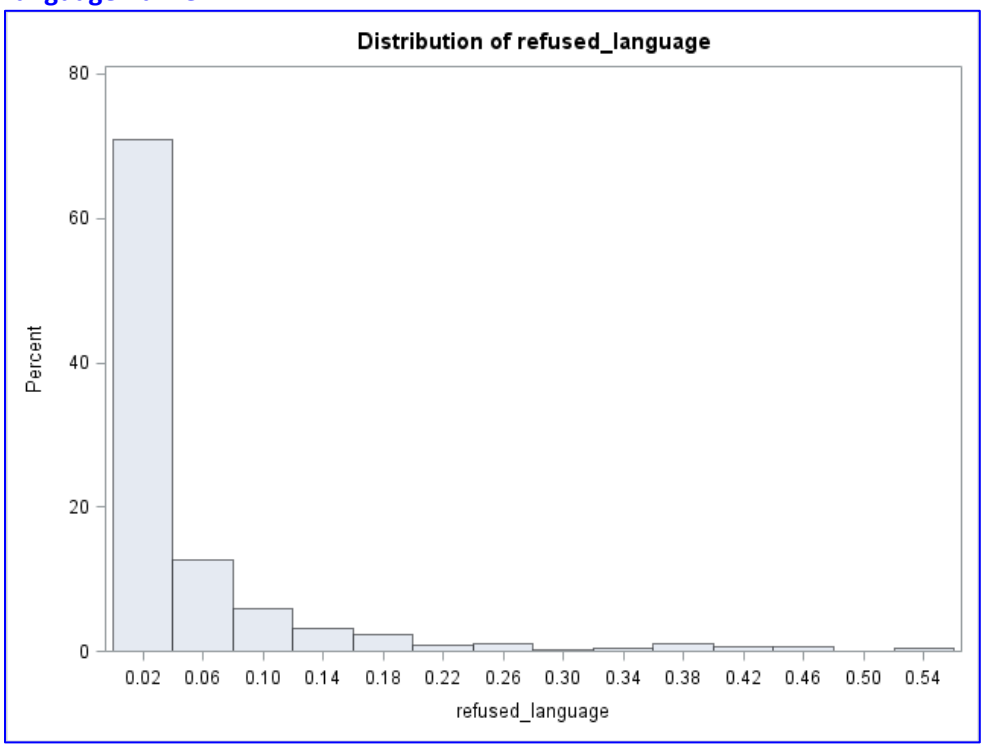
Mean = 0.2%
Median = 0%

Cognitive Impairment:



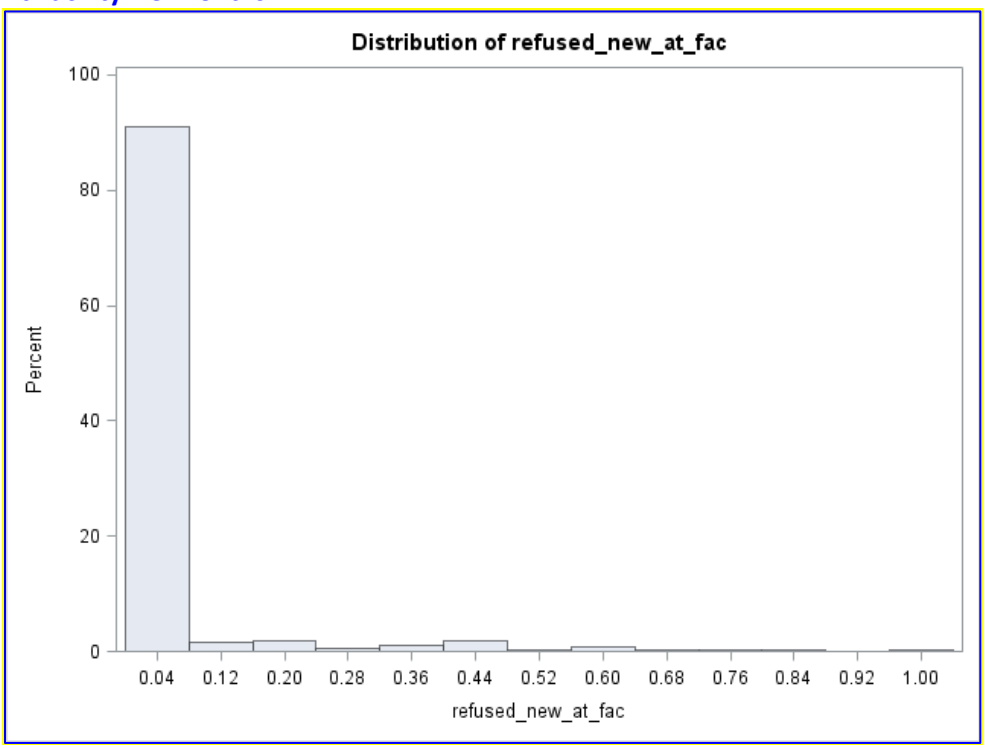
Mean = 24.8%

Language Barrier:



Mean = 4.1%
Median =

At Facility < 3 Months:



Mean = 3.7%
Median = 0%

2b3.3. 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*)

We found that individuals reported to have been excluded based on age <18 years were calculated to be <18 years approximately 90% of the time.

The findings of completion rates across facilities showed that males were significantly *less* likely to complete the KDQOL-36 survey (rather than refuse) compared to female patients, while older patients, minorities, and those whose race is unreported (missing) were significantly *more* likely to complete the KDQOL-36 survey (rather than refuse) compared to white race.

The findings of completion rates across facilities showed that males were significantly *more* likely to complete the KDQOL-36 survey (rather than be excluded) compared to female patients when compared to those who were excluded, while older patients, minorities, and those whose race is unreported (missing) were significantly *less* likely to complete the KDQOL-36 survey (rather than be excluded) when compared to white race.

So far as exclusions, in facilities with 10 or more patients, staff excluded those with cognitive impairment/dementia/active psychosis in the greatest numbers (mean 24.8%; median 20%). Staff exclusion for age <18 was very low (mean 0.2%; median 0). Staff seldom excluded those who had been at the facility <3 months (mean 3.7%; median 0). Staff exclusion for language showed the greatest variation across facilities (mean 4.1%; median 8.5%).

The rate of exclusion for age <18 is not surprising, with so few pediatric dialysis patients. Regarding cognitive impairment, a 2008 study reported that in hemodialysis patients age 55 and older, up to 70% had some degree of cognitive impairment, and in most cases this is undiagnosed [Murray AM, Cognitive impairment in chronic kidney disease, *Adv Chronic Kidney Dis* 2008;Apr 15(2):123-132]. The FHN Trial, which measured cognitive function reported that 16% of patients had impairments, while executive function was impaired in an additional 29% [Tamura MK, Larive B, Unruh ML, et al. Prevalence and correlates of cognitive impairment in hemodialysis patients: The Frequent Hemodialysis Network Trials. *Clin J Am Soc Nephrol* 2010;5(8):1429-1438].

Staff were more likely to exclude Asian patients for language. There is no KDQOL translation for all Asian languages.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

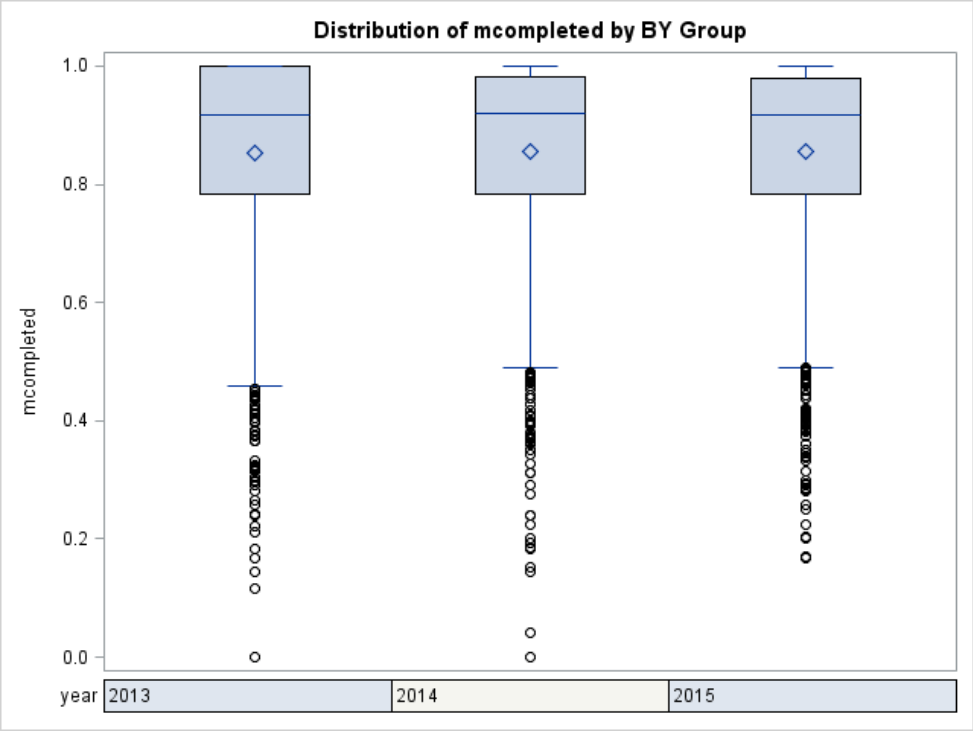
2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*
Using KDQOL-Complete data, we are able to test whether there are statistically significant differences in the performance measure between facilities. In the box plot below we present the overall distributions of scores by year for facilities with at least 10 patients. Facilities with fewer than 10 patients gave unstable estimates.

Linear regression was employed to test for significant differences across all facilities using the facility ID as a dummy variable and assessing the Type III SS for overall variation, adjusted for year. Models were run both adjusted and unadjusted for patient characteristics.

2b5.2. 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*)
F = 6.66 (p<0.0001)

Facilities with 10+ patients:



Summary Statistics:

Measure	2013	2014	2015
N	1,240	1,222	1,261
Mean	85.4	85.5	85.6
Median	91.8	92.1	91.8
Range	0-100	0-100	16.7-100
Interquartile Range	78.3-100	78.3-98.1	78.3-100
Quantiles			
100% Max	100	100	100
99%	100	100	100
95%	100	100	100
90%	100	100	100
75% Q3	100	98.1	97.9
50% Median	91.8	92.1	91.8
25% Q1	78.3	78.3	78.3
10%	59.4	61.4	61.2
5%	46.9	50.0	48.9
1%	28.1	29.1	29.3
0% Min	0	0	16.7

2b5.3. 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?*)

There is enough significant variation across the facilities to assess for meaningful differences in scores.

This measure identifies both an opportunity for improvement in survey completion among eligible patients who are in those groups of eligible patients whose response rate is lower by providing education to staff to help them increase the response rate among racial minorities who may be predisposed to distrust what they perceive to be research. It also presents an opportunity for improvement in addressing barriers to survey completion by those who are currently excluded for language if resources can be identified for survey translation into more languages, especially Asian ones.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

Every patient in KDQOL-Complete was successfully categorized into “Completed,” “Refused,” or “Excluded”. However, since we cannot know whether facilities have patients whose data is never entered into KDQOL-Complete, we cannot assess whether data are missing.

2b7.2. 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*)
Please see 2b7.1

2b7.3. 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*)
Please see 2b7.1

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)
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)*

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.**

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.**
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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
- Data collection: Although self-administration is encouraged, staff may assist patients who need it. The response rate is lower when patients require help to complete the survey and don´t get it. Patients should be encouraged to complete the survey at the clinic during dialysis or during a home clinic visit. When patients take the survey home they frequently fail to return it and when they return it, there is no guarantee the answers are the patient´s. Staff need to explain the survey´s purpose and use. Patients may decline to complete the survey if they don´t believe it is relevant to them or their care. During dialysis fluid and electrolyte shifts

contribute to changes in cognition. Patients may decline to complete the survey late if they don´t feel well or during diaysis when their brain is "foggy." Staff should offer the survey at a different time during the dialysis or on a different day rather than waiting a year.
- Availability of Data: KDQOL-Complete reports data on survey completions and exclusions by category.
- Missing data: Data completeness relies on dialysis facility staff entering data completely and accurate reporting of patients on the facility census. On the KDQOL-36, the first 12 questions yield the physical component summary (PCS) score and a mental component summary (MCS) score and must all be completed to get these scores. Otherwise, missing data is handled by averaging completed responses in each domain.
- Timing/Frequency: The minimum frequency of survey is annually. A patient can complete the survey in <30 minutes. If the social social worker assists, it may take longer and the patient may provide the social worker with a fuller picture of his/her life and HRQOL.
- Patient confidentiality: Survey instructions inform patients that the survey is used internally to help plan care. Some providers require signed patient consent while others accept completion of the survey as implied consent.

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)*.
The RAND website posts the survey instrument, instructions, and a free Excel template that staff can download to the facility computer for scoring and retention. Large dialysis organizations have developed their own scoring and reporting programs. KDQOL-Complete is a subscription-based data management and analysis service offers online scoring of the KDQOL-36 in multiple languages. Arbor Research Collaborative for Health collected KDQOL-36 data from 1,282 U.S. prevalent in-center hemodialysis (HD) patients. Arbor statisticians determined that gender (M/F), diabetes (Y/N), and age (<45, 45-64, 65-74, 75+) were the demographic characteristics associated with the greatest variability in KDQOL-36 scores. Using the case mix adjustment, KDQOL-Complete reports individual patient domain scores by tertiles (thirds): Above average (color coded green) = More than one standard deviation above the mean; Average (color coded yellow) = The mean +/- one standard deviation; Below average (color coded red) = More than one standard deviation below the mean. KDQOL-Complete also provides facility-level reports and a report with aggregate data for all facilities using KDQOL-Complete that allow dialysis facilities using KDQOL Complete to compare their population demographics, response rate, and scores to others.

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.

Planned	Current Use (for current use provide URL)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) KDQOL-Complete www.kdqol-complete.org Quality Improvement (Internal to the specific organization) KDQOL-Complete www.kdqol-complete.org Other dialysis facilities http://www.rand.org/health/surveys_tools/kdqol.html Large dialysis organizations use their own scoring programs internal to those dialysis facilities

4a.1. **For each CURRENT use, checked above, provide:**

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

For 4.1f and g above, over 6,000 U.S. dialysis facilities treat almost 500,000 dialysis patients. The Federal regulation, ESRD Conditions for Coverage at 42 CFR 494.90(a)(6), requires dialysis facilities to assess each eligible (not excluded) U.S. dialysis patient´s physical and mental functioning (HRQOL) at least annually. Further, there is a requirement at 42 CFR 494.110(a) for dialysis facilities to have a quality assessment and performance improvement program that "achieves measurable improvement in health outcomes." To do this, the facility must "measure, analyze, and track quality indicators." State ESRD surveyors use the CMS Measures Assessment Tool and the ESRD Core Survey Field Manual´s interview worksheets ask patients and staff about the survey and its use (https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/GuidanceforLawsAndRegulations/Dialysis.html) as part of all dialysis facility recertification surveys that determine compliance with the individual patient plan of care requirement (V552) and the facility-level Quality Assessment and Performance Improvement (V628).

4a.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?*)

Dialysis facilities are required to offer the health-related quality of life survey and use the results of the survey plan care for patients under 42 CFR 494.90(a)(6) in the ESRD Conditions for Coverage (CfC). State Agency ESRD surveyors review facilities requesting recertification and cite facilities that are not in compliance with the ESRD CfC.

4a.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.*)

Janis Grady, CMS project officer for CROWNWeb, has been contacted and has shared our desire for CROWNWeb to collect data on health-related quality of life with multiple CMS program directors.

The Kidney Disease Quality of Life (KDQOL) survey is administered by IMPAQ International, LLC, on behalf of the Centers for Medicare & Medicaid Services (CMS), to assess self-reported quality of life among Medicare end-stage renal disease (ESRD) beneficiaries included in the Comprehensive ESRD Care (CEC) Model. IMPAQ will use data from the two censuses to further develop and test quality of life measures suitable for monitoring changes in the impact of ESCO-provided care on the health and well-being of aligned beneficiaries over time.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Looking at all patients in each calendar year, we calculated the patient level completion rate. In 2013, 62,620 patients out of 74,410 eligible (not excluded) patients (84.2% completed the survey. In 2014 66,386 patients out of 78,401 eligible (not excluded) patients (84.7%) completed the survey. In 2015, 69,746 patients out of 82,221 eligible (not excluded) patients (84.8%) completed the survey.

If completion rates are calculated within each facility and averaged over all facilities, completion rates were 85.4% in 2013, 85.5% in 2014, and 85.6% in 2015. These completion rates do not take into account differences in patient mix across time. Therefore we calculated the odds of KDQOL-36 survey completion over time using a logistic model adjusted for age, gender, race, and diabetes status and found that patients were 3% more likely to complete the survey each year ($p<0.0001$).

4b.2. **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.**

KDQOL-Complete data showed improvement.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

None reported

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.

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

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 completely harmonized?

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:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Witten and Associates, LLC
Co.2 **Point of Contact:** Beth, Witten, beth@wittenllc.com, 913-642-0269-
Co.3 **Measure Developer if different from Measure Steward:** UCLA
Co.4 **Point of Contact:** Ron, Hays, drhays@ucla.edu, 310-794-2294-

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.
NA

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 1994
Ad.3 **Month and Year of most recent revision:** 04, 2016
Ad.4 **What is your frequency for review/update of this measure?** Ad hoc
Ad.5 **When is the next scheduled review/update for this measure?** 02, 2017

Ad.6 **Copyright statement:** The KDQOL-36 survey states: "Copyright © 2000 by RAND and the University of Arizona"
Ad.7 **Disclaimers:** NA

Ad.8 **Additional Information/Comments:** The release date for KDQOL-SF survey was 1994 and the KDQOL-36 was published in January 2004. The date the NQF endorsed the measure of percent of eligible patients who completed the HRQOL survey was 2007. 4/2016 IS most recent date for this measure maintenance.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0369
Measure Title: Standardized Mortality Ratio for Dialysis Facilities
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: Standardized mortality ratio for dialysis facility patients. This measure is calculated as a ratio but can also be expressed as a rate.
Developer Rationale: US chronic dialysis patients are much more likely to die than age-matched individuals without ESRD. The excess mortality associated with ESRD patients on dialysis is influenced by dialysis facility practices, and is one of several important health outcomes used by providers, health consumers, and insurers to evaluate the quality of care provided in dialysis facilities.

Numerator Statement: Number of deaths among eligible patients at the facility during the time period.
Denominator Statement: Number of deaths that would be expected among eligible dialysis patients at the facility during the time period, given the national average mortality rate and the patient mix at the facility.
Denominator Exclusions: N/A

Measure Type: Outcome
Data Source: Administrative claims, Electronic Clinical Data
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Apr 02, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This measure calculates the standardized mortality ratio for dialysis facility patients. This measure is calculated as a ratio but can also be expressed as a rate.
- The developer indicates that there are numerous dialysis care processes that can influence the likelihood of a patient's dying. These processes include:
    - Inadequate processes related to fluid management/removal. Inadequate control of total body fluid balance and fluid removal can result in fluid overload and congestive heart failure, increasing the possibility of death.

- Inadequate infection prevention. Inadequate infection prevention processes, including suboptimal management of vascular access, can lead to bacteremia or septicemia, increasing the possibility of death.
- Inadequate dialysis. Failure to maintain processes to ensure adequate dialysis can lead to low Kt/v, increasing the possibility of death.

Guidance from the Evidence Algorithm: Measure assesses performance of a health outcome (Box 1)→ Relationship established between measured health outcome and at least one healthcare action (Box 2)→ PASS

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence?*

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states the Standardized Mortality Ratio (SMR) for Dialysis Facilities varies widely across facilities. For the period 2010 – 2013, the 4 year SMR varied from 0.00 to 3.1.
- The mean value for 4-year SMR was 1.02 and the standard deviation was 0.28.
- The data used to calculate these rates is limited to those facilities with at least 3 expected deaths.
- The developer reports following distribution of the SMR, 2010-2014:

Year	Facilities	Mean SMR	Standard Error	10th percentile	90th percentile
2011	5004	1.02	.39	.057	1.52
2012	5155	1.02	.39	.058	1.52
2013	5279	1.02	.39	.057	1.51
2014	5409	1.02	.40	.056	1.53

Disparities

- The developer provides the following information:
  - There is evidence indicating that mortality for Hispanic ESRD patients is lower than mortality for non-Hispanic ESRD patients. Mortality for female ESRD patients is lower than mortality for male ESRD patients. This might suggest absence of a disparity with respect to ethnicity and female sex.
  - The SMR is adjusted for these patient characteristics (sex and ethnicity)as well as race to avoid masking disparities in care across groups.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Why is there no change in the distribution of results even though more facilities are now reporting?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **SMR measures multiple factors contributing to dialysis patient survival - - most of which are out of the hands of managing clinicians. Those factors that are in the hands of providers and can show improvement have not to my mind been shown

to be of adequate importance to allow SMR to be a useful measure of dialysis quality. In other words, the signal:noise may be quite low.

The last SMR review entailed the same discussion - - I don't think anything has changed - - but I'd like to once again engage this conversation.

My preliminary sense is that the evidence that this is a useful tool to measure quality of care delivery is NO PASS

**This measure calculates the standardized mortality ratio for dialysis facility patients. This measure is calculated as a ratio but can also be expressed as a rate. The developer indicates that there are numerous dialysis care processes that can influence the likelihood of a patient's dying. These processes include: Inadequate processes related to fluid management/removal. Inadequate control of total body fluid balance and fluid removal can result in fluid overload and congestive heart failure, increasing the possibility of death. Inadequate infection prevention; Inadequate infection prevention processes, including suboptimal management of vascular access, can lead to bacteremia or septicemia, increasing the possibility of death; Inadequate dialysis. Failure to maintain processes to ensure adequate dialysis can lead to low Kt/v, increasing the possibility of death.

The measure directly relates to outcomes.

Rationale is the same as previous and would agree to approve.

**There is clearly a link between care in the dialysis facility and mortality rate, so passes on Medical Evidence.

**The evidence supports there are modifiable practice patterns a dialysis unit can do to prevent mortality. However, given this is an endstage population, mortality may occur as a result of natural disease progression as well. It is important to tease these two contributing factors apart for such a measure.

**20 citations for evidence including case control studies, observational cohorts, a randomized clinical trial, and 6 review articles -- several use USRDS data -- meets evidence criteria

**Provider should be able to achieve a change in the measure results. I agree there is no need for repeat discussion

**It is possible, but not guaranteed, that the provider can improve the situation beyond the current state. I agree there is no need for repeat discussion.

**New studies, information - no.

Healthcare action related to outcome - yes

**adequate evidence base

**Evidence met.

1b. Performance Gap

Comments: **There is wide variation across facilities in SMR - - Of course by design the overall SMR has not changed over the years 2011-2014.

If SMR were "actionable" then the clear "gap" in care wold warrant a national performance measure. Again, I question the premise.

Preliminary rating opportunity for improvement: Insufficient Evidence

**Standardized Mortality Ratio (SMR) for Dialysis Facilities varies widely across facilities.

There is a gap among ethnic groups and also for sex of patient.

**I give moderate ranking , as there are significant differences between facilities, but facilities may also vary substantially by chance in any given year. Using 4 year data improves identification of poor performers.

While there are known SDS disparities in mortality, it is appropriate to adjust for SDS as the comparison is between facilities.

**1. We are assuming that the SMR provided in the performance gap analysis are using the updated claims based risk adjustment and no the prior 2728 related method. If that is true a gap exists.

2. Since 2728 data is used we are aware the literature has documented systematic differences between the accuracy of the 2728 for various demographic categories. It would be helpful to review the accuracy of the 2728 data being used vs the historical pre dialysis claims data to determine if an SES gap exists.

**Data demonstrates gap with fair range in SMR between 10% and 90% -- data as it pertains to subgroups also provided

**The opportunity for improvement would be easier to understand as a rate. There does appear to be a gap. It may not be possible to improve beyond the current state.

**There does appear to be a performance gap, especially given the ethnic and gender differences. The facilities have increased by < 10%, so I am not surprised that the data are stable.

**Gap in care - unchanging range 10-90th percentile SMR over 4 years (~0.06 to 1.5) with mean 1.02 and SD of 0.28 in 2013. Hence, opportunity to decrease variability is evident.

Disparity - contradicts commonly understood meaning: decreased mortality amongst Hispanics, Asians, and females.

**variation by gender and race

**No concerns

1c. High Priority (previously referred to as High Impact)
Comments: **yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data

Specifications:

- The following updates have been made since the last submission
  - The model now adjusts for each incident comorbidity separately rather than using a comorbidity index.
  - The indicators for diabetes were modified by consolidating the individual indicators.
  - Adjustments for 210 prevalent comorbidities (identified through Medicare claims) were included
  - The measure is now limited to Medicare patients
- The numerator of this measure is: *Number of deaths among eligible patients at the facility during the time period.*
- The denominator of this measure is: *Number of deaths that would be expected among eligible dialysis patients at the facility during the time period, given the national average mortality rate and the patient mix at the facility.*
- The ICD-9 and ICD-10 codes have been included in the Data Dictionary Code Table.
- The calculation algorithm is stated in S.18 and appears straightforward.
- This outcome measure is risk adjusted, using a statistical risk model.

Questions for the Committee :

- *Are the changes in the specification appropriate?*
- *When were the changes made? Does the data provided in 1b reflect the performance of the measure with the old specifications? If so, has analysis of the measure after the specifications were changed been done?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing-Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- To assess reliability, the developer assessed the degree to which the SMR was consistent year to year. Year to year variability in the SMR values was assessed across the years 2006, 2007, 2008 and 2009 based on the 5,280 dialysis centers for which an SMR is reported in the 2010 Dialysis Facility Reports (DFRs).
- The Committee had some concerns but agreed the data elements to identify dialysis patients and those that died are not subject to much error.

Describe any updates to testing:

- The reliability of the Standardized Mortality Ratio (SMR) was assessed using data among ESRD dialysis patients during 2010-2013.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- Since the SMR is not a simple average, the developer estimated the inter-unit reliability (IUR) using a bootstrap approach, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by an one-way analysis of variance (ANOVA).
- The developer states 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.

Results of reliability testing:

IUR for One-year SMR Overall and by Facility Size, 2010-2013

	2010		**2011**		**2012**		**2013**	
Facility Size (Number of patients)	IUR	N	IUR	N	IUR	N	IUR	N
All Facilities	0.32	5004	0.26	5155	0.30	5279	0.28	5409
Small (<=45)	0.07	1137	0.06	1205	0.03	1241	0.10	1256
Medium (46–85)	0.19	1924	0.16	1967	0.17	2018	0.17	2132
Large (>=86)	0.48	1943	0.39	1983	0.47	2020	0.42	2022

IUR for Four-year SMR Overall and by Facility Size, 2010-2013

Facility Size (Number of patients)	IUR	N
All Facilities	0.59	5935
Small (<=135)	0.30	1242
Medium (136–305)	0.45	2320
Large (>=306)	0.73	2373

- IURs for the one-year SMR ranged from 0.26-0.32 across the years 2010, 2011, 2012, and 2013, which the developer interprets as indicating that about thirty percent of the variation in the one-year SMR can be attributed to the between-facility differences (signal) and about seventy percent to within-facility variation (noise). Based on the definition proved above, this value of IUR indicates a relatively low degree of reliability. When stratified by facility size, larger facilities have greater IUR.
- The developer found the reliability improved when four-year data were used, with the IUR for the four-year SMR for 2010-2013 being 0.59 which indicates that about sixty percent of the variation in the four-year SMR can be attributed to the between-facility differences (signal) and about forty percent to within-facility variation (noise). This value of IUR indicates a moderate degree of reliability. When stratified by facility size, larger facilities have greater IUR.

Guidance from the Reliability Algorithm:

Submitted specifications precise, unambiguous and complete (Box 1)→ Empirical reliability testing conducted (Box 2)→ Testing conducted with measure score at entity level (Box 4) →Method described and appropriate (Box 5) → Level of certainty or confidence that performance measure score is reliable (Box 6): eligible for MODERATE rating for larger sample sizes.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Caveat: Reliability is moderate only for larger sample sizes and also improves when utilizing multiple years of data

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Adjusted mortality and fractions of patients achieving K/DOQI guidelines for urea reduction ratios (URRs; > or =65%) and hematocrit levels (> or =33%) were computed for 2,858 dialysis facilities from 1999 to 2002 using national data for patients with end-stage renal disease. Linear and Poisson regression were used to study the relationship between K/DOQI compliance and mortality and between changes in compliance and changes in mortality.
- Measure validity is also demonstrated by the relationship of the Standardized Mortality Ratio to other quality of care indicators, including hemoglobin greater than 10 g/dL, urea reduction ratio >= 65%, percent of patients dialyzing with a fistula, and percent of patients dialyzing with a catheter.
- The 2011 Committee expressed the following concerns: "The risk-adjustment model includes race and ethnicity, which the NQF criteria suggest not be included in risk models because it tends to obscure disparities in care. Therefore, the developer was asked for justification of including race/ethnicity in the risk model. In the case of dialysis patients, black patients have a lower death rate than whites, which is not consistent with disparities in access and quality and lower survival of black CKD patients. A committee member asked if the difference by race was primarily due to age differences in dialysis patients by race. The developer noted that race interacts with age, but that a model including age but not race/ethnicity was not sufficient. The developer reported that the identification of patients who died had been validated in prior studies. The Committee agreed that race/ethnicity should be in the model. The measure takes into account race/ethnicity as discussed regarding the risk model. However, mortality rates are not reported separately by race."

Describe any updates to validity testing: Measure validity was demonstrated by the relationship of the Standardized Mortality Ratio to other quality of care indicators. New face validity was conducted.

SUMMARY OF TESTING
Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Standardized Mortality Ratio was compared to other quality of care indicators, including the Standardized Hospitalization Ratio (SHR) – Admissions, the Standardized Readmission Ratio (SRR), the Standardized Transfusion Ratio (STrR), percent of patients dialyzing with a fistula, percent of patients dialyzing with a catheter, and percent of patients with Kt/V >=1.2. Spearman's rho is reported for all variables. Because the correlations were approximately the same for the four years 2010-2013, the developer only reported the 2013 correlations.
- Face validity was assessed by a TEP in 2006 for potential implementation on Dialysis Facility Compare (DFC). In 2015, a TEP was held specifically to consider prevalent comorbidity adjustments for inclusion in the measure. The TEP's recommendations are reflected in the risk adjustment methodology.

Validity testing results:

	Rho	P
SHR-Admissions	0.20	<.0001
SRR-Readmissions	0.10	<.0001
STrR	0.21	<.0001
AV Fistula	-0.11	<.0001
Catheter	0.13	<.0001
Hemodialysis patients with Kt/V>=1.2	-0.04	<.0001

- The developer provided the following analysis:
  - The SMR is positively correlated with the SHR-Admissions (rho=0.20, p<.0001), SRR-Readmissions (rho=0.10, p<.0001), and the STrR (rho=0.21, p<.0001); higher standardized mortality rates in facilities are associated with higher standardized hospitalization rates, higher standardized readmissions rates and higher standardized transfusion rates.
  - The SMR is negatively correlated with percent of patients in the facility with AV Fistula (rho= -0.11, p<.0001); lower standardized mortality rates are associated with higher rates of AV Fistula use.
  - The SMR is positively correlated with catheter use (rho=0.13, p<.0001), indicating that higher values of SMR are associated with increased use of catheters.
  - The SMR is negatively correlated (rho= -0.04, p<.0001) with the percent of hemodialysis patients with Kt/V>=1.2. Lower SMRs are associated with a higher percentage of patients receiving adequate dialysis dose.
  - The general consensus of the 2012 TEP was the SMR captured meaningful information on survival that DFC users could use to assess facility quality.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Are the correlations with other quality indicators as expected?*
- *The validity testing provided by the developer has been updated but has not significantly changed. Does the Committee agree there is no need for repeat discussion and vote on Validity?*

2b3-2b7. Threats to Validity

2b3. Exclusions: No exclusions.

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

- The final risk adjustment is based on a Cox or relative risk model. In this model, covariates are taken to act multiplicatively on the death rate and the adjustment model is fitted with facility defining strata so as to provide valid estimates even if the distribution of adjustment variables differs across facilities. All analyses are performed using SAS.
- In the SMR, adjustment is made for patient age, sex, race, ethnicity, cause of ESRD, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, prevalent comorbidities, and calendar year. The SMR is also adjusted for state population death rates.
- Risk adjustment factors were selected for testing based on several considerations, specifically clinical criteria, expert input, factors identified in the literature as associated with mortality, and data availability. The developer began with a large set of patient characteristics, comorbidities (at ESRD incidence and prevalent), anthropometrics, and other characteristics. Facility characteristics were also considered. Risk factors were evaluated for appropriateness of the adjustment.
- Factors considered appropriate and supported in the literature were then investigated with statistical models, including interactions between sets of adjusters, to determine if they were empirically related to mortality. Risk factors were also evaluated for face validity as potential predictors of mortality. Finally, SDS/SES factors were evaluated based on appropriateness (whether related to disparities in care), empirical association with the outcome, and support in published literature.
- Table 3a presents the SMR model coefficients. Of note, it shows the coefficients on the prevalent comorbidities that were recommended by the TEP as additional risk adjusters (i.e., in addition to the risk adjusters in the SMR model since the 2011 endorsement maintenance review).
- In addition to clinical factors, the developer evaluated patient and area-level SDS/SES factors as risk adjusters. These were in addition to the current SDS factors of race, ethnicity, and sex. Race and sex were included in the original SMR calculation and ethnicity was added to the model in 2005.
- The following relationships were observed by the developer:
    - Strong relationships among individual SDS factors, socioeconomic disadvantage and mortality in the general population.
    - Individual and market or area-level measures of deprivation have been shown to contribute independently to higher mortality.
    - Area-level income and residential segregation specifically have been shown to be associated with poorer outcomes, but particularly so for racial minorities, suggesting the interplay of patient-level (race) and area-level factors related to lower income, neighborhood poverty, segregation, levels of educational attainment, and unemployment levels that jointly influence key health outcomes in mortality and morbidity.
    - The relationship between race and mortality, as well as both race and area-level SES factors and mortality in the dialysis population, is also well documented.
    - Differences based on clinical factors and Hispanic ethnicity have also been observed to impact lower mortality.
    - Females in the general population have lower mortality rates than males.
    - Maintaining employment is a challenge for dialysis patients which in turn can influence well-being and may have a proximal impact on outcomes such as mortality.
    - Insurance status is also related to health outcomes but this has not been studied extensively within the dialysis population as it relates to mortality.
- In sum these studies suggest notable associations with mortality differences when taking into account patient level SDS factors (race, sex, ethnicity), and area level SES factors. Additionally, employment status and type of insurance coverage (specifically Medicare-Medicaid dual eligibility) suggest a proximate relationship to health outcomes that may have downstream impacts on mortality.
- Table 4a below presents a sensitivity analysis assessing the inclusion of additional SES measures (the base model already includes race, sex, and ethnicity). It compares coefficients in the original (baseline) SMR model with and without adjustment for the SES measures.
- Table 4b presents a sensitivity analysis of inclusion of additional SES measures. It compares coefficients for the prevalent comorbidities that were added into the baseline SMR model to the model with adjustment for additional SES measures.
- **Patient-level SDS:**
    - Compared with men, women were less likely to die (OR=0.92; p<0.01).

- Patients of Asian/PI, Native American and Other/unknown race, respectively, all had lower odds of mortality compared to the reference group of white patients (OR=0.72, p<0.01; OR= 0.87, p<0.01; OR=0.78, p<0.01).
- Mortality in Black patients was not significantly different from the reference group.
- Hispanic patients had lower odds of mortality (OR=0.73, p<0.01), consistent with observations in previous studies

- **Patient-level SES:**
  - Patients employed prior to ESRD incidence, and patients with unknown employment status (OR=1.13, p<0.01) had higher odds of mortality (OR=1.05; p<0.01) compared to unemployed patients.
  - Compared with Medicare-only patients, patients with both Medicare and Medicaid (OR=1.01; p=.004) and patients with Medicare as secondary/Medicare HMO (OR=1.31; p<0.01) had higher odds of mortality.
- **Area-level SES:**
  - Areas with high measures of deprivation are likely to have higher mortality
  - Overall the results provide nominal support for the postulated relationships between indicators of area-level deprivation and mortality.
- Figure 1 shows the correlation between facility SMRs with and without adjustment for patient and area-level SES.
- After adjustment for patient and area-level SES, 199 facilities (3.4%) changed performance categories. Ninety (1.5%) facilities were down-graded, and 109 (1.8%) were upgraded.
  - These analyses indicate that some patient-level SES variables affect expected death rates, while most patient and area-level SES indicators have at most minimal effect.
- SMRs with and without adjustment for patient SES and area SES are highly correlated (0.9885, p<0.0001), and adjustment for SES shifts facility performance only slightly.
  - This suggests SES does not contribute much to the flagging profiles for facility performance.
- In the final SMR model, the developer continues to include race, ethnicity, and sex (SDS factors) for risk adjustment.
  - Patient level SES factors are not included in the final risk adjusted model. Given the very small impact of area-level SES factors, the developer decided not to include these as risk adjustments in the final model. While other studies have shown the association between these patient and area-level SES factors and mortality, further work is needed to demonstrate that differences based on these factors are not related to facility care, in order to prevent disparities in care.
  - In this model, the C-Index was 0.724, which suggests good predictive ability of the risk model.
- Figure 2 is the decile plot showing estimates of cumulative rates by years. The plot shows 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 the best survival rates). The absolute differences between the groups is also large with survival at one year ranging from 96% for those patients predicted to have the lowest mortality rates (group 1) down to 60% for those predicted to have the lowest rates of survival (group 10).

Questions for the Committee:

- *Is there a conceptual relationship between the SDS factor(s) and the measure focus?*
- *Does empirical analysis (as provided by the measure developer) show that the SDS factor(s) has a significant and unique effect on the outcome in question?*
- *Does the reliability and validity testing match the final measure specifications?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *How well do the SDS variables that were available and analyzed align with the conceptual description provided?*
- *Are these variables available and generally accessible for the measured patient population?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- For both the one-year SMR and four-year SMR, a majority of facilities had mortality that was "As Expected."
    - Overall, for the 2013 SMR, approximately 3.6% of facilities had SMR that was "Better than expected," while 3.9% of all facilities had SMR that was "Worse than expected." Across all facilities, for the 2010-2013 SMR, approximately 7.7% of facilities had a SMR that was "Better than expected," while 7.6% of facilities had a SMR that was "Worse than expected."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Does this measure provide meaningful information for patients and other stakeholders about quality of care in dialysis facilities?*

2b6. Comparability of data sources/methods: Not needed.

2b7. Missing Data : An analysis of missing data was not provided on this measure.

Preliminary rating for validity: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Guidance from Validity Algorithm:

Specifications consistent with evidence (Box 1)→Potential threats to validity assessed (Box 2) →Empirical validity testing of measure as specified (Box 3) →Testing performed with measure score (Box 6) → Method described and appropriate (Box 7) →Level of certainty or confidence that measure score is a valid indicator of quality (Box 8) →Moderate

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **The Specifications are consistent with the evidence

2a2. Reliability Testing

Comments:

**SMR very consistent year to year - -as a ratio it should be.

Only LDOs have IUR > 0.41; SDOs have IUR 0.10. Thus this measure is not a very good one to reflect non-random mortality differences - - perhaps with the exception of the LDOs.

Preliminary rating of reliability: Low

**The reliability of the Standardized Mortality Ratio (SMR) was assessed using data among ESRD dialysis patients during 2010-2013. Minimal error is possible in identifying these patients.

The value of IUR indicates a moderate degree of reliability. When stratified by facility size, larger facilities have greater IUR.

**Reliability clearly better in the four year sample.

Rating moderate due to inability to eliminate all "noise", understanding that may not be possible.

**1. Reliability testing was done by unit size. However the denominator is a national average. Given that dialysis mortality may vary from geography to geography based on patient demogratphics it would advisable to test the reliability based on some geographic parameter rather than just dialysis unit size. For example the southeast may contain more African American patients which the 2011 committee described as having lower mortality.

**reliability testing better for 4 yr data than 1 yr data and for large units than smaller units -- given reliability testing may want time frame to be longer period (4 yrs) -- unclear as to ramifications of reliability and unit size on use of measure in smaller facilities since results of reliability testing rather lackluster for this sized facility

**There does appear to be some heterogeneity that could allow for improvement

**Reliability testing is adequate. Specifications are consistent with the evidence.

**IUR suggests that differences are largely accounted for by noise and the correlation between SMRs within facilities one year to the next is not highly correlated. Ability of SMR to detect true differences in facility related practices/processes that influence this outcome was greater in larger units and in calculating SMR over a 4 year period.

**Reliable

**No concerns

2b2. Validity Testing

Comments:
**SMR correlated with SHR, SR(readmission)R, ST(transfusion)R, AVF and catheter frequency - - I remain skeptical that this demonstrates validity of the testing but rather may sow that healthier people have less complications and longer lives.
Preliminary rating for validity: Low
**Race/ethnicity should be in the model. The measure takes into account race/ethnicity as discussed regarding the risk model. However, mortality rates are not reported separately by race.
SMR captured meaningful information on survival
**Validity testing shows strong correlations in the expected directions with SMR, STrR, AVF, catheter, Kt/V.
Rating: High
**Yes. However the 2015 TEP narrowly addressed the question as to what comborbidities to include and was not shown the model or allowed to have other discussions around the assumptions. This may weaken the statement that the 2015 TEP gave face validity to this measure.
**TEP face validity claimed
Also provided statistically significant relationship linking performance with SMR to other indicators of quailty
**No need for repeat discussion.
**Other standardized ratios (hospitalization, readmission, transfusion) purporting to demonstrate facility quality, AVF percentage, catheter percentage, Kt/V at or above the target correlate with rho 0.13 to p.21 except for Kt/V at -0.04, all significant. This is mildly supportive. Face validity was substantiated by TEP panels.
**Validity established
**No concerns

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**SES clearly does impact SMR - - -
**studies suggest notable associations with mortality differences when taking into account patient level SDS factors (race, sex, ethnicity), and area level SES factors. Additionally, employment status and type of insurance coverage (specifically Medicare-Medicaid dual eligibility) suggest a proximate relationship to health outcomes that may have downstream impacts on mortality.
**Extensive discussion of risk adjustment and the inclusion vs exclusion of SDS in the modeling. Unsure of their bottom line conclusion, but the final model should give the most valid comparison between facilities.
Meaningful differences - do not see criteria for "Worse than expected" and "Better than expected". I recal;l the definition being related to above or below some number of s.d. - this should be clarified.
**1) In reviewing the discussion conducted by the measure developer and TEP, the developer has purposely excluded comorbidities that may the result of the natural progression of disease or the result of dialysis unit medical management. As a result and the ability to proportionately account for dialysis unit action and natural disease this measure has a threat to validity.
2. the use of 2728 data has been described in the literature has lacking data integrity. Despite this the 2728 is still being used.
3. Since coding changes affect the underlying data, it is possible that the unless this measure is continuously improved, the accuracy of the model will vary in future years
**The adjustment for SDS appears to be warranted as this appears to impact many paramaters. However, race itself does not appear to be an issue. It is not clear why some patient-level SES factors which are significant (e.g. insurance) are not included.
**There is a conceptual relationship with SDS. SDS has been shown via empirical evidence.
**No exclusions (Medicare population only). 2b4 Comorbidity adjustment occur at incidence and prevalent over the prior year. Race, ethnicity, sex have a conceptual basis to link to this outcome. SES factors (employment, insurance status, area level factors) link to SMR when tested. Some of the impact is counterintuitive - employment and Medicare as secondary insurance increased SMR. Area level SES adjustment had a modest effect on SMR in the expected direction TEP panel judged whether the comorbidities selected would have impact on mortality and statistical tests were then carried out.
**No

**No concerns

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the 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.

- All data elements are in defined fields in electronic form and generated or collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: None

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- This measure is publically reported nationally in Dialysis Facility Compare (DFC).

Improvement results: The developer states mortality rates have decreased over time as evidenced by the coefficients for calendar year from the SMR model. The mortality rate for 2011 was 2.6% lower compared to 2010 (p-value<0.0001), and the rates for 2012 and 2013 were lower compared to 2010 at 12.4% and 13.0%, respectively (p-value <0.0001).

Year	Coefficient	P-value
2011	-0.026	<0.0001
2012	-0.124	<0.0001
2013	-0.130	<0.0001

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms: In the past, a concern has been raised about patient selection relating to ensuring access to care for

sicker patients. The SMR measure incorporates a risk adjustment methodology that levels the playing field for facilities with different patient case-mixes in order to dis-incentivize cherry-picking of healthier patients over sicker patients at higher risk of mortality. Given the extensive list of adjustments for patients' prevalent comorbidities, which reflect a substantial modification to the SMR, the developer thinks this additional risk adjustment strategy would discourage avoidance of treating sicker and more complex patients that may also have limited life expectancy, and that may be more likely to die while receiving care at their facility.

Feedback : No feedback provided on QPS. During the 2012-2013 MAP review, MAP supported this measure for inclusion in the End-Stage Renal Disease Quality Reporting. They stated the mortality is an important outcome for patients; however, the measure should be linked to structural and process measures.

***Questions for the Committee**:*

- *How do the new specifications and adjustment affect the measure results? Are the measure results comparable from year to year?*
- *Is the SMR useful and understandable for patients and other stakeholders? Would rates be more useful?*
- *If most facilities' results are "as expected" how useful is this measure for patients and other stakeholders?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**New specifications probably will not have a substantial effect on this measure.
Regarding patients' understanding and some clinicians' understanding - - rates would be clearer than the ratio - - however, I favor use of the ratio.
relatively small number of facility show "higher than expected mortality." Despite the small number - - I would say this is an important metric - - IF we are convenced that this high SMR by facility can be improved by specific process improvements. My own choice is to have the PROCESS improvements (eg., no catheter use, infection rates, adequate fluid management) metrics and not SMR
Preliminary usability rating: Moderate
**This measure is publically reported nationally in Dialysis Facility Compare (DFC).
**In use for public reporting
Apparently the measurement results are improving supporting usefulness
Unintended consequence (cherry picking patients) has been addressed by the risk adjustment modeling
Rating: High
**The data is being reported as a ratio, but could and should be reported as a rate.
The data used for statistical adjustment is not readily available to the dialysis units, making continuous improvement cycles long as data much be generated by CMS and provided to the dialysis unit. It would be helpful to provide this data on a monthly basis to the dialysis units using the six month lagged claim file.
**part of Dialysis Facility Compare
**There is concern about avoiding treatment of sicker patients, though this is at least partially addressed through risk adjustment.
**This data is usable. The as expected data IS useful, more useful than unadjusted numbers as population factors play an important role, as do comorbid disease factors.
**Dialysis Facility Compare; Star system,
**Publicly reported

**Currently in use.

Criterion 5: Related and Competing Measures

Related or competing measures

- 1463 : Standardized Hospitalization Ratio for Dialysis Facilities
- 2496 : Standardized Readmission Ratio (SRR) for dialysis facilities

Harmonization

- The developer states that the specifications are not completely harmonized. Each measure assesses different outcomes as reflected in certain differences across the measure specifications. SMR, and SHR and SRR are harmonized to the population they measure (Medicare-covered ESRD patients), methods (SMR and SHR) and certain risk adjustment factors specific to the ESRD population. SMR and SHR adjust for the same comorbidity risk factors, a similar set of patient characteristics, and use fixed effects in their modeling approach. The differences between SMR and SHR and SRR reflect adjustment for factors specific to the outcome of each respective measure. Both SMR and SHR adjust for a set of prevalent comorbidities (observed in a prior year), however the complete set of comorbidities for SMR differs from SRR. SRR, a measure of hospital utilization adjusts for planned readmissions; and for discharging hospital, acknowledging that for readmission, hospitals also bear accountability for properly coordinating care with the dialysis facility. These risk adjustments in SRR account for those characteristics specifically associated with readmission, and do not apply to SMR. Only SMR adjusts for state death rates, race, and ethnicity to account for these respective differences related to mortality outcomes and that are deemed outside of a facility's control.

Pre-meeting public and member comments

Comment by Daniel E. Weiner, MD, MS
Organization Dialysis Clinic, Inc. (DCI)
Comment #5670: I appreciate the opportunity to comment on NQF 0369 and NQF 1463, the Risk-Adjusted SMR and SHR. These are important outcome measures and the use of risk adjustment for comorbid conditions based on claims data is an important advance. The adjustment methodology has important validity issues.

Model selection needs to incorporate background knowledge about the relationship of a variable to the outcome of interest. Unfortunately, adjustment for prevalent comorbid conditions proposed in these metrics relied almost entirely on automatic variable selection techniques, resulting in a model that may be robust only for the data on which it was generated and that will rapidly lose validity as coders and codes change. In defending the methodology, the developer stated that the TEP agreed with the inclusion of this set of prevalent comorbidities. In discussing with TEP members, this is misleading, with members noting the same concerns as raised in this letter about the final models.

Examples include:

1. Cancer is good. The constellation of prostate, thyroid, and kidney cancer together has twice the protective effect against death that gangrene has for harm. This of course is ridiculous; however, the coefficients generated for these comorbid coefficients reflect multicollinearity among variables; coding habits; survival, indication and lead time biases; and, critically, lack of incorporation of existing knowledge into the predictive modeling approach.

2. Peripheral vascular disease codes for important conditions like gangrene, ulcers and osteomyelitis are messy, with numerous positive and negative coefficients that are likely to deviate from the truth with each passing year as coding habits change, providing a classic example of the pitfalls of multicollinearity in predictive models.

3. Codes for diabetic eye disease are highly protective. Why? Likely because these codes indicate that a

dialysis patient has seen an ophthalmologist, which is likely an indicator of care coordination. Inclusion of these 3 variables will harm ESCOs for example, where an eye exam is a process measure. This makes no physiologic sense.

The examples above illustrate where, although statistically correct at the time of model development, the adjustment process is destined to lose robustness rapidly with time.

In evaluating these proposed measures, I hope NQF calls attention to the details of the adjustment model and suggests a refined approach moving forward that incorporates both the advanced statistical techniques that were used in the proposed model along with existing knowledge on the relationships of clinical conditions with outcomes and awareness of the biases inherent in the use of these administrative data to develop future comorbidity adjustment models that will remain robust for their intended purpose.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5679: KCP believes mortality is an important outcome to measure, but has concerns about the specifications, reliability, validity (risk model), and harmonization issues.

SPECIFICATIONS. The specifications for the time period state "at least one year." KCP believes specifications should be unambiguous, so this construction is imprecise. We believe the time period should be an exact period, and we further believe the 1-year period is inappropriate based on the reliability testing data and, at minimum, should be a 4-year period.

As we discuss further in the following section, KCP has significant concerns about the SMR's reliability for small- and medium-sized facilities. The SMR specifications do not address a minimum sample size by excluding facilities of "x" or fewer patients, as we are aware other measures do.

The specifications do not exclude incident hospice patients. The NQF's Measure Applications Partnership (MAP) recently did not recommend the SMR, in part because the measure did not exclude patients who are already in hospice when they initiate dialysis. During the deliberations, it was noted that occasionally incident patients begin dialysis treatments while in hospice, but then choose to discontinue them after a period of time. KCP supports MAP's recommendation that patients who initiate dialysis while also in hospice be excluded from the SMR. As currently constructed, such patients are attributed to the facility providing the dialysis.

The SMR documentation indicates at least three expected deaths must occur for inclusion in the SMR calculations, but no justification or empirical analyses are offered to justify this threshold—e.g., how many clinics were excluded using this approach and what is the impact on scoring because of the exclusion?

Finally, the SMR specifications indicate the measures can be expressed as a rate, but is calculated as a ratio. KCP prefers normalized rates or year-over-year improvement in rates instead of a standardized ratio. We believe comprehension, transparency, and utility to all stakeholders is superior with a scientifically valid rate methodology. We note that MAP also did not support the SMR because, in addition to the lack of a hospice exclusion, MAP felt "mortality rates would be more meaningful to consumers and actionable for facilities."

(cont.)

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5680: RELIABILITY. Based on the testing results, KCP has serious concerns about the SMR's reliability. We note a reliability statistic of 0.70 is often considered as "good" reliability,[1] though we

recognize the characterization also depends on the analytic method. Testing results for the 1-year SMR yielded IURs of 0.26-0.32 for each of 2010, 2011, 2012, and 2013—a low degree of reliability, where only about 30% of the variation in a score can be attributed to between-facility differences, yet the specifications permit this 1-year measure. The 4-year SMR yielded an IUR of 0.66 for 2009-2012 and only 0.59 for 2010-2013 data. Even with the 4-year SMR, less than 60% of a facility's score is attributable to between-facility differences for the overall sample. Moreover, 4-year SMR testing results specifically for small- and medium-sized facilities indicate very poor reliability, with IURs of 0.30 and 0.45, respectively. Only large facilities have a reasonable IUR of 0.73 for 2010-2013 data. As noted earlier, KCP thus believes the specifications must specifically require a minimum sample as identified through the developer's empirical testing.

VALIDITY. KCP has strongly advocated for the use of prevalent co-morbidities in the SMR's risk model, and commends the developer for moving to incorporate prevalent co-morbidities in the specifications. We continue to be concerned about the validity of the Medical Evidence Form (CMS 2728) as a data source for incident co-morbidities, however, and urge that the Committee recommend that CMS assess this matter.

In previous comments to CMS, KCP noted that many of the prevalent co-morbidities in the final model had p-values significantly greater than 0.05—e.g., paralytic ileus (p=0.5007), episodic mood disorder NOS (p=0.8254). CMS responded that these were included because: "Most of the coefficient estimates for the prevalent co-morbidities are positive and statistically significant, but several do not obtain statistical significance. The very large number of clinical factors in the model expectedly generates multi-collinearity among co-variates, likely resulting in some unexpected results in direction of coefficient sign and levels of statistical significance. Inclusion of this set of prevalent co-morbidities reflects the consensus of the TEP that adjustment for all of these prevalent co-morbidities, in addition to incident co-morbidities, is important to reflect the initial and current health condition of the patient in risk adjustment."

[1]Adams, JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California:RAND Corporation. TR-653-NCQA, 2009.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5681: VALIDITY (cont.). We do not believe this approach is sufficient. Our conversations with TEP members for the SMR/SHR indicate they did not advocate for model building in a vacuum without accounting for the meaning of the coded co-morbid conditions, but rather for including as many co-morbid conditions as possible. This is a very different interpretation than is offered by the developer's explanation and far more appropriate when dealing with administrative coding habits that are not static over time. It may require, for example, grouping certain individual codes together to develop a more appropriate overarching description of true co-morbidity burden.

KCP is concerned that the strategy adopted for the SMR (and SHR) results in a model that will not be generalizable. Currently, for example, having thyroid cancer is protective to the same magnitude that atrial fibrillation is harmful. This makes no sense, and we posit is a function of collinearity and coding idiosyncrasy. Similarly, in the current model, osteomyelitis NOS-ankle is associated with a lower risk of death while ulcer of lower limb NOS is harmful. In actual medical practice, osteomyelitis is far worse than an ulcer of the lower limb. In the current model, lower extremity amputation is protective while 'status amput below knee' is harmful. Again, KCP supports prevalent co-morbidity adjustment, but we are concerned that the proposed collection of adjusters will be less robust with each year that passes from initial model development.

KCP also notes that while the SMR applies to all patients, the current list of co-morbidities does not account for those that may be unique to pediatrics. We recommend the Standing Committee suggest to the developer

that such should be considered and included when indicated.

KCP also notes that the validity testing yielded a c-statistic for the SMR of 0.724. We are concerned the model will not adequately discriminate performance—particularly that smaller units, including pediatric units, might look worse than reality. We believe a minimum c-statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

(cont.)

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5682: VALIDITY (cont.). Information on the risk model states that determination of a prevalent co-morbidity requires at least two outpatient claims or one inpatient claim, but no justification or empirical analyses are offered to support this algorithm over other approaches. We are aware this approach has been validated for diabetes,[2] but we are not that it has been validated for the large number of other co-morbidities or is broadly generalizable.

Finally, the risk model includes ambiguous language. The submission indicates patient characteristics included in the stage 1 model include "nursing home status in previous year." It is unclear if this means patients moving into a nursing home for the first time during the measurement year would not be adjusted for "nursing home status." Specifically, it is unclear as to whether the look-back is one year prior to the given event (inclusive of the data year) or if this verbiage means the look-back is in the previous calendar year (not inclusive of the data year). KCP believes such ambiguity should be addressed and that the current reporting year be included, not just the previous one.

HARMONIZATION ISSUES. The risk models for the groupings used for patient age and duration of ESRD differ among the SMR, SHR, and STrR. For example, the age groups for the SMR is n=3, but for the SHR and STrR the age groupings are the same, but n=6. Similarly, the number of groups for ESRD duration for the SMR (n=4) differs from that for the SHR (n=6). No justification or empirical analyses are offered to justify these differences.

There also are significant inconsistencies in how facility size is defined when assessing reliability for the SMR, SHR, and STrR. Specifically, for the SMR, the definitions were <=45, 46-85, >=86 for the 1-year reliability analyses, but were <=135, 136-305, and >=306 for the 4-year analyses. For the SHR, <=50, 51-87, and >=88 were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as <=46, 47-78, and >=79, respectively. We understand reliability for a given measure depends on sample size, but find the varying demarcations analytically troubling. We posit a more appropriate analytic approach would be to analyze reliability using consistent "bins" of size (i.e., small, medium, and large are consistently defined) and identify the facility size at which reliability for that particular measure can be confidently inferred—and then reflect the minimum size in the actual specifications.

[2]Hebert PL, Geiss LS, et al. Identifying persons with diabetes using Medicare claims data. Am J Med Qual. 1999;14(4):270-277.

Comment by Joseph Vassalotti
Organization National Kidney Foundation
Comment #5696: Per our comments to the Measures Application Partnership (MAP), the National Kidney Foundation (NKF) does not support this measure as it does not clearly encourage quality improvement nor

provide meaningful information to patients. This measure does not stratify by causes of mortality that are attributable to the care that a patient receives by the dialysis care team and does not adequately adjust for comorbidities. For example, clinics caring for patients with high levels of comorbidity, poor functional status and frailty will be penalized. This may create disincentives to accept patients with complex illness. In addition and as the MAP recommended, patients who begin dialysis while in hospice should be excluded from the measure. Some patients may elect to begin dialysis while under hospice before later deciding to discontinue. This is a difficult decision for patients and families and should not be unintentionally discouraged by including these patients in the SMR. Lastly, patients acknowledge concern about dialysis units that have high SMRs, but do not have enough information to interpret SMR as it reflects quality dialysis care delivery.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0369
Measure Title: Standardized Mortality 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: 4/15/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
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

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

2011 Submission

The Standardized Mortality Ratio (SMR) is used by ESRD state surveyors in conjunction with other standard criteria for prioritizing and selecting facilities to survey. This patient survival classification measure is reported publicly on the DFC web site to assist patients in selecting dialysis facilities.

2016 Submission

There are numerous dialysis care processes that can influence the likelihood of a patient's dying. These processes include:

(1) Inadequate processes related to fluid management/removal. Inadequate control of total body fluid balance and fluid removal can result in fluid overload and congestive heart failure, increasing the possibility of death.

(2) Inadequate infection prevention. Inadequate infection prevention processes, including suboptimal management of vascular access, can lead to bacteremia or septicemia, increasing the possibility of death.

(3) Inadequate dialysis. Failure to maintain processes to ensure adequate dialysis can lead to low Kt/v, increasing the possibility of death.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

2011 Submission

This was not a question on the 2011 Submission Form.

2016 Submission

ESRD patients on chronic dialysis experience all cause mortality far in excess of age matched controls [1]. Patients in some dialysis facilities have consistently higher mortality than patients in other facilities, even after controlling for multiple patient characteristics [2]. Selection of dialysis modality, sometimes the result of dialysis facility practices, likely influences mortality [3]. Furthermore, mortality from certain conditions resulting from kidney failure and chronic dialysis care, including uremic toxin accumulation, volume overload/HTN and its treatment, bone/mineral disease, and infections related to dialysis access, have been described in detail [4-6].

Specific dialysis practices have been identified for several of these ESRD-related conditions that can improve patient survival and morbidity, including provision of adequate small solute clearance [7], control of total body volume while guarding against rapid ultrafiltration [8-11] and appropriate management of mineral and bone disorders [12-14]. In addition, improved infection prevention efforts by dialysis providers can result in reduced infection-related hospitalization and mortality [15-20].

[1]. United States Renal Data System. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015.

[2]. Kalbfleisch J, Wolfe R, Bell S, Sun R, Messana J, Shearon T, Ashby V, Padilla R, Zhang M, Turenne M, Pearson J, Dahlerus C, Li Y. Risk Adjustment and the Assessment of Disparities in Dialysis Mortality Outcomes. J Am Soc Nephrol. 2015; Nov;26(11):2641-5.

Abstract: Standardized mortality ratios (SMRs) reported by Medicare compare mortality at individual dialysis facilities with the national average, and are currently adjusted for race. However, whether the adjustment for race obscures or clarifies disparities in quality of care for minority groups is unknown. Cox model-based SMRs were computed with and without adjustment for patient race for 5920 facilities in the United States during 2010. The study population included virtually all patients treated with dialysis during this period. Without race adjustment, facilities with higher proportions of black patients had better survival outcomes; facilities with the highest percentage of black patients (top 10%) had overall mortality rates approximately 7% lower than expected. After adjusting for within-facility racial differences, facilities with higher proportions of black patients had poorer survival outcomes among black and non-black patients; facilities with the highest percentage of black patients (top 10%) had mortality rates approximately 6% worse than expected. In conclusion, accounting for within-facility racial differences in the computation of SMR helps to clarify disparities in quality of health care among patients with ESRD. The adjustment that accommodates within-facility comparisons is key, because it could also clarify relationships between patient characteristics and health care provider outcomes in other settings.

[3]. Weinhandl ED, Nieman KM, Gilbertson DT, Collins AJ. Hospitalization in daily home hemodialysis and matched thrice-weekly in-center hemodialysis patients. Am J Kidney Dis. 2015 Jan;65(1):98-108.

BACKGROUND: Cardiovascular disease is a common cause of hospitalization in dialysis patients. Daily hemodialysis improves some parameters of cardiovascular function, but whether it associates with lower hospitalization risk is unclear.

STUDY DESIGN: Observational cohort study using US Renal Data System data.

SETTING & PARTICIPANTS: Medicare-enrolled daily (5 or 6 sessions weekly) home hemodialysis (HHD) patients initiating NxStage System One use from January 1, 2006, through December 31, 2009, and contemporary thrice-weekly in-center hemodialysis patients, matched 5 to 1.

PREDICTOR: Daily HHD or thrice-weekly in-center hemodialysis.

OUTCOMES & MEASUREMENTS: All-cause and cause-specific hospital admissions, hospital readmissions, and hospital days assessed from Medicare Part A claims.

RESULTS: For 3,480 daily HHD and 17,400 thrice-weekly in-center hemodialysis patients in intention-to-treat analysis, the HR of all-cause admission for daily HHD versus in-center hemodialysis was 1.01 (95%CI, 0.98-1.03). Cause-specific admission HRs were 0.89 (95%CI, 0.86-0.93) for cardiovascular disease, 1.18 (95%CI, 1.13-1.23) for infection, 1.01 (95%CI, 0.93-1.09) for vascular access dysfunction, and 1.02 (95%CI, 0.99-1.06) for other morbidity. Regarding cardiovascular disease, first admission and readmission HRs for daily HHD versus in-center hemodialysis were 0.91 and 0.87, respectively. Regarding infection, first admission and readmission HRs were 1.35 and 1.03, respectively. Protective associations of daily HHD with heart failure and hypertensive disease were most pronounced, as were adverse associations of daily HHD with bacteremia/sepsis, cardiac infection, osteomyelitis, and vascular access infection.

LIMITATIONS: Results may be confounded by unmeasured factors, including vascular access type; information about dialysis frequency, duration, and dose was lacking; causes of admission may be misclassified; results may not apply to patients without Medicare coverage.

CONCLUSIONS: All-cause hospitalization risk was similar in daily HHD and thrice-weekly in-center hemodialysis patients. However, risk of cardiovascular-related admission was lower with daily HHD, and risk of infection-related admission was higher. More attention should be afforded to infection in HHD patients.

[4]. Himmelfarb J, Ikizler T. Hemodialysis N Engl J. 2010 Nov; 363:1833–1845.

Abstract: Fifty years ago, Belding Scribner and his colleagues at the University of Washington developed a blood-access device using Teflon-coated plastic tubes, which facilitated the use of repeated hemodialysis as a life-sustaining treatment for patients with uremia.1,2 The introduction of the Scribner shunt, as it became known, soon led to the development of a variety of surgical techniques for the creation of arteriovenous fistulas and grafts. Consequently, hemodialysis has made survival possible for more than a million people throughout the world who have end-stage renal disease (ESRD) with limited or no kidney function. The expansion of dialysis into a form of long-term renal-replacement therapy transformed the field of nephrology and also created a new area of medical science, which has been called the physiology of the artificial kidney. This review describes the medical, social, and economic evolution of hemodialysis therapy.

[5]. Kliger AS. Maintaining Safety in the Dialysis Facility. Clin J Am Soc Nephrol. 2015 Apr 7;10(4):688-95.

Abstract: Errors in dialysis care can cause harm and death. While dialysis machines are rarely a major cause of morbidity, human factors at the machine interface and suboptimal communication among caregivers are common sources of error. Major causes of potentially reversible adverse outcomes include medication errors, infections, hyperkalemia, access-related errors, and patient falls. Root cause analysis of adverse events and "near misses" can illuminate care processes and show system changes to improve safety. Human factors engineering and simulation exercises have strong potential to define common clinical team purpose, and improve processes of care. Patient observations and their participation in error reduction increase the effectiveness of patient safety efforts.

[6]. Hung AM, Hakim RM. Dialysate and Serum Potassium in Hemodialysis. Am J Kidney Dis. 2015 Jul;66(1):125-32.

Abstract: Most patients with end-stage renal disease depend on intermittent hemodialysis to maintain levels of serum potassium and other electrolytes within a normal range. However, one of the challenges has been the safety of using a low-potassium dialysate to achieve that goal, given the concern about the effects that rapid and/or large changes in serum potassium concentrations may have on cardiac electrophysiology and arrhythmia. Additionally, in this patient population, there is a high prevalence of structural cardiac changes and ischemic heart disease, making them even more susceptible to acute arrhythmogenic triggers. This concern is highlighted by the knowledge that about two-thirds of all cardiac deaths in dialysis are due to sudden cardiac death and that sudden cardiac death accounts for 25% of the overall death for end-stage renal disease. Developing new approaches and practice standards for potassium removal during dialysis, as well as understanding other modifiable triggers of sudden cardiac death, such as other electrolyte components of the dialysate (magnesium and calcium), rapid ultrafiltration rates, and safety of a number of medications (ie, drugs that prolong the QT interval or use of digoxin), are critical in order to decrease the unacceptably high cardiac mortality experienced by hemodialysis-dependent patients.

[7]. Port FK, Ashby VB, Dhingra RK, Roys EC, Wolfe RA: Dialysis dose and body mass index are strongly associated with survival in hemodialysis patients. J Am Soc Nephrol 13:1061-1066, 2002

Abstract: Low dose of hemodialysis (HD) and small body size are independent risk factors for mortality. Recent changes in clinical practice, toward higher HD doses and use of more high-flux dialyzers, suggest the need to redetermine the dose level above which no benefit from higher dose can be observed. Data were analyzed from

45,967 HD patients starting end-stage renal disease (ESRD) therapy during April 1, 1997, through December 31, 1998. Data from Health Care Financing Administration (HCFA) billing records during months 10 to 15 of ESRD were used to classify each patient into one of five categories of HD dose by urea reduction ratio (URR) ranging from <60% to >75%. Cox regression models were used to calculate relative risk (RR) of mortality after adjustment for demographics, body mass index (BMI), and 18 comorbid conditions. Of the three body-size groups, the lowest BMI group had a 42% higher mortality risk than the highest BMI tertile. In each of three body-size groups by BMI, the RR was 17%, 17%, and 19% lower per 5% higher URR category among groups with small, medium, and large BMI, respectively ($P < 0.0001$ for each group). Patients treated with URR >75% had a substantially lower RR than patients treated with URR 70 to 75% ($P < 0.005$ each, for medium and small BMI groups). It is concluded that a higher dialysis dose, substantially above the Dialysis Outcomes Quality Initiative guidelines (URR >65%), is a strong predictor of lower patient mortality for patients in all body-size groups. Further reductions in mortality might be possible with increased HD dose.

[8]. Saran R, Bragg-Gresham JL, Levin NW, Twardowski ZJ, Wizemann V, Saito A, Kimata N, Gillespie BW, Combe C, Bommer J, Akiba T, Mapes DL, Young EW, Port FK. Longer Treatment Time and Slower Ultrafiltration in Hemodialysis: Associations With Reduced Mortality in the DOPPS. Kidney Int. 2006 Apr;69(7):1222-8.

Abstract: Longer treatment time (TT) and slower ultrafiltration rate (UFR) are considered advantageous for hemodialysis (HD) patients. The study included 22,000 HD patients from seven countries in the Dialysis Outcomes and Practice Patterns Study (DOPPS). Logistic regression was used to study predictors of TT > 240 min and UFR > 10 ml/h/kg bodyweight. Cox regression was used for survival analyses. Statistical adjustments were made for patient demographics, comorbidities, dose of dialysis (Kt/V), and body size. Europe and Japan had significantly longer ($P < 0.0001$) average TT than the US (232 and 244 min vs 211 in DOPPS I; 235 and 240 min vs 221 in DOPPS II). Kt/V increased concomitantly with TT in all three regions with the largest absolute difference observed in Japan. TT > 240 min was independently associated with significantly lower relative risk (RR) of mortality (RR = 0.81; $P = 0.0005$). Every 30 min longer on HD was associated with a 7% lower RR of mortality (RR = 0.93; $P < 0.0001$). The RR reduction with longer TT was greatest in Japan. A synergistic interaction occurred between Kt/V and TT ($P = 0.007$) toward mortality reduction. UFR > 10 ml/h/kg was associated with higher odds of intradialytic hypotension (odds ratio = 1.30; $P = 0.045$) and a higher risk of mortality (RR = 1.09; $P = 0.02$). Longer TT and higher Kt/V were independently as well as synergistically associated with lower mortality. Rapid UFR during HD was also associated with higher mortality risk. These results warrant a randomized clinical trial of longer dialysis sessions in thrice-weekly HD.

[9]. FHN Trial Group, Chertow GM, Levin NW, Beck GJ, Depner TA, Eggers PW, Gassman JJ, Gorodetskaya I, Greene T, James S, Larive B, Lindsay RM, Mehta RL, Miller B, Ornt DB, Rajagopalan S, Rastogi A, Rocco MV, Schiller B, Sergeyeva O, Schulman G, Ting GO, Unruh ML, Star RA, Kliger AS. In-center hemodialysis six times per week versus three times per week. N Engl J Med. 2010 Dec 9;363(24):2287-300.

BACKGROUND: In this randomized clinical trial, we aimed to determine whether increasing the frequency of in-center hemodialysis would result in beneficial changes in left ventricular mass, self-reported physical health, and other intermediate outcomes among patients undergoing maintenance hemodialysis.

METHODS: Patients were randomly assigned to undergo hemodialysis six times per week (frequent hemodialysis, 125 patients) or three times per week (conventional hemodialysis, 120 patients) for 12 months. The two coprimary composite outcomes were death or change (from baseline to 12 months) in left ventricular mass, as assessed by cardiac magnetic resonance imaging, and death or change in the physical-health composite score of the RAND 36-item health survey. Secondary outcomes included cognitive performance; self-reported depression; laboratory markers of nutrition, mineral metabolism, and anemia; blood pressure; and rates of hospitalization and of interventions related to vascular access.

RESULTS: Patients in the frequent-hemodialysis group averaged 5.2 sessions per week; the weekly standard Kt/V(urea) (the product of the urea clearance and the duration of the dialysis session normalized to the volume

of distribution of urea) was significantly higher in the frequent-hemodialysis group than in the conventional-hemodialysis group (3.54±0.56 vs. 2.49±0.27). Frequent hemodialysis was associated with significant benefits with respect to both coprimary composite outcomes (hazard ratio for death or increase in left ventricular mass, 0.61; 95% confidence interval [CI], 0.46 to 0.82; hazard ratio for death or a decrease in the physical-health composite score, 0.70; 95% CI, 0.53 to 0.92). Patients randomly assigned to frequent hemodialysis were more likely to undergo interventions related to vascular access than were patients assigned to conventional hemodialysis (hazard ratio, 1.71; 95% CI, 1.08 to 2.73). Frequent hemodialysis was associated with improved control of hypertension and hyperphosphatemia. There were no significant effects of frequent hemodialysis on cognitive performance, self-reported depression, serum albumin concentration, or use of erythropoiesis-stimulating agents.

CONCLUSIONS: Frequent hemodialysis, as compared with conventional hemodialysis, was associated with favorable results with respect to the composite outcomes of death or change in left ventricular mass and death or change in a physical-health composite score but prompted more frequent interventions related to vascular access. (Funded by the National Institute of Diabetes and Digestive and Kidney Diseases and others; ClinicalTrials.gov number, NCT00264758.).

[10]. Flythe JE, Curhan GC, Brunelli SM. Disentangling the Ultrafiltration Rate–Mortality Association: The Respective Roles of Session Length and Weight Gain. Clin J Am Soc Nephrol. 2013 Jul;8(7):1151-61

BACKGROUND AND OBJECTIVES: Rapid ultrafiltration rate is associated with increased mortality among hemodialysis patients. Ultrafiltration rates are determined by interdialytic weight gain and session length. Although both interdialytic weight gain and session length have been linked to mortality, the relationship of each to mortality, independent of the other, is not adequately defined. This study was designed to evaluate whether shorter session length independent of weight gain and larger weight gain independent of session length are associated with increased mortality.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Data were taken from a national cohort of 14,643 prevalent, thrice-weekly, in-center hemodialysis patients dialyzing from 2005 to 2009 (median survival time, 25 months) at a single dialysis organization. Patients with adequate urea clearance and delivered dialysis session ≥240 and <240 minutes were pair-matched on interdialytic weight gain (n=1794), and patients with weight gain ≤3 and >3 kg were pair-matched on session length (n=2114); mortality associations were estimated separately.

RESULTS: Compared with delivered session length ≥240, session length <240 minutes was associated with increased all-cause mortality (adjusted hazard ratio [95% confidence interval], 1.32 [1.03 to 1.69]). Compared with weight gain ≤3, weight gain >3 kg was associated with increased mortality (1.29 [1.01 to 1.65]). The associations were consistent across strata of age, sex, weight, and weight gain and session length. Secondary analyses demonstrated dose-response relationships between both and mortality.
CONCLUSIONS: Among patients with adequate urea clearance, shorter dialysis session length and greater interdialytic weight gain are associated with increased mortality; thus, both are viable targets for directed intervention.

[11]. Weiner DE, Brunelli SM, Hunt A, Schiller B, Glassrock R, Maddux FW, Johnson D, Parker T, Nissenson A. Improving clinical outcomes among hemodialysis patients: a proposal for a "volume first" approach from the chief medical officers of US dialysis providers. Am J Kidney Dis. 2014 Nov;64(5):685-95.

Abstract: Addressing fluid intake and volume control requires alignment and coordination of patients, providers, dialysis facilities, and payers, potentially necessitating a "Volume First" approach. This article reports the consensus opinions achieved at the March 2013 symposium of the Chief Medical Officers of 14 of the largest dialysis providers in the United States. These opinions are based on broad experience among participants, but often reinforced by only observational and frequently retrospective studies, highlighting the lack of high-quality clinical trials in nephrology. Given the high morbidity and mortality rates among dialysis patients and the absence of sufficient trial data to guide most aspects of hemodialysis therapy, participants believed that

immediate attempts to improve care based on quality improvement initiatives, physiologic principles, and clinical experiences are warranted until such time as rigorous clinical trial data become available. The following overarching consensus opinions emerged. (1) Extracellular fluid status should be a component of sufficient dialysis, such that approaching normalization of extracellular fluid volume should be a primary goal of dialysis care. (2) Fluid removal should be gradual and dialysis treatment duration should not routinely be less than 4 hours without justification based on individual patient factors. (3) Intradialytic sodium loading should be avoided by incorporating dialysate sodium concentrations set routinely in the range of 134-138 mEq/L, avoidance of routine use of sodium modeling, and avoidance of hypertonic saline solution. (4) Dietary counseling should emphasize sodium avoidance.

[12]. Block GA, Kilpatrick RD, Lowe KA, Wang W, Danese MD. CKD-mineral and bone disorder and risk of death and cardiovascular hospitalization in patients on hemodialysis. Clin J Am Soc Nephrol. 2013 Dec;8(12):2132-40.

BACKGROUND AND OBJECTIVES: Parathyroid hormone, calcium, and phosphate have been independently associated with cardiovascular event risk. Because these parameters may be on the same causal pathway and have been proposed as quality measures, an integrated approach to estimating event risks is needed.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Prevalent dialysis patients were followed from August 31, 2005 to December 31, 2006. A two-stage modeling approach was used. First, the 16-month probabilities of death and composite end point of death or cardiovascular hospitalization were estimated and adjusted for potential confounders. Second, patients were categorized into 1 of 36 possible phenotypes using average parathyroid hormone, calcium, and phosphate values over a 4-month baseline period. Associations among phenotypes and outcomes were estimated and adjusted for the underlying event risk estimated from the first model stage.
RESULTS: Of 26,221 patients, 98.5% of patients were in 22 groups with at least 100 patients and 20% of patients were in the reference group defined using guideline-based reference ranges for parathyroid hormone, calcium, and phosphate. Within the 22 most common phenotypes, 20% of patients were in groups with significantly ($P<0.05$) higher risk of death and 54% of patients were in groups with significantly higher risk of the composite end point relative to the in-target reference group. Increased risks ranged from 15% to 47% for death and from 8% to 55% for the composite. More than 40% of all patients were in the three largest groups with elevated composite end point risk (high parathyroid hormone, target calcium, and high phosphate; target high parathyroid hormone, target calcium, and high phosphate; and target high parathyroid hormone, target calcium, and target phosphate).
CONCLUSION: After adjusting for baseline risk, phenotypes defined by categories of parathyroid hormone, calcium, and phosphate identify patients at higher risk of death and cardiovascular hospitalization. Identifying common high-risk phenotypes may inform clinical interventions and policies related to quality of care.

[13]. Pun PH, Horton JR, Middleton JP. Dialysate calcium concentration and the risk of sudden cardiac arrest in hemodialysis patients. Clin J Am Soc Nephrol. 2013 May;8(5):797-803.

BACKGROUND AND OBJECTIVES: The optimal dialysate calcium concentration to maintain normal mineralization and reduce risk of cardiovascular events in hemodialysis patients is debated. Guidelines suggest that dialysate Ca concentration should be lowered to avoid vascular calcification, but cardiac arrhythmias may be more likely to occur at lower dialysate Ca. Concurrent use of QT-prolonging medications may also exacerbate arrhythmic risk. This study examined the influence of serum Ca, dialysate Ca, and QT interval-prolonging medications on the risk of sudden cardiac arrest in a cohort of hemodialysis patients.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: This case-control study among 43,200 hemodialysis patients occurred between 2002 and 2005; 510 patients who experienced a witnessed sudden cardiac arrest were compared with 1560 matched controls. This study examined covariate-adjusted sudden cardiac arrest risk associations with serum Ca, dialysate Ca, serum dialysate Ca gradient, and prescription of QT-prolonging medications using logistic regression techniques.
RESULTS: Patients assigned to low Ca dialysate<2.5 mEq/L were more likely to be exposed to larger serum dialysate Ca gradient and had a greater fall in BP during dialysis treatment. After accounting for covariates and baseline differences, low Ca dialysate<2.5 mEq/L (odds ratio=2.00, 95% confidence interval=1.40-2.90), higher corrected serum Ca (odds ratio=1.10, 95% confidence interval=1.00-1.30), and increasing serum dialysate Ca gradient (odds ratio=1.40, 95% confidence interval=1.10-1.80) were associated with increased risk of sudden cardiac arrest, whereas there were no significant risk associations with QT-prolonging medications.

CONCLUSIONS: Increased risk of sudden cardiac arrest associated with low Ca dialysate and large serum dialysate Ca gradients should be considered in determining the optimal dialysate Ca prescription.

[14]. Ishani A, Liu J, Wetmore JB, Lowe KA, Do T, Bradbury BD, Block GA, Collins AJ. Clinical outcomes after parathyroidectomy in a nationwide cohort of patients on hemodialysis. Clin J Am Soc Nephrol. 2015 Jan 7;10(1):90-7.

BACKGROUND AND OBJECTIVES: Patients receiving dialysis undergo parathyroidectomy to improve laboratory parameters in resistant hyperparathyroidism with the assumption that clinical outcomes will also improve. However, no randomized clinical trial data demonstrate the benefits of parathyroidectomy. This study aimed to evaluate clinical outcomes up to 1 year after parathyroidectomy in a nationwide sample of patients receiving hemodialysis.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Using data from the US Renal Data System, this study identified prevalent hemodialysis patients aged ≥18 years with Medicare as primary payers who underwent parathyroidectomy from 2007 to 2009. Baseline characteristics and comorbid conditions were assessed in the year preceding parathyroidectomy; clinical events were identified in the year preceding and the year after parathyroidectomy. After parathyroidectomy, patients were censored at death, loss of Medicare coverage, kidney transplant, change in dialysis modality, or 365 days. This study estimated cause-specific event rates for both periods and rate ratios comparing event rates in the postparathyroidectomy versus preparathyroidectomy periods.

RESULTS: Of 4435 patients who underwent parathyroidectomy, 2.0% died during the parathyroidectomy hospitalization and the 30 days after discharge. During the 30 days after discharge, 23.8% of patients were rehospitalized; 29.3% of these patients required intensive care. In the year after parathyroidectomy, hospitalizations were higher by 39%, hospital days by 58%, intensive care unit admissions by 69%, and emergency room/observation visits requiring hypocalcemia treatment by 20-fold compared with the preceding year. Cause-specific hospitalizations were higher for acute myocardial infarction (rate ratio, 1.98; 95% confidence interval, 1.60 to 2.46) and dysrhythmia (rate ratio 1.4; 95% confidence interval1.16 to 1.78); fracture rates did not differ (rate ratio 0.82; 95% confidence interval 0.6 to 1.1).

CONCLUSIONS: Parathyroidectomy is associated with significant morbidity in the 30 days after hospital discharge and in the year after the procedure. Awareness of clinical events will assist in developing evidence-based risk/benefit determinations for the indication for parathyroidectomy.

[15]. Gilbertson DT, Unruh M, McBean AM, Kausz AT, Snyder JJ, Collins AJ. Influenza vaccine delivery and effectiveness in end-stage renal disease. Kidney Int. 2003 Feb;63(2):738-43.

BACKGROUND: Influenza vaccination rates in the general population have been associated with improved outcomes, yet high-risk populations, such as end-stage renal disease (ESRD) patients, have received little attention in determining the potential benefits. This report assessed the frequency and effectiveness of influenza vaccination, while also assessing disparities in vaccination rates in the ESRD population.

METHODS: Using the United States Renal Data System research files containing claims for all Medicare ESRD patients, vaccination rates and outcomes among vaccinated and unvaccinated persons for the 1997 to 1998 and 1998 to 1999 influenza seasons were compared after adjustment for baseline demographic factors and health characteristics.

RESULTS: Vaccination rates in the ESRD population were less than 50% for each season. Influenza vaccination rates were lower in non-whites, women, younger patients, and peritoneal dialysis patients. Influenza vaccination was associated with a lower risk for hospitalization and death.
CONCLUSIONS: Despite universal coverage of free influenza vaccination, the ESRD population had a less than 50% vaccination rate for the years 1997 to 1998 and 1998 to 1999 as demonstrated by Medicare billing data. Substantial differences were found in vaccination rates among non-whites and peritoneal dialysis patients. This study confirms that the ESRD populations benefit from influenza vaccination, suggesting that dialysis providers should take advantage of all opportunities to immunize this high-risk group.

[16]. Rosenblum A, Wang W, Ball LK, Latham C, Maddux FW, Lacson E Jr. Hemodialysis catheter care strategies: a cluster-randomized quality improvement initiative. Am J Kidney Dis. 2014 Feb;63(2):259-67.

BACKGROUND: The prevalence of central venous catheters (CVCs) for hemodialysis remains high and, despite infection-control protocols, predisposes to bloodstream infections (BSIs).

STUDY DESIGN: Stratified, cluster-randomized, quality improvement initiative.

SETTING & PARTICIPANTS: All in-center patients with a CVC within 211 facility pairs matched by region, facility size, and rate of positive blood cultures (January to March 2011) at Fresenius Medical Care, North America.

QUALITY IMPROVEMENT PLAN: Incorporate the use of 2% chlorhexidine with 70% alcohol swab sticks for exit-site care and 70% alcohol pads to perform "scrub the hubs" in dialysis-related CVC care procedures compared to usual care.

OUTCOME: The primary outcome was positive blood cultures for estimating BSI rates.

MEASUREMENTS: Comparison of 3-month baseline period from April 1 to June 30 and follow-up period from August 1 to October 30, 2011.

RESULTS: Baseline BSI rates were similar (0.85 vs 0.86/1,000 CVC-days), but follow-up rates differed at 0.81/1,000 CVC-days in intervention facilities versus 1.04/1,000 CVC-days in controls (P = 0.02). Intravenous antibiotic starts during the follow-up period also were lower, at 2.53/1,000 CVC-days versus 3.15/1,000 CVC-days in controls (P < 0.001). Cluster-adjusted Poisson regression confirmed 21%-22% reductions in both (P < 0.001). Extended follow-up for 3 successive quarters demonstrated a sustained reduction of bacteremia rates for patients in intervention facilities, at 0.50/1,000 CVC-days (41% reduction; P < 0.001). Hospitalizations due to sepsis during 1-year extended follow-up were 0.19/1,000 CVC-days (0.069/CVC-year) versus 0.26/1,000 CVC-days (0.095/CVC-year) in controls (~27% difference; P < 0.05).

LIMITATIONS: Inability to capture results from blood cultures sent to external laboratories, underestimation of sepsis-specific hospitalizations, and potential crossover adoption of the intervention protocol in control facilities.

CONCLUSIONS: Adoption of the new catheter care procedure (consistent with Centers for Disease Control and Prevention recommendations) resulted in a 20% lower rate of BSIs and intravenous antibiotic starts, which were sustained over time and associated with a lower rate of hospitalizations due to sepsis.

[17]. Patel PR, Kallen AJ. Bloodstream infection prevention in ESRD: forging a pathway for success. Am J Kidney Dis. 2014 Feb;63(2):180-2.

Abstract: There should be little doubt regarding the importance of infections in the hemodialysis patient population. For years, the US Renal Data System has reported increasing hospitalization rates for all infectious diagnoses and for bacteremia/sepsis in patients treated with hemodialysis.1 In 2011, the Centers for Disease Control and Prevention (CDC) reported that although the burden of central line–associated bloodstream infections (BSIs) in hospitalized patients had declined nationally, the estimated burden of central line–associated BSIs in people treated with outpatient hemodialysis was substantial, possibly reaching 37,000 in 2008.2 Soon after, the US Department of Health and Human Services released their National Action Plan to Prevent Healthcare-Associated Infections (HAIs) for End Stage Renal Disease (ESRD) Facilities.3 The Action Plan, which was developed by the Federal Steering Committee for the Prevention of HAIs in ESRD Facilities with dialysis community stakeholder input, highlighted BSIs as a top priority for national prevention efforts.

[18]. Dalrymple LS, Mu Y, Romano PS, Nguyen DV, Chertow GM, Delgado C, Grimes B, Kaysen GA, Johansen KL. Outcomes of infection-related hospitalization in Medicare beneficiaries receiving in-center hemodialysis. Am J Kidney Dis. 2015 May;65(5):754-62.

BACKGROUND: Infection is a common cause of hospitalization in adults receiving hemodialysis. Limited data are available about downstream events resulting from or following these hospitalizations.

STUDY DESIGN: Retrospective cohort study using the US Renal Data System.

SETTING & PARTICIPANTS: Medicare beneficiaries initiating in-center hemodialysis therapy in 2005 to 2008.

FACTORS: Demographics, dual Medicare/Medicaid eligibility, body mass index, comorbid conditions, initial vascular access type, nephrology care prior to dialysis therapy initiation, residence in a care facility, tobacco use, biochemical measures, and type of infection.

OUTCOMES: 30-day hospital readmission or death following first infection-related hospitalization.

RESULTS: 60,270 Medicare beneficiaries had at least one hospitalization for infection. Of those who survived the initial hospitalization, 15,113 (27%) were readmitted and survived the 30 days following hospital discharge, 1,624 (3%) were readmitted to the hospital and then died within 30 days of discharge, and 2,425 (4%) died without hospital readmission. Complications related to dialysis access, sepsis, and heart failure accounted for 12%, 9%, and 7% of hospital readmissions, respectively. Factors associated with higher odds of 30-day readmission or death without readmission included non-Hispanic ethnicity, lower serum albumin level, inability to ambulate or transfer, limited nephrology care prior to dialysis therapy, and specific types of infection. In comparison, older age, select comorbid conditions, and institutionalization had stronger associations with death without readmission than with readmission.

LIMITATIONS: Findings limited to Medicare beneficiaries receiving in-center hemodialysis.

CONCLUSIONS: Hospitalizations for infection among patients receiving in-center hemodialysis are associated with exceptionally high rates of 30-day hospital readmission and death without readmission.

[19]. Dalrymple LS, Mu Y, Nguyen DV, Romano PS, Chertow GM, Grimes B, Kaysen GA, Johansen KL. Risk Factors for Infection-Related Hospitalization in In-Center Hemodialysis. Clin J Am Soc Nephrol. 2015 Dec 7;10(12):2170-80.

> BACKGROUND AND OBJECTIVES: Infection-related hospitalizations have increased dramatically over the last 10 years in patients receiving in-center hemodialysis. Patient and dialysis facility characteristics associated with the rate of infection-related hospitalization were examined, with consideration of the region of care, rural-urban residence, and socioeconomic status.
> DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: The US Renal Data System linked to the American Community Survey and Rural-Urban Commuting Area codes was used to examine factors associated with hospitalization for infection among Medicare beneficiaries starting in-center hemodialysis between 2005 and 2008. A Poisson mixed effects model was used to examine the associations among patient and dialysis facility characteristics and the rate of infection-related hospitalization.
> RESULTS: Among 135,545 Medicare beneficiaries, 38,475 (28%) had at least one infection-related hospitalization. The overall rate of infection-related hospitalization was 40.2 per 100 person-years. Age ≥85 years old, cancer, chronic obstructive pulmonary disease, inability to ambulate or transfer, drug dependence, residence in a care facility, serum albumin <3.5 g/dl at dialysis initiation, and dialysis initiation with an access other than a fistula were associated with a ≥20% increase in the rate of infection-related hospitalization. Patients residing in isolated small rural compared with urban areas had lower rates of hospitalization for infection (rate ratio, 0.91; 95% confidence interval, 0.86 to 0.97), and rates of hospitalization for infection varied across the ESRD networks. Measures of socioeconomic status (at the zip code level), total facility staffing, and the composition of staff (percentage of nurses) were not associated with the rate of hospitalization for infection.
> CONCLUSIONS: Patient and facility factors associated with higher rates of infection-related hospitalization were identified. The findings from this study can be used to identify patients at higher risk for infection and inform the design of infection prevention strategies.

[20]. Gilbertson DT, Wetmore JB. Infections Requiring Hospitalization in Patients on Hemodialysis. Clin J Am Soc Nephrol. 2015 Dec 7;10(12):2101-3.

> Introduction: Although the past decade has witnessed significant improvements in survival or patients receiving hemodialysis (HD) (1), hospitalization rates, particularly for infection, have not improved commensurately. Notable lack of progress is evident regarding hospitalizations for bacteremia/septicemia and pulmonary infections, such as pneumonia and influenza (2). For bacteremia/septicemia, first–year (incident) admission rates showed a 39% relative increase between 2003 and 2010 from 12.9% to 18.0%. Similarly, admission rates for prevalent patients increased 36% from 8.6% to 11.6%. Pneumonia/influenza hospitalization rates also did not improve between 2003 and 2010; although first–year admission rates decreased slightly (from 10.2% to 9.0%), rates for prevalent patients increased from 8.3% to 9.0%.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A

_____

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0369_Evidence_form-635963092729659495.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
US chronic dialysis patients are much more likely to die than age-matched individuals without ESRD. The excess mortality associated with ESRD patients on dialysis is influenced by dialysis facility practices, and is one of several important health outcomes used by providers, health consumers, and insurers to evaluate the quality of care provided in dialysis facilities.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The Standardized Mortality Ratio for Dialysis Facilities varies widely across facilities. For example, for the period 2010 – 2013, the 4 year SMR varied from 0.00 to 3.1. The mean value for 4-year SMR was 1.02 and the standard deviation was 0.28. The data used to calculate these rates is limited to those facilities with at least 3 expected deaths (reflecting how the measure is currently calculated on DFC).

Distribution of the SMR, 2010-2013:

2011: Facilities = 5004, Mean SMR = 1.02, Standard Error = .39, 10th = .057, 25th = .76, 50th = .98, 75th = 1.24, 90th = 1.52

2012: Facilities = 5155, Mean SMR = 1.02, Standard Error = .39, 10th = .058, 25th = .76, 50th = .99, 75th = 1.23, 90th = 1.52

2013: Facilities = 5279, Mean SMR = 1.02, Standard Error = .39, 10th = .057, 25th = .76, 50th = .98, 75th = 1.23, 90th = 1.51

2014: Facilities = 5409, Mean SMR = 1.02, Standard Error = .40, 10th = .056, 25th = .75, 50th = .98, 75th = 1.24, 90th = 1.53

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
There is evidence indicating that mortality for Hispanic ESRD patients is lower than mortality for non-Hispanic ESRD patients, and mortality for female ESRD patients is lower than mortality for male ESRD patients (see references below). This might suggest absence of a disparity with respect to ethnicity and female sex. However, Kalbfleisch et al (2015) demonstrate that when accounting for within facility differences in racial and ethnic composition, SMRs will vary depending on the percent of patients by race and ethnicity. Without an ethnicity adjustment, identical SMRs for one facility with predominantly Hispanic patients and one facility with predominantly non-Hispanic patients, for example, would give the false impression that quality of care at the two facilities was equivalent, when in fact ethnicity-adjusted mortality at the facility with more Hispanic patients would be lower if performance was identical. This same result holds for sex. As such the SMR is adjusted for these patient characteristics to avoid masking disparities in

care across groups. It is also adjusted for race, since historically the issue described above also applied to black patients.

To examine other sociodemographic disparities we included quintiles of socioeconomic status (defined for each patient as the median zipcode household income). This had little effect on the resulting expected deaths counts from the model.

See the section on risk adjustment for further details on adjustments for race, ethnicity, and sex based on the findings of Kalbfleisch et al (2015).

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Epidemiological: At the end of 2013 there were 661,648 patients being dialyzed of which 117,162 were new (incident) End Stage Renal Disease (ESRD) patients (USRDS 2015). ESRD mortality in the US was 33% higher than in Europe (Goodkin, 2004), suggesting that this improvement of this outcome is -possible. The components of unexplained or unexpected mortality that are actionable and associated with treatment and overall management of ESRD and other conditions are important to identify. For example, through effective volume control and fluid weight management' management of mineral and bone disease.

There is substantial evidence on the association between dialysis facility care practices, intermediate outcomes and mortality. For example, these include practices related to adequate dialysis, volume control, and appropriate management of mineral and bone disorder. Port et al, reported that dose of dialysis and BMI were both associated with mortality among hemodialysis patients. [Port 2002.] Flythe and Brunelli (2013) report that high ultrafiltration rates have been shown in several studies to be independently associated with increased risk of mortality. Rivara et al, found that high concentrations of serum calcium and phosphorus were associated with increased mortality (Rivara 2015).

Financial: Inefficient and inappropriate management of all aspects of patient ESRD care carries a high costs for both providers and payers. In 2013, total Medicare costs for the ESRD program were $30.9 billion (a 1.6% increase from 2012) (USRDS 2015).

Policy: This measure has been in use in the Dialysis Facility Reports since 1995 and on the Dialysis Facility Compare (DFC) web site (www.medicare.gov) since 2001, when the Balanced Budget Act (1997) required a system to measure and report the quality of dialysis services under Medicare.

The Dialysis Facility Reports are used by the dialysis facilities and ESRD Networks for quality improvement, and by ESRD state surveyors for monitoring and surveillance. The Standardized Mortality Ratio for Dialysis Facilities (SMR) in particular is used by ESRD state surveyors in conjunction with other standard criteria for prioritizing and selecting facilities to survey. This patient survival classification measure is reported publicly on the DFC web site to assist patients in selecting dialysis facilities.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
United States Renal Data System, 2015 annual data report: An overview of the epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015.

Goodkin DA, Young EW, Kurokawa K, Prutz K-G, Levin NW: Mortality among hemodialysis patients in Europe, Japan, and the United States: Case-mix effects. Am J Kidney Dis 2004; 44[Suppl 2]: S16–S21.

Port FK, Ashby VB, Dhingra RK, Roys EC, Wolfe RA: Dialysis dose and body mass index are strongly associated with survival in hemodialysis patients. J Am Soc Nephrol 13:1061-1066, 2002

Rivara M, Ravel V, Kalantar-Zadeh K et al. Uncorrected and Albumin-Corrected Calcium, Phosphorus, and Mortality in Patients Undergoing Maintenance Dialysis. J Am Soc Nephrol 26: 2015

Flythe JE, Curhan GC, Brunelli SM. Disentangling the Ultrafiltration Rate–Mortality Association: The Respective Roles of Session Length and Weight Gain. Clin J Am Soc Nephrol. 2013 Jul;8(7):1151-61

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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:** 0369_Data_Dictionary_Code_Table.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
This form is being used for endorsement maintenance. Updates include:
•The model now adjusts for each incident comorbidity separately rather than using a comorbidity index.
•We have also modified the indicators for diabetes by consolidating the individual indicators.
•We have included adjustments for 210 prevalent comorbidities (identified through Medicare claims)
•The measure is now limited to Medicare patients

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of deaths among eligible patients at the facility during the time period.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
This measure was developed with 12 months of data. The time window can be specified from one to four years. Currently, the measure is publicly reported using four years of data.

S.6. **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, 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.
Information on death is obtained from several sources which include the CMS ESRD Program Medical Management Information System, the Death Notification Form (CMS Form 2746), and the Social Security Death Master File. The number of deaths that occurred among eligible dialysis patients during the time period is calculated. This count includes only Medicare patients, as detailed below. It does not include deaths from street drugs or accidents unrelated to treatment: Deaths from these causes varied by facility, with certain facilities (in particular, urban facilities that treated large numbers of male and young patients) reporting large numbers of deaths from these causes and others reporting extremely low numbers (Turenne, 1996). Since these deaths are unlikely to have been due to treatment facility characteristics, they are excluded from the calculations.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of deaths that would be expected among eligible dialysis patients at the facility during the time period, given the national average mortality rate and the patient mix at the facility.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
UM-KECC's treatment history file provides a complete history of the status, location, and dialysis treatment modality of an ESRD patient from the date of the first ESRD service until the patient dies or the data collection cutoff date is reached. For each patient, a new record is created each time he/she changes facility or treatment modality. Each record represents a time period associated with a specific modality and dialysis facility. SIMS/CROWNWeb is the primary basis for placing patients at dialysis facilities and dialysis claims are used as an additional source. Information regarding first ESRD service date, death and transplant is obtained from additional sources including the CMS Medical Evidence Form (Form CMS-2728), transplant data from the Organ Procurement and Transplant Network (OPTN), the Death Notification Form (Form CMS-2746) and the Social Security Death Master File.

The denominator for SMR for a facility is the total number of expected deaths identified using all patient-records at the facility meeting inclusion criteria. The number of days at risk in each of these patient-records is used to calculate the expected number of deaths for that patient-record.

The denominator is based on expected mortality calculated from a Cox model (Cox, 1972; SAS Institute Inc., 2004; Kalbfleisch and Prentice, 2002; Collett, 1994). The model used is fit in two stages. The stage 1 model is a Cox model stratified by facility and adjusted for patient age, race, ethnicity, sex, diabetes, duration of ESRD, nursing home status, patient comorbidities, calendar year, and body mass index (BMI) at incidence. This model allows the baseline survival probabilities to vary between strata (facilities), and assumes that the regression coefficients are the same across all strata. Stratification by facility at this stage avoids biases in estimating regression coefficients that can occur if the covariate distributions vary substantially across centers. The results of this analysis are estimates of the regression coefficients in the Cox model and these provide an estimate of the relative risk for each patient. This is based on a linear predictor that arises from the Cox model, and is then used as an offset in the stage 2 model, which is unstratified and includes an adjustment for the race-specific age-adjusted state population death rates.

Assignment of Patients to Facilities
We detail patient inclusion criteria, facility assignment and how to count days at risk, all of which are required for the risk adjustment model. As patients can receive dialysis treatment at more than one facility in a given year, we assign each patient day to a facility (or no facility, in some cases) based on a set of conventions below.

General Inclusion Criteria for Dialysis Patients
Since a patient's follow-up in the database can be incomplete during the first 90 days of ESRD therapy, we only include a patient's follow-up into the tabulations after that patient has received chronic renal replacement therapy for at least 90 days. Thus, hospitalizations, mortality and survival during the first 90 days of ESRD do not enter into the calculations. This minimum 90-day period also assures that most patients are eligible for Medicare, either as their primary or secondary insurer. It also excludes from analysis patients who die or recover renal function during the first 90 days of ESRD.

In order to exclude patients who only received temporary dialysis therapy, we assign patients to a facility only after they have been on dialysis there for the past 60 days. This 60 day period is used both for patients who started ESRD for the first time and for those who returned to dialysis after a transplant. That is, deaths and survival during the first 60 days of dialysis at a facility do not affect the SMR of that facility.

Identifying Facility Treatment Histories for Each Patient
For each patient, we identify the dialysis provider at each point in time. Starting with day 91 after onset of ESRD, we attribute patients to facilities according to the following rules. A patient is attributed to a facility once the patient has been treated there for the past 60 days. When a patient transfers from one facility to another, the patient continues to be attributed to the original facility for 60 days and then is attributed to the destination facility from day 61. In particular, a patient is attributed to their current facility on day 91 of ESRD if that facility had treated him or her for the past 60 days. If on day 91, the facility had not treated a patient for the past 60 days, we wait until the patient reaches day 60 of continuous treatment at that facility before attributing the patient to that facility. When a patient is not treated in a single facility for a span of 60 days (for instance, if there were two switches within 60 days of each other), we do not attribute that patient to any facility. Patients were removed from a facility's analysis upon receiving a transplant. Patients who withdrew from dialysis or recovered renal function remain assigned to their treatment facility for 60 days after withdrawal or recovery.

If a period of one year passes with neither paid dialysis claims nor SIMS information to indicate that a patient was receiving dialysis treatment, we consider the patient lost to follow-up and do not include that patient in the analysis. If dialysis claims or other evidence of dialysis reappears, the patient is entered into analysis after 60 days of continuous therapy at a single facility.

Days at Risk for Each Patient-Record
After patient treatment histories are defined as described above, periods of follow-up time (or patient-records) are created for each patient. A patient-record begins each time the patient is determined to be at a different facility or at the start of each calendar year. The number of days at risk starts over at zero for each patient record so that the number of days at risk for any patient-record is always a number between 0 and 365 (or 366 for leap years). Therefore, a patient who is in one facility for all four years gives rise to four patient-records and is analyzed the same way as would be four separate patients in that facility for one year each. When patients are treated at the same facility for two or more separate time periods during a year, the days at risk at the facility is the sum of all time spent at the facility for the year so that a given patient can generate only one patient-record per year at a given facility. For example, consider a patient who spends two periods of 100 days assigned to a facility, but is assigned to a different facility for the 165 days between these two 100-day periods. This patient will give rise to one patient-record of 200 days at risk at the first facility, and a separate patient-record of 165 days at risk at the second facility.

This measure is limited to Medicare dialysis patients. We require that patients reach a certain level of Medicare-paid dialysis bills to be included in the mortality statistics, or that patients have Medicare-paid inpatient claims during the period. Specifically, months within a given dialysis patient-period are used for SMR calculation when they meet the criterion of being within two months after a month with either: (a) $900+ of Medicare-paid dialysis claims OR (b) at least one Medicare-paid inpatient claim. The intention of this criterion is to assure completeness of information on hospitalizations for all patients included in the analysis.

Then we use the number of days at risk in each of these patient-records to calculate the expected number of deaths for that patient-record, and sum the total number of expected deaths during all patient-records at the facility as the expected number of death for that facility. Detailed methodology is described in Statistical Risk Model and Variables S.14.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
N/A

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
N/A

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The SMR is based on expected mortality calculated from a Cox model (Cox, 1972; SAS Institute Inc., 2004; Kalbfleisch and Prentice, 2002; Collett, 1994). The model used is fit in two stages. The stage 1 model is a Cox model stratified by facility and adjusted for patient age, race, ethnicity, sex, diabetes as cause of ESRD, duration of ESRD, nursing home status from previous year, patient comorbidities at incidence, prevalent comorbidities, calendar year and body mass index (BMI) at incidence. This model allows the baseline survival probabilities to vary between strata (facilities), and assumes that the regression coefficients are the same across all strata. Stratification by facility at this stage avoids biases in estimating regression coefficients that can occur if the covariate distributions vary substantially across centers.

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. Age is included as a piecewise continuous variable with different coefficients based on whether the patient is 0-13 years old, 14-60 years old, or 61+ years old.
•Sex: We determine each patient's sex from his/her Medical Evidence Form (CMS-2728).
•Race (White, Black, Asian/PI, Native American or other): We determine race from REBUS/PMMIS, the EDB (Enrollment Data Base), and SIMS.
•Ethnicity (Hispanic, non-Hispanic or unknown): We determine ethnicity from his/her CMS-2728.
•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 less than one year, 1-2 years, 2-3 years, or 3+ years as of the period start date.
•Nursing home status in previous year: 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. The logarithm of BMI is included as a piecewise continuous log-linear term with different coefficients based on whether the log of BMI is greater or less than 3.5.
•Comorbidities at incidence: We determine each patient's comorbidities at incidence from his/her 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 indicator in the model, having a value of 1 if the patient has that comorbidity, and a value of 0 otherwise. Another categorical indicator variable is included as a covariate in the stage 1 model to flag records where patients have at least one comorbidities. This variable has a value of 1 if the patient has at least one comorbidity and a value of 0 otherwise.
•Prevalent comorbidities: We identify a patient's prevalent comorbidities based on claims from the previous calendar year. The comorbidities adjusted for include those included in Appendix A.
•Calendar year: 2010-2013
•Missing indicator variables: Categorical indicator variables are included as covariates in the stage I model to account for records with missing values for cause of ESRD, comorbidity at incidence(missing CMS-2728 form), and BMI. These variables have a value of 1 if the patient is missing the corresponding variable and a value of 0 otherwise. BMI is imputed when either missing, or outside the range of [10,70) for adults or [5,70) for children. To impute BMI, we used the average values of the group of patients with similar characteristics (age, race, sex, diabetes) when data for all four of these characteristics were available. If either race or diabetes was also missing, the imputation was based on age and sex only. If either age or sex is missing, the patient is excluded from computations.

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

•Age*Race: Black
•Ethnicity*Race: Non-White
•Diabetes as cause of ESRD*Race
•Diabetes as cause of ESRD*Vintage
•Duration of ESRD: less than or equal to 1 year *Race
•Duration of ESRD: less than or equal to 1 year* Sex
•Diabetes as cause of ESRD*Sex
•Sex*Race: Black

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
See flowchart in Appendix.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(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.

S.25. 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.26. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Dialysis Facility
If other:

S.28. 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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0369_Testing_form-635963110373772247.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0369
Measure Title: Standardized Mortality Ratio for Dialysis Facilities
Date of Submission: 4/15/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *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.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient

factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/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), 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.

1.3. What are the dates of the data used in testing? Click here to enter date range

Data from calendar years 2010 through 2013 were used for testing.

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.26*)	**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 <u>measured entities</u> 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 of the four years from 2010-2013, there were 5,004, 5,155, 5,279, and 5,409 facilities, respectively.

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*)

For each year of the four years from 2010-2013, there were 373,002, 382,145, 390,893, and 397,804 patients, respectively.

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.

N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patient level:

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

**Assessed at the start of time at risk based on calendar year and facility assignment. Medicare coverage in the model was defined as:*
1. Medicare as primary and Medicaid
2. Medicare as primary and NO Medicaid
3. Medicare as secondary or Medicare HMO

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

Proxy/Area level: 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+) without high school diploma (%)

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*)

2011 Submission

To assess reliability, we assessed the degree to which the SMR was consistent year to year. If one looks at two adjacent time intervals, one should expect that a reliable measure will exhibit correlation over these periods since large changes in patterns affecting the measure should not occur for most centers over shorter periods. Year to year variability in the SMR values was assessed across the years 2006, 2007, 2008 and 2009 based on the 5,280 dialysis centers for which an SMR is reported in the 2010 DFRs.

2016 Submission

The reliability of the Standardized Mortality Ratio (SMR) was assessed using data among ESRD dialysis patients during 2010-2013. 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 total variation of a measure that is attributable to the between-facility variation. The SMR, 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 SMR for these facilities. Within each facility, select at random and with replacement B (say 100) bootstrap samples. 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 SMR_i and repeat the process B times. Thus, for the ith facility, we have bootstrapped SMRs of $T_{i1}^*,\ldots, T_{i200}^*$. Let S_i^* be the sample variance of this bootstrap sample. From this it can be seen that

$$s_{t,w}^2 = \frac{\sum_{i=1}^{N}[(n_i-1)S_i^{*2}]}{\sum_{i=1}^{N}(n_i-1)}$$

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

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

where

$$\bar{T} = \sum n_i T_i / \sum n_i$$

is the weighted mean of the observed SMR 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 the total variation of SMR and 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 estimated IUR, which is defined by

$$IUR = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{t,w}^2}$$

can be estimated with $(s_t^2 - s_{t,w}^2)/s_t^2$.

The SMR calculation only included facilities with at least 3 expected deaths for each year.

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*)

2011 Submission

The correlation between SMR across adjacent years (2006 vs. 2007, 2007 vs 2008, and 2008 vs. 2009) ranged from 0.26 to 0.33, indicating that centers with large or small SMR tended to have larger or smaller SMR on the following year. These correlations were highly significant. Similarly, there was persistence in SMRs that were significant from year to year.

For example, there were 4.6% of facilities that had an SMR significantly greater than 1.0 in 2006 (18.3% did not have an SMR). Among those facilities, 30% were again significantly larger than 1.0 in 2007. Of the 3.1% of facilities that were significantly less than 1.0 in 2006, 18% were found to be significantly less than 1.0 in 2007. Among the 74% of facilities that had an SMR not significantly different from 1.0 in 2006, 87% remained in that category in 2007. The measure is based on complete data and is not subject to judgment or rater variability. Hence the measures of inter-rater variability are not relevant here.

2016 Submission

Table 1: IUR for One-year SMR Overall and by Facility Size, 2010-2013

	2010		2011		2012		2013	
Facility Size (Number of patients)	IUR	N	IUR	N	IUR	N	IUR	N
All Facilities	**0.32**	5004	**0.26**	5155	**0.30**	5279	**0.28**	5409
Small (<=45)	**0.07**	1137	**0.06**	1205	**0.03**	1241	**0.10**	1256
Medium (46–85)	**0.19**	1924	**0.16**	1967	**0.17**	2018	**0.17**	2132
Large (>=86)	**0.48**	1943	**0.39**	1983	**0.47**	2020	**0.42**	2022

Table 2: IUR for Four-year SMR Overall and by Facility Size, 2010-2013

Facility Size (Number of patients)	IUR	N
All Facilities	**0.59**	5935
Small (<=135)	**0.30**	1242
Medium (136–305)	**0.45**	2320
Large (>=306)	**0.73**	2373

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?*)

2011 Submission

This was not a question on the 2011 Submission Form.

2016 Submission

Overall, we found that IURs for the one-year SMR have a range of 0.26-0.32 across the years 2010, 2011, 2012, and 2013, which indicates that about thirty percent of the variation in the one-year SMR can be attributed to the between-facility differences and about seventy percent to within-facility variation. This value of IUR indicates a relatively **low degree of reliability**. When stratified by facility size, we find that, as expected, larger facilities have greater IUR.

Reliability improved when four-year data were used. Overall, we found that IUR for the four-year SMR for 2010-2013 is 0.59 which indicates that about sixty percent of the variation in the four-year SMR can be attributed to the between-facility differences (signal) and about forty percent to within-facility variation (noise). 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.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

2011 Submission

Adjusted mortality and fractions of patients achieving K/DOQI guidelines for urea reduction ratios (URRs; > or =65%) and hematocrit levels (> or =33%) were computed for 2,858 dialysis facilities from 1999 to 2002 using national data for patients with end-stage renal disease. Linear and Poisson regression were used to study the relationship between K/DOQI compliance and mortality and between changes in compliance and changes in mortality.

Measure validity is also demonstrated by the relationship of the Standardized Mortality Ratio to other quality of care indicators, including hemoglobin greater than 10 g/dL, urea reduction ratio >= 65%, percent of patients dialyzing with a fistula, and percent of patients dialyzing with a catheter.

2016 Submission

Measure validity is demonstrated by the relationship of the Standardized Mortality Ratio to other quality of care indicators, including the Standardized Hospitalization Ratio (SHR) – Admissions, the Standardized Readmission Ratio (SRR), the Standardized Transfusion Ratio (STrR), percent of patients dialyzing with a fistula, percent of patients dialyzing with a catheter, and percent of patients with Kt/V >=1.2. Spearman's rho is reported for all variables. Because the correlations were approximately the same for the four years 2010-2013, we are reporting only the 2013 correlations.

The measure is also maintained on face validity. It was reviewed by a TEP in 2006 for potential implementation on DFC. The general consensus was the SMR captured meaningful information on survival that DFC users could use to assess facility quality. In 2015, a TEP was held specifically to consider prevalent comorbidity adjustments for inclusion in the measure. The TEP's recommendations are reflected in the risk adjustment methodology.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

2011 Submission

In 2002, facilities in the lowest quintile of K/DOQI compliance for urea reduction ratio (URR) and hematocrit guidelines had 22% and 14% greater mortality rates ($P < 0.0001$) than facilities in the highest quintile, respectively. A 10-percentage point increase in fraction of patients with a URR of 65% or greater was associated with a 2.2% decrease in mortality ($P = 0.0006$), and a 10-percentage point increase in percentage of patients with a hematocrit of 33% or greater was associated with a 1.5% decrease in mortality ($P = 0.003$). Facilities in the highest tertiles of improvement for URR and hematocrit had a change in mortality rates that was 15% better than those observed for facilities in the lowest tertiles ($P < 0.0001$).

Please see the following publication for further details: Wolfe RA, Hulbert-Shearon TE, Ashby VB, Mahadevan S, Port FK. Improvements in dialysis patient mortality are associated with improvements in urea reduction ratio and hematocrit, 1999 to 2002. Am J Kidney Dis. 2005 Jan;45(1):127-35.

2016 Submission

SHR-Admissions: rho=0.20, p<.0001

SRR-Readmissions: rho=0.10, p<.0001

STrR: rho=0.21, p<.0001

AV Fistula: rho= -0.11, p<.0001

Catheter: rho=0.13, p<.0001

Hemodialysis patients with Kt/V>=1.2: rho= -0.04, p<.0001

2b2.4. 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?*)

2011 Submission

This was not a question on the 2011 Submission Form.

2016 Submission

As expected, the SMR is positively correlated with the SHR-Admissions (rho=0.20, p<.0001), SRR-Readmissions (rho=0.10, p<.0001), and the STrR (rho=0.21, p<.0001); higher standardized mortality rates in facilities are associated with higher standardized hospitalization rates, higher standardized readmissions rates and higher standardized transfusion rates. The SMR is negatively correlated with percent of patients in the facility with AV Fistula (rho= -0.11, p<.0001); lower standardized mortality rates are associated with higher rates of AV Fistula use. On the other hand, the SMR is positively correlated with catheter use (rho=0.13, p<.0001), indicating that higher values of SMR are associated with increased use of catheters. The SMR is also found to be negatively correlated (rho= -0.04, p<.0001) with the percent of hemodialysis patients with Kt/V>=1.2, again in the direction expected. Lower SMRs are associated with a higher percentage of patients receiving adequate dialysis dose.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. 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*)

N/A

2b3.2. 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*)

N/A

2b3.3. 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*)

N/A

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with** 232 **risk factors** (diabetes, sex, age, race, ethnicity, duration of ESRD, BMI at incidence, calendar year, nursing home status, 13 comorbidities at incidence and 210 prevalent comorbidities)
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

The methods for development of the risk factor models have been published and documented previously (Wolfe 1992; Wolfe 2001). The final risk adjustment is based on a Cox or relative risk model. In this model, covariates are taken to act multiplicatively on the death rate and the adjustment model is fitted with facility defining strata so as to provide valid estimates even if the distribution of adjustment variables differs across facilities. Relevant references are Cox (1972) and Kalbfleisch and Prentice (2002). All analyses are performed using SAS.

In the SMR, adjustment is made for patient age, sex, race, ethnicity, cause of ESRD, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, prevalent comorbidities, and calendar year. The SMR is also adjusted for state population death rates.

Below we discuss factors considered for inclusion in the statistical risk model, with emphasis on new factors considered since the last cycle of NQF maintenance endorsement in 2011. We present results and discussion supporting the selection of specific risk factors in the model.

Risk adjustment factors were selected for testing based on several considerations, specifically clinical criteria, expert input, factors identified in the literature as associated with mortality, and data availability. We began with a large set of patient characteristics, comorbidities (at ESRD incidence and prevalent), anthropometrics, and other characteristics. Facility characteristics were also considered. Risk factors were evaluated for appropriateness of the adjustment. For instance, it is important not to adjust for factors that reflect the results of treatment. Factors considered appropriate and supported in the literature were then investigated with statistical models, including interactions between sets of adjusters, to determine if they were empirically related to mortality. Risk factors were also evaluated for face validity as potential predictors of mortality. Finally, SDS/SES factors were evaluated based on appropriateness (whether related to disparities in care), empirical association with the outcome, and support in published literature.

Consideration of prevalent comorbidities as risk adjusters, in addition to incident comorbidities, is in part a response to stakeholder interest to adjust for more current (prevalent) comorbidities to reflect the current health status of dialysis patients, and conditions associated with mortality. CMS contracted with UM-KECC to convene a Technical Expert Panel

(TEP) in September 2015 to consider the addition of prevalent comorbidity risk adjustment. The summary report for the TEP can be found here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TechnicalExpertPanels.html.

The TEP was charged with evaluating the potential of including prevalent comorbidities in the SMR and SHR risk adjustment models. In developing its recommendations, the TEP was asked to apply the criteria for risk-adjusters developed by the National Quality Forum (NQF): (1) Risk adjustment should be based on patient factors that influence the measured outcome and are present at the start of care; (2) Measures should not be adjusted for factors related to disparities in care or the quality of care; (3) Risk adjustment factors must be substantially related to the outcome being measured; (4) Risk adjustment factors should not reflect quality of care by the provider/facility being evaluated.

The TEP evaluated a list of prevalent comorbidities derived through the following process. First, the ESRD Hierarchical Condition Categories (ESRD-HCCs) were used as a starting point to identify ICD-9 diagnosis codes related to dialysis care. Those individual ICD-9 conditions that comprised the respective ESRD HCCs, with a prevalence of at least 0.1% in the patient population, were then selected for analysis to determine their statistical relationship to mortality and/or hospitalization. This step resulted in 555 comorbidity diagnoses (out of over 3000 ICD-9 diagnosis codes in the ESRD-HCCs). Next, an adaptive lasso variable selection method was applied to these 555 diagnoses to identify those with a statistically significant relationship to mortality and/or hospitalization ($p<0.05$). This process identified 242 diagnoses. The TEP members then scored each of these diagnoses as follows:

1. Very likely the result of dialysis facility care
2. Likely the result of dialysis facility care
3. May or may not be the result of dialysis facility care
4. Unlikely to be the result of dialysis facility care
5. Very likely not the result of dialysis facility care

The TEP established that comorbidities scored as "unlikely" or "very unlikely the result of facility care" by at least half of TEP members (simple majority) were judged as appropriate for inclusion as risk-adjusters. This process resulted in 210 conditions as risk adjustors. The TEP further recommended that: (1) comorbidities for inclusion as risk-adjusters in a particular year should be present in Medicare claims in the preceding calendar year; and (2) determination of a prevalent comorbidity required at least two outpatient claims or one inpatient claim. The set of prevalent comorbidities recommended by the TEP for inclusion as risk-adjusters is presented in the model results section.

Consideration of SES/SDS risk factors:

In addition to clinical factors, we evaluated patient and area-level SDS/SES factors as risk adjusters. These were in addition to the current SDS factors of race, ethnicity, and sex. Race and sex were included in the original SMR calculation and ethnicity was added to the model in 2005.

The relationships among individual SDS factors, socioeconomic disadvantage and mortality is well-established in the general population (Singh and Siahpush, 2006; Williams, 2006; Williams and Collins, 2001). Further, individual and market or area-level measures of deprivation have been shown to contribute independently to higher mortality (Smith et al., 1998).

Area-level income and residential segregation specifically have been shown to be associated with poorer outcomes, but particularly so for racial minorities, suggesting the interplay of patient-level (race) and area-level factors related to lower income, neighborhood poverty, segregation, levels of educational attainment, and unemployment levels that jointly influence key health outcomes in mortality and morbidity (Williams, 2006; Williams and Collins, 2001). For example, Williams (2006) explains that differences in health outcomes and mortality by race persist, even after accounting for levels of SES. This suggests the potential added effect of historical and institutional discrimination (e.g., segregation; restricted educational access; fewer health-related resources in poor neighborhoods; no insurance or Medicaid status)

that have cumulatively over time led to reduced access to care. Residential segregation of blacks in the U.S., Williams and Collins argue, is a primary cause of SES differences that in turn have resulted in a high prevalence of chronic diseases and related differences in health care outcomes such as mortality (Williams and C Collins 2001, p 404-406).

The relationship between race and mortality, as well as both race and area-level SES factors and mortality in the dialysis population, is also well documented (e.g., Burrows et al, 2014; Crews et al , 2001; Eisenstein et al, 2009; Johns et al , 2014; Kucirka et al, 2010; Ricks et al, 2011; Kalbfleisch et al., 2015; Rodriguez et al, 2007; Kimmel et al, 2013; Streja et al, 2011; Yan et al., 2013; Yan et al, 2013). However, the direction of the relationship between race and mortality is inverted relative to the general population, with lower observed mortality in blacks on chronic dialysis compared to whites, although the relationship is mediated by sociodemographic and clinical factors (Norris et al., 2008; Powe, 2006; Cowie et al. 1994). For example, while black ESRD patients overall have been observed to have lower mortality compared to whites, some studies have shown this difference is attenuated or disappears once accounting for one or more area level SES factors (Eisenstein et al 2009; Johns et al 2014; Rodriguez et al 2007; Crews et al., 2011; Ricks et al., 2011; Streja et al 2011; Johns et al 2014; Yan 2013; Yan et al 2014).

Differences based on clinical factors and Hispanic ethnicity have also been observed to impact lower mortality (Streja et al 2011; Johns et al 2014; Yan 2013; Yan et al 2013; Ricks et al 2011). Taken together race and ethnicity are shown to be strongly associated with mortality but in different clinical pathways after accounting for specific clinical markers of health status. Race was included as an adjuster in the prior version of SMR because accounting for within-facility racial differences helps to clarify disparities in quality of healthcare provided to patients with ESRD (Kalbfleisch et al., 2015).

Females in the general population have lower mortality rates (CDC National Vital Statistics Reports, 2012) than males. Adjustment for sex allows for a fair comparison between dialysis facilities with patient populations that have a different mix of males and females.

Maintaining employment is a challenge for dialysis patients which in turn can influence well-being and may have a proximal impact on outcomes such as mortality. For example, Curtin et al (AJKD 1996) found that measures of functional status were higher in patients that were employed.

Insurance status is also related to health outcomes but this has not been studied extensively within the dialysis population as it relates to mortality. However some evidence suggests a link between dual eligibles and hospital utilization (Wright et al., 2015).

In sum these studies suggest notable associations with mortality differences when taking into account patient level SDS factors (race, sex, ethnicity), and area level SES factors. Additionally, employment status and type of insurance coverage (specifically Medicare-Medicaid dual eligibility) suggest a proximate relationship to health outcomes that may have downstream impacts on mortality.

Given these observed linkages, we tested these patient- and area-level SDS/SES variables based on the conceptual relationships as described above and demonstrated in the literature, as well as on the availability of data for the analyses. Measures of area-level socioeconomic deprivation are included as individual components from the Area Deprivation Index (Singh, 2003).

References:

Burrows N, Cho P, Bullard KM, Narva A, and Eggers P. Survival on Dialysis Among American Indians and Alaska Natives With Diabetes in the United States, 1995–2010. American Journal of Public Health. Supplement 3, 2014, Vol 104, No. S3. S490.

CDC National Vital Statistics Reports, Vol. 61, No. 6, October 10, 2012, Table A

Cowie C, Port F, Rust K, Harris M: Differences In Survival Between Black And White Patients With Diabetic End-Stage Renal Disease. Diabetes Care 17: 681–687, 1994

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

Crews D, Sozio S, Liu Y, Coresh J, and Powe N. Inflammation and the Paradox of Racial Differences in Dialysis Survival. J Am Soc Nephrol 22: 2279–2286, 2011.

Curtin R, Oberley E, Sacksteder P, and Friedman A. Differences Between Employed and Nonemployed Dialysis Patients. AJKD Vol 27:4. (April) 1996. 533-540.

Eisenstein E, Sun J, Anstrom K, Stafford J, Szczech L, Muhlbaier L, Mark D. Do Income Level and Race Influence Survival in Patients Receiving Hemodialysis? The American Journal of Medicine (2009) 122, 170-180.

Johns T, Estrella M, Crews D, Appel L, Anderson C, Ephraim P, Cook C, and Boulware L. Neighborhood Socioeconomic Status, Race, and Mortality in Young Adult Dialysis Patients. Am Soc Nephrol 25: epub, 2014.

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

Kalbfleisch J, Wolfe R, Bell S, Sun R, Messana J, Shearon T, Ashby V, Padilla R, Zhang M, Turenne M, Pearson J, Dahlerus C, and Li Y. Risk Adjustment and the Assessment of Disparities in Dialysis Mortality Outcomes. J Am Soc Nephrol 26: 2641–2645, 2015

Kimmel P, Fwu CW, and Eggers P. Segregation, Income Disparities, and Survival in Hemodialysis Patients. JASN. February 2013 vol. 24 no. 2 293-301.

Kucirka L, Grams M, Lessler J, Hall E, James J, Massie A, Montgomery R, and Segev D. Age and Racial Disparities in Dialysis Survival. JAMA. 2011 August 10; 306(6): 620–626. doi:10.1001/jama.2011.1127

Norris K, Mehrotra R, Nissenson A. Racial Differences in Mortality and ESRD. American Journal of Kidney Diseases, Volume 52, Issue 2, August 2008, Pages 205–208.

Powe, NR. Reverse Race And Ethnic Disparities In Survival Increase With Severity Of Chronic Kidney Disease: What Does This Mean? Clin J Am Soc Nephrol 1: 905–906, 2006;

Ricks R, Molnar M, Kovesdy C, Kopple J, Norris K, Mehrotra R, Nissenson A, Arah O, Greenland S, and Kalantar-Zadeh K. Racial and Ethnic Differences in the Association of Body Mass Index and Survival in Maintenance Hemodialysis Patients. Am J Kidney Dis. 2011 October ; 58(4): 574–582.

Rodriguez R, Sen S, Mehta K, Moody-Ayers S, Bacchetti P, and O'Hare A. Geography Matters: Relationships among Urban Residential Segregation, Dialysis Facilities, and Patient Outcomes. Ann Intern Med. 2007;146:493-501.

Singh, G. Area Deprivation and Widening Inequalities In US Mortality, 1969–1998. Am J Public Health. 2003;93(7):1137–1143

Singh G and Siahpush M. Widening Socioeconomic Inequalities In US Life Expectancy, 1980–2000. Int. J. Epidemiol. (August 2006) 35 (4): 969-979

Smith G, Hart C, Watt G, Hole D, Hawthorne V. Individual Social Class, Area-Based Deprivation, Cardiovascular Disease Risk Factors, And Mortality: the Renfrew and Paisley Study. J Epidemiol Community Health 1998; 52:399-405

Streja E, Kovesdy C, Molnar M, Norris K, Greenland S, Nissenson A, Kopple J, and Kalantar-Zadeh K. Role of Nutritional Status and Inflammation in Higher Survival of African American and Hispanic Hemodialysis Patients. Am J Kidney Dis. 2011 June ; 57(6): 883–893.

Williams D. "Race, Socioeconomic Status, and Health: The Added Effects of Racism and Discrimination. Annals of the New York Academy of Sciences. Volume 896, Issue 1, Article first published online: 6 February 2006.

Williams D, and Collins C, Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. Public Health Reports / September–October 2001. Volume 116. 404-416.

Wolfe RA et al: New Dialysis Facility Mortality Statistics (Smrs) Adjust For More Patient Characteristics. J Am Soc Nephrol 2001; 12; A1802

Wolfe R et al. Using USRDS Generated Mortality Tables To Compare Local ESRD Mortality Rates To National Rates. Kidney Int 1992; 42: 991-96

Wright B, Potter A, and Trivedi A. Federally Qualified Health Center Use Among Dual Eligibles: Rates Of Hospitalizations And Emergency Department Visits Health Affairs, 34, no.7 (2015):1147-1155

Yan G, Norris K, Xin W, Ma J, Yu A, Greene T, Yu W, and Cheung A. Facility Size, Race and Ethnicity, and Mortality for In-Center Hemodialysis. J Am Soc Nephrol 24: 2062–2070, 2013. doi: 10.1681.

Yan G, Norris K, Yu A, Ma J, Greene T, Yu W, and Cheung A. The Relationship of Age, Race, and Ethnicity with Survival in Dialysis Patients. Clin J Am Soc Nephrol 8: 953–961, 2013. doi: 10.2215.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Analyses of Comorbidities and other Clinical Factors

Table 3a presents the SMR model coefficients. Of note, it shows the coefficients on the prevalent comorbidities that were recommended by the TEP as additional risk adjusters (i.e., in addition to the risk adjusters in the SMR model since the 2011 endorsement maintenance review).

Table 3a. Model Coefficients, Data Years 2010–2013

Covariate	Coefficient	p-value
Comorbidities at start of ESRD		
At least of the comorbidities listed below	0.15783	<.0001
Atherosclerotic heart disease	0.04559	<.0001
Other cardiac disease	0.06736	<.0001
Diabetes (all types including diabetic retinopathy)*	0.01596	0.0389
Congestive heart failure	0.12221	<.0001
Inability to ambulate	0.14953	<.0001
Chronic obstructive pulmonary disease	0.07399	<.0001
Inability to transfer	0.11727	<.0001
Malignant neoplasm, cancer	0.10791	<.0001
Peripheral vascular disease	0.05252	<.0001
Cerebrovascular disease, CVA, TIA	0.01484	0.0311
Tobacco use (current smoker)	0.10783	<.0001
Alcohol dependence	0.03135	0.0989
Drug dependence	0.07436	0.0008
No Medical Evidence (CMS-2728) Form	0.0115	0.7696
Cause of ESRD		
Diabetes	0.14834	<.0001

Covariate	Coefficient	p-value
Missing	-0.02574	0.2855
Sex: Female	-0.07704	<.0001
Age		
Age (continuous)	-0.05786	0.0003
Age spline at 14	0.08753	<.0001
Age spline at 60	0.00651	<.0001
Race: black X age interaction		
Age (continuous)	-0.0371	0.1983
Age spline at 14	0.03412	0.2384
Age spline at 60	0.0009396	0.4437
Patient in nursing home	0.31026	<.0001
Incident BMI		
Log of BMI (continuous)	-0.48904	<.0001
Log of BMI spline at 35	0.57016	<.0001
BMI Missing	0.14771	<.0001
Race		
White	Reference	-
Black	0.31856	0.4275
Asian/PI	-0.33283	<.0001
Native American	-0.12939	0.0015
Other	-0.25062	<.0001
Time on ESRD		
< 1 year	-0.18009	<.0001
1 to 2 years	-0.21764	<.0001
2 to 3 years	-0.17079	<.0001
3+ years	Reference	-
Calendar year		
2010	0.1289	<.0001
2011	0.10334	<.0001
2012	0.00509	0.3735
2013	Reference	-
Ethnicity		
Hispanic	-0.31125	<.0001
Non-Hispanic ethnicity	Reference	
Unknown ethnicity	0.09259	0.0082
Ethnicity X race: nonwhite interaction		
Hispanic ethnicity	0.30208	<.0001
Unknown ethnicity	0.12773	0.0004
Race X diabetes as cause of ESRD interaction		
Asian/PI	0.04491	0.0405
Black	-0.08505	<.0001
Native American	-0.00639	0.8865
Other	0.10269	0.0266
Time with ESRD X diabetes as cause of ESRD interaction		
< 1 year	-0.20115	<.0001
1 to 2 years	-0.11321	<.0001
2 to 3 years	-0.04516	0.0004
3+ years	Reference	-
Time on ESRD: < 1 year X race interaction		
Asian/PI	-0.13672	<.0001
Black	0.03974	0.0003
Native American	-0.10883	0.0344
Other	0.26902	<.0001
Time on ESRD: < 1 year X sex: female interaction	0.00915	0.3193
Sex: female X cause of ESRD: diabetes interaction	-0.00839	0.3009

Covariate	Coefficient	p-value
Race: black X sex: female interaction	0.06686	<.0001

*The diabetes indicator includes all diabetes comorbidities on CMS-2728 and diabetes as cause of ESRD

Table 3b. Prevalent Comorbidity Coefficients, Data Years 2010–2013

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Sarcoidosis	135	0.0498	0.1881
Malign neopl prostate	185	-0.06496	<.0001
Malign neopl thyroid	193	-0.24613	<.0001
Oth severe malnutrition	262	0.17484	<.0001
Chr airway obstruct NEC	496	0.16266	<.0001
Postinflam pulm fibrosis	515	0.15118	<.0001
Malignant neopl rectum	1541	0.30273	<.0001
Mal neo liver, primary	1550	0.36764	<.0001
Mal neo upper lobe lung	1623	0.27901	<.0001
Mal neo bronch/lung NOS	1629	0.41213	<.0001
Malig neo bladder NOS	1889	0.19631	<.0001
Malig neopl kidney	1890	-0.04592	0.0198
Secondary malig neo lung	1970	0.5234	<.0001
Second malig neo liver	1977	0.90921	<.0001
Secondary malig neo bone	1985	0.71735	<.0001
Malignant neoplasm NOS	1991	0.35314	<.0001
Protein-cal malnutr NOS	2639	0.19068	<.0001
Dis urea cycle metabol	2706	-0.01549	0.7273
Senile dementia uncomp	2900	0.07334	<.0001
Drug withdrawal	2920	0.13901	0.0014
Mental disor NEC oth dis	2948	0.16473	<.0001
Cereb degeneration NOS	3319	0.10725	<.0001
Aut neuropthy in oth dis	3371	0.02175	0.1983
Grand mal status	3453	-0.00454	0.8984
Anoxic brain damage	3481	0.2873	<.0001
Cerebral edema	3485	0.21974	<.0001
Idio periph neurpthy NOS	3569	0.03128	0.0003
Neuropathy in diabetes	3572	0.0258	0.0042
Intermed coronary synd	4111	0.05768	<.0001
Angina pectoris NEC/NOS	4139	0.00621	0.5314
Prim pulm hypertension	4160	0.05884	0.0002
Chr pulmon heart dis NEC	4168	0.1898	<.0001
Prim cardiomyopathy NEC	4254	0.23084	<.0001
Cardiomyopath in oth dis	4258	0.04292	0.0329
Atriovent block complete	4260	0.15129	<.0001
Parox ventric tachycard	4271	0.18283	<.0001
Parox tachycardia NOS	4272	0.07202	0.0747
Subdural hemorrhage	4321	0.13039	<.0001
Aortic atherosclerosis	4400	0.03595	0.0233

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Lower extremity aneurysm	4423	0.02375	0.4642
Periph vascular dis NOS	4439	0.16444	<.0001
Stricture of artery	4471	-0.02833	0.0635
Oth inf vena cava thromb	4532	0.30687	<.0001
Emphysema NEC	4928	0.07809	<.0001
Bronchiectas w/o ac exac	4940	0.03515	0.3221
Food/vomit pneumonitis	5070	0.1607	<.0001
Lung involv in oth dis	5178	0.15956	0.0088
Regional enteritis NOS	5559	0.12126	0.0002
Ulceratve colitis unspcf	5569	0.02044	0.5561
Chr vasc insuff intest	5571	0.13302	<.0001
Paralytic ileus	5601	-0.01047	0.5007
Intestinal obstruct NOS	5609	0.08494	<.0001
Alcohol cirrhosis liver	5712	0.15572	<.0001
Cirrhosis of liver NOS	5715	0.41697	<.0001
Hepatic encephalopathy	5722	0.31225	<.0001
Portal hypertension	5723	0.22903	<.0001
Oth sequela, chr liv dis	5728	0.2376	<.0001
Chronic pancreatitis	5771	0.17966	<.0001
Chronic skin ulcer NEC	7078	0.14188	<.0001
Syst lupus erythematosus	7100	0.19554	<.0001
Systemic sclerosis	7101	0.39484	<.0001
Rheumatoid arthritis	7140	0.0896	<.0001
Inflamm polyarthrop NOS	7149	-0.02268	0.6699
Sacroiliitis NEC	7202	0.04558	0.2878
Gangrene	7854	0.17237	<.0001
Cachexia	7994	0.33328	<.0001
Fracture of pubis-closed	8082	0.11422	0.0001
Pelvic fracture NOS-clos	8088	0.05103	0.1367
Fx neck of femur NOS-cl	8208	0.04397	0.0051
Amput below knee, unilat	8970	-0.09002	<.0001
Amputat bk, unilat-compl	8971	-0.01234	0.7926
Amput above knee, unilat	8972	-0.11732	<.0001
Amputat leg, unilat NOS	8974	-0.08497	0.064
Candidal esophagitis	11284	0.21728	<.0001
Oth lymp unsp xtrndl org	20280	0.20078	<.0001
Mult mye w/o achv rmson	20300	0.41084	<.0001
Ch lym leuk wo achv rmsn	20410	0.37957	<.0001
Essntial thrombocythemia	23871	0.12789	0.0003
Low grde myelody syn les	23872	0.15381	0.0017
Myelodysplastic synd NOS	23875	0.20555	<.0001
DMII wo cmp nt st uncntr	25000	0.0721	<.0001
DMII wo cmp uncntrld	25002	-0.01161	0.0705
DMII keto nt st uncntrld	25010	0.0982	0.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
DMII ketoacd uncontrold	25012	0.14458	<.0001
DMI ketoacd uncontrold	25013	0.28449	<.0001
DMII hprosmlr uncontrold	25022	0.04571	0.2251
DMII renl nt st uncntrld	25040	0.03375	<.0001
DMI renl nt st uncntrld	25041	0.07679	<.0001
DMII ophth nt st uncntrl	25050	0.00575	0.482
DMI ophth uncntrld	25053	0.0629	0.0443
DMII neuro nt st uncntrl	25060	-0.00885	0.2742
DMI neuro nt st uncntrld	25061	0.03226	0.0203
DMII neuro uncntrld	25062	-0.004	0.7193
DMI neuro uncntrld	25063	0.05321	0.037
DMII circ nt st uncntrld	25070	-0.01444	0.0857
DMI circ nt st uncntrld	25071	-0.02272	0.1652
DMII circ uncntrld	25072	0.00435	0.7765
DMII oth nt st uncntrld	25080	0.12132	<.0001
DMI oth nt st uncntrld	25081	0.09973	<.0001
DMII oth uncntrld	25082	0.05006	0.0001
DMI oth uncntrld	25083	0.14618	<.0001
Glucocorticoid deficient	25541	0.31984	<.0001
Amyloidosis NEC	27739	0.32816	<.0001
Metabolism disorder NEC	27789	0.13233	0.0078
Morbid obesity	27801	0.00932	0.3779
Obesity hypovent synd	27803	-0.02953	0.3107
Sickle cell disease NOS	28260	0.61472	<.0001
Antin chemo indcd pancyt	28411	0.39212	<.0001
Other pancytopenia	28419	0.17159	<.0001
Neutropenia NOS	28800	0.19529	<.0001
Drug induced neutropenia	28803	0.29116	<.0001
Prim hypercoagulable st	28981	0.15977	<.0001
Senile delusion	29020	0.1114	0.0105
Vascular dementia,uncomp	29040	0.10829	<.0001
Dementia w/o behav dist	29410	0.10461	<.0001
Dementia w behavior dist	29411	0.12167	<.0001
Demen NOS w/o behv dstrb	29420	0.15134	<.0001
Schizophrenia NOS-unspec	29590	0.16904	<.0001
Depress psychosis-unspec	29620	0.08783	<.0001
Recurr depr psychos-unsp	29630	0.04595	0.0459
Recur depr psych-severe	29633	0.04953	0.0214
Bipolar disorder NOS	29680	0.03951	0.0718
Bipolar disorder NEC	29689	0.0765	0.1406
Episodic mood disord NOS	29690	-0.0061	0.8254
Alcoh dep NEC/NOS-unspec	30390	0.02262	0.4481
Alcoh dep NEC/NOS-remiss	30393	-0.0592	0.1194
Opioid dependence-unspec	30400	0.23963	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Opioid dependence-contin	30401	0.10216	0.0083
Drug depend NOS-unspec	30490	0.09283	0.0412
Psymotr epil w/o int epi	34540	-0.05696	0.1739
Epilep NOS w/o intr epil	34590	0.10419	<.0001
Critical illness myopthy	35981	-0.10948	0.0009
Prolif diab retinopathy	36202	-0.056	<.0001
Mod nonprolf db retinoph	36205	-0.10539	0.0017
Diabetic macular edema	36207	-0.16216	<.0001
Hyp ht dis NOS w ht fail	40291	-0.01224	0.5579
Subendo infarct, initial	41071	0.28073	<.0001
AMI NEC, unspecified	41080	-0.00835	0.8738
AMI NOS, unspecified	41090	0.04091	0.0037
Ac ischemic hrt dis NEC	41189	0.07088	0.0013
Pulm embol/infarct NEC	41519	0.02084	0.2221
Atrial fibrillation	42731	0.24876	<.0001
Atrial flutter	42732	0.06245	<.0001
Sinoatrial node dysfunct	42781	-0.04157	<.0001
Crbl emblsm w infrct	43411	0.18777	<.0001
Crbl art ocl NOS w infrc	43491	0.12749	<.0001
Athscl extrm ntv art NOS	44020	0.02718	0.0013
Ath ext ntv at w claudct	44021	0.02956	0.0173
Ath ext ntv at w rst pn	44022	0.0837	<.0001
Ath ext ntv art ulcrtion	44023	0.05416	<.0001
Dsct of thoracic aorta	44101	0.11966	0.0452
Periph vascular dis NEC	44389	0.02878	0.0596
Deep phlebitis-leg NEC	45119	-0.04641	0.1151
Ac DVT/emb prox low ext	45341	0.08701	<.0001
Ch DVT/embl low ext NOS	45350	0.05663	0.1025
Ch DVT/embl prox low ext	45351	0.03822	0.3528
Ch emblsm subclav veins	45375	0.16767	<.0001
Ac DVT/embl up ext	45382	0.07744	0.0026
Ac emblsm axillary veins	45384	0.07944	0.049
Ac embl internl jug vein	45386	0.08068	0.0006
Ac embl thorac vein NEC	45387	0.07384	0.0288
Esoph varice oth dis NOS	45621	0.18859	<.0001
Obs chr bronc w(ac) exac	49121	0.13193	<.0001
Obs chr bronc w ac bronc	49122	-0.0088	0.5824
Chronic obst asthma NOS	49320	0.01834	0.1388
Ch obst asth w (ac) exac	49322	0.01286	0.4885
Ac resp flr fol trma/srg	51851	0.02845	0.355
Ot pul insuf fol trm/srg	51852	-0.06297	0.3178
Other pulmonary insuff	51882	0.09857	<.0001
Chronic respiratory fail	51883	0.11434	<.0001
Acute & chronc resp fail	51884	0.12628	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Gastrostomy comp - mech	53642	0.15365	<.0001
Fecal impaction	56032	0.04821	0.1281
Pressure ulcer, low back	70703	0.22465	<.0001
Pressure ulcer, hip	70704	0.24053	<.0001
Pressure ulcer, buttock	70705	0.09838	<.0001
Ulcer of lower limb NOS	70710	0.09412	<.0001
Ulcer other part of foot	70715	0.08756	<.0001
Ulcer oth part low limb	70719	0.16587	<.0001
Pyogen arthritis-unspec	71100	-0.04327	0.3753
Pyogen arthritis-l/leg	71106	0.02859	0.4542
Ac osteomyelitis-unspec	73000	-0.04987	0.131
Ac osteomyelitis-ankle	73007	-0.08917	<.0001
Ac osteomyelitis NEC	73008	-0.03235	0.307
Osteomyelitis NOS-hand	73024	0.24478	<.0001
Osteomyelitis NOS-ankle	73027	-0.12149	<.0001
Path fx vertebrae	73313	0.22531	<.0001
Aseptic necrosis femur	73342	0.10754	0.0188
Asept necrosis bone NEC	73349	0.15539	0.006
Coma	78001	0.21242	<.0001
Convulsions NEC	78039	0.09323	<.0001
Fx femur intrcaps NEC-cl	82009	-0.00952	0.7647
Fx femur NOS-closed	82100	-0.02136	0.4055
React-indwell urin cath	99664	0.05432	0.0555
Compl heart transplant	99683	0.09947	0.1582
Asymp hiv infectn status	V08	0.46221	<.0001
Heart transplant status	V421	0.19932	0.0002
Liver transplant status	V427	0.03733	0.2656
Trnspl status-pancreas	V4283	0.1358	0.0026
Gastrostomy status	V441	0.02576	0.2534
Ileostomy status	V442	-0.07135	0.0349
Colostomy status	V443	0.01882	0.4186
Urinostomy status NEC	V446	0.27221	<.0001
Respirator depend status	V4611	0.08244	<.0001
Status amput othr toe(s)	V4972	-0.02421	0.1067
Status amput below knee	V4975	0.14259	<.0001
Status amput above knee	V4976	0.09281	<.0001
Atten to gastrostomy	V551	-0.05311	0.0197
Long-term use of insulin	V5867	0.0585	<.0001
BMI 40.0-44.9, adult	V8541	-0.03968	0.0375
Less than 6 months of Medicare eligible claims in the previous calendar year		0.53332	<.0001

Most of the coefficient estimates for the prevalent comorbidities are positive and statistically significant, but several do not obtain statistical significance. The very large number of clinical factors in the model expectedly generates multicollinearity among covariates, likely resulting in some unexpected results in direction of coefficient sign and levels

of statistical significance. Inclusion of this set of prevalent comorbidities reflects the consensus of the TEP that adjustment for all of these prevalent comorbidities, in addition to incident comorbidities, is important to reflect the initial and current health condition of the patient in risk adjustment.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Table 4a below presents a sensitivity analysis assessing the inclusion of additional SES measures (the base model already includes race, sex, and ethnicity). It compares coefficients in the original (baseline) SMR model with and without adjustment for the SES measures.

Table 4a. Comparing coefficients between sensitivity models with and without SES adjustors, 2010-2013: Model coefficients

	Baseline SMR			SES-adjusted SMR	
Covariate	**Coefficient**	**P-value**		**Coefficient**	**P-value**
Medicare coverage*					
Medicare primary + Medicaid	NA	NA		0.01461	0.0044
Medicare primary + no Medicaid	NA	NA		Reference	-
Medicare secondary/HMO	NA	NA		0.27131	<.0001
Employment status 6 months prior to ESRD					
Unemployed	NA	NA		Reference	-
Employed	NA	NA		0.04617	<.0001
Other/Unknown	NA	NA		0.12512	<.0001
ADI element					
Home value (median)	NA	NA		0.02098	<.0001
Family income (median)	NA	NA		-0.01099	<.0001
Income disparity**	NA	NA		-0.00043	0.8072
Monthly mortgage (median)	NA	NA		-0.01234	0.3707
< 9 years of education (%)	NA	NA		-0.00135	0.0257
No high school diploma (%)	NA	NA		0.00346	<.0001
Home ownership rate (%)	NA	NA		0.00115	<.0001
Families below the poverty level (%)	NA	NA		0.00149	0.0093
Gross rent (median)	NA	NA		-0.03188	0.0617
Single-parent households with children <18 (%)	NA	NA		-0.00172	<.0001
Unemployment rate (%)	NA	NA		0.00194	0.1061
Comorbidities at start of ESRD					
At least one of the comorbidities listed below	0.15783	<.0001		0.15872	<.0001
Atherosclerotic heart disease	0.04559	<.0001		0.04497	<.0001
Other cardiac disease	0.06736	<.0001		0.06610	<.0001
Diabetes***	0.01596	0.0389		0.00909	0.2402
Congestive heart failure	0.12221	<.0001		0.12053	<.0001
Inability to ambulate	0.14953	<.0001		0.14973	<.0001
Chronic obstructive pulmonary disease	0.07399	<.0001		0.07118	<.0001
Inability to transfer	0.11727	<.0001		0.11738	<.0001
Malignant neoplasm, cancer	0.10791	<.0001		0.10938	<.0001
Peripheral vascular disease	0.05252	<.0001		0.05068	<.0001
Cerebrovascular disease, CVA, TIA	0.01484	0.0311		0.01500	0.0295
Tobacco use (current smoker)	0.10783	<.0001		0.10764	<.0001
Alcohol dependence	0.03135	0.0989		0.03031	0.1118
Drug dependence	0.07436	0.0008		0.07526	0.0008
No Medical Evidence (CMS-2728) Form	0.0115	0.7696		0.02392	0.5432
Cause of ESRD					
Diabetes	0.14834	<.0001		0.14697	<.0001
Missing	-0.02574	0.2855		-0.02566	0.2876

Covariate	Baseline SMR			SES-adjusted SMR	
	Coefficient	P-value		Coefficient	P-value
Sex: Female	-0.07704	<.0001		-0.07910	<.0001
Age					
Continuous (years)	-0.05786	0.0003		-0.04705	0.0049
Spline at 14 years	0.08753	<.0001		0.07640	<.0001
Spline at 60 years	0.00651	<.0001		0.00687	<.0001
Race: black X age interaction					
Continuous (years)	-0.0371	0.1983		-0.04956	0.0899
Spline at 14 years	0.03412	0.2384		0.04682	0.1104
Spline at 60 years	0.0009396	0.4437		0.00019	0.8764
In nursing home the previous year	0.31026	<.0001		0.30617	<.0001
Incident BMI					
Log BMI (continuous)	-0.48904	<.0001		-0.49342	<.0001
Log BMI (spline at 35)	0.57016	<.0001		0.57780	<.0001
BMI missing	0.14771	<.0001		0.09123	<.0001
Race					
White	Reference	-		Reference	-
Black	0.31856	0.4275		0.47373	0.2443
Asian/PI	-0.33283	<.0001		-0.32944	<.0001
Native American	-0.12939	0.0015		-0.14447	0.0004
Other/unknown	-0.25062	<.0001		-0.24259	<.0001
Time on ESRD					
< 1 year	-0.18009	<.0001		-0.15762	<.0001
1 to 2 years	-0.21764	<.0001		-0.22296	<.0001
2 to 3 years	-0.17079	<.0001		-0.17220	<.0001
3+ years	Reference	-		Reference	-
Calendar year					
2010	0.1289	<.0001		0.12868	<.0001
2011	0.10334	<.0001		0.10466	<.0001
2012	0.00509	0.3735		0.00637	0.2659
2013	Reference	-		Reference	-
Ethnicity					
Hispanic	-0.31125	<.0001		-0.31963	<.0001
Non-Hispanic ethnicity	Reference	-		Reference	-
Unknown ethnicity	0.09259	0.0082		0.04305	0.2247
Ethnicity X race: nonwhite interaction				0.04305	0.2247
Hispanic ethnicity	0.30208	<.0001		0.29982	<.0001
Unknown ethnicity	0.12773	0.0004		0.13890	0.0001
Race X diabetes as cause of ESRD interaction					
Asian/PI	0.04491	0.0405		0.04655	0.0342
Black	-0.08505	<.0001		-0.08224	<.0001
Native American	-0.00639	0.8865		-0.00422	0.9251
Other	0.10269	0.0266		0.09440	0.0422
Time with ESRD X diabetes as cause of ESRD interaction					
< 1 year	-0.20115	<.0001		-0.20451	<.0001
1 to 2 years	-0.11321	<.0001		-0.11674	<.0001
2 to 3 years	-0.04516	0.0004		-0.04722	0.0002
3+ years	Reference	-		Reference	-
Time on ESRD: < 1 year X race interaction					
Asian/PI	-0.13672	<.0001		-0.12823	<.0001
Black	0.03974	0.0003		0.03854	0.0005
Native American	-0.10883	0.0344		-0.08779	0.0889
Other	0.26902	<.0001		0.28112	<.0001
Time on ESRD: < 1 year X sex: female interaction	0.00915	0.3193		0.01012	0.2716
Sex: female X cause of ESRD: diabetes interaction	-0.00839	0.3009		-0.00766	0.3454
Race: black X sex: female interaction	0.06686	<.0001		0.06466	<.0001

*Patients without Medicare coverage or with unknown coverage type were excluded from the model.

**Log(100)*(the ratio of the number of households with less than $10,000 in income to the number of households with $50,000 or more in income).

***The diabetes indicator includes all diabetes comorbidities on CMS-2728 and diabetes as cause of ESRD.

Table 4b presents a sensitivity analysis of inclusion of additional SES measures. It compares coefficients for the prevalent comorbidities that were added into the baseline SMR model to the model with adjustment for additional SES measures.

Table 4b. Comparing coefficients between sensitivity models with and without SDS/SES adjustors, 2010-2013: Prevalent comorbidity coefficients

		Baseline SMR		SES-adjusted SMR	
ICD-9 Description	**ICD-9 Code**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Protein-cal malnutr NOS	2639	0.19068	<.0001	0.18507	<.0001
Aut neuropthy in oth dis	3371	0.02175	0.1983	0.01961	0.2463
Epilep NOS w/o intr epil	34590	0.10419	<.0001	0.09632	<.0001
Cerebral edema	3485	0.21974	<.0001	0.21941	<.0001
Subendo infarct, initial	41071	0.28073	<.0001	0.26653	<.0001
AMI NEC, unspecified	41080	-0.00835	0.8738	-0.00041	0.9938
AMI NOS, unspecified	41090	0.04091	0.0037	0.05808	<.0001
Intermed coronary synd	4111	0.05768	<.0001	0.05824	<.0001
Ac ischemic hrt dis NEC	41189	0.07088	0.0013	0.07115	0.0013
Angina pectoris NEC/NOS	4139	0.00621	0.5314	0.01037	0.2964
Cardiomyopath in oth dis	4258	0.04292	0.0329	0.04335	0.0312
Atriovent block complete	4260	0.15129	<.0001	0.15412	<.0001
Parox ventric tachycard	4271	0.18283	<.0001	0.18208	<.0001
Parox tachycardia NOS	4272	0.07202	0.0747	0.07677	0.0578
Atrial fibrillation	42731	0.24876	<.0001	0.24872	<.0001
Atrial flutter	42732	0.06245	<.0001	0.05850	<.0001
Sinoatrial node dysfunct	42781	-0.04157	<.0001	-0.03410	0.0007
Subdural hemorrhage	4321	0.13039	<.0001	0.13410	<.0001
Stricture of artery	4471	-0.02833	0.0635	-0.02009	0.1885
Paralytic ileus	5601	-0.01047	0.5007	-0.01566	0.3137
Convulsions NEC	78039	0.09323	<.0001	0.09773	<.0001
Gangrene	7854	0.17237	<.0001	0.16491	<.0001
Cachexia	7994	0.33328	<.0001	0.32915	<.0001
Candidal esophagitis	11284	0.21728	<.0001	0.21573	<.0001
Sarcoidosis	135	0.0498	0.1881	0.05122	0.1762
Malignant neopl rectum	1541	0.30273	<.0001	0.30444	<.0001
Mal neo liver, primary	1550	0.36764	<.0001	0.36945	<.0001
Mal neo upper lobe lung	1623	0.27901	<.0001	0.27482	<.0001
Mal neo bronch/lung NOS	1629	0.41213	<.0001	0.41821	<.0001
Malign neopl prostate	185	-0.06496	<.0001	-0.05553	0.0002
Malig neo bladder NOS	1889	0.19631	<.0001	0.20432	<.0001
Malig neopl kidney	1890	-0.04592	0.0198	-0.04201	0.0332
Malign neopl thyroid	193	-0.24613	<.0001	-0.24139	<.0001
Secondary malig neo lung	1970	0.5234	<.0001	0.51907	<.0001
Second malig neo liver	1977	0.90921	<.0001	0.89766	<.0001
Secondary malig neo bone	1985	0.71735	<.0001	0.72095	<.0001
Malignant neoplasm NOS	1991	0.35314	<.0001	0.35642	<.0001
Oth lymp unsp xtrndl org	20280	0.20078	<.0001	0.19980	<.0001
Mult mye w/o achv rmson	20300	0.41084	<.0001	0.41119	<.0001
Ch lym leuk wo achv rmsn	20410	0.37957	<.0001	0.37275	<.0001
Essntial thrombocythemia	23871	0.12789	0.0003	0.12778	0.0003
Low grde myelody syn les	23872	0.15381	0.0017	0.15872	0.0012
Myelodysplastic synd NOS	23875	0.20555	<.0001	0.20504	<.0001
DMII wo cmp nt st uncntr	25000	0.0721	<.0001	0.08063	<.0001
DMII wo cmp uncntrld	25002	-0.01161	0.0705	-0.00322	0.616
DMII keto nt st uncntrld	25010	0.0982	0.0001	0.10744	<.0001
DMII ketoacd uncontrold	25012	0.14458	<.0001	0.13872	<.0001

ICD-9 Description	ICD-9 Code	Baseline SMR		SES-adjusted SMR	
		Coefficient	P-value	Coefficient	P-value
DMI ketoacd uncontrold	25013	0.28449	<.0001	0.27018	<.0001
DMII hprosmlr uncontrold	25022	0.04571	0.2251	0.03856	0.3067
DMII renl nt st uncntrld	25040	0.03375	<.0001	0.03346	<.0001
DMI renl nt st uncntrld	25041	0.07679	<.0001	0.08050	<.0001
DMII ophth nt st uncntrl	25050	0.00575	0.482	0.00487	0.5519
DMI ophth uncntrld	25053	0.0629	0.0443	0.05910	0.0592
DMII neuro nt st uncntrl	25060	-0.00885	0.2742	-0.00427	0.5978
DMI neuro nt st uncntrld	25061	0.03226	0.0203	0.03699	0.0078
DMII neuro uncntrld	25062	-0.004	0.7193	-0.00338	0.7615
DMI neuro uncntrld	25063	0.05321	0.037	0.05173	0.0429
DMII circ nt st uncntrld	25070	-0.01444	0.0857	-0.00987	0.2409
DMI circ nt st uncntrld	25071	-0.02272	0.1652	-0.01331	0.4165
DMII circ uncntrld	25072	0.00435	0.7765	0.00623	0.6842
DMII oth nt st uncntrld	25080	0.12132	<.0001	0.11796	<.0001
DMI oth nt st uncntrld	25081	0.09973	<.0001	0.09945	<.0001
DMII oth uncntrld	25082	0.05006	0.0001	0.04745	0.0003
DMI oth uncntrld	25083	0.14618	<.0001	0.14627	<.0001
Glucocorticoid deficient	25541	0.31984	<.0001	0.31685	<.0001
Oth severe malnutrition	262	0.17484	<.0001	0.16782	<.0001
Dis urea cycle metabol	2706	-0.01549	0.7273	-0.01721	0.6988
Amyloidosis NEC	27739	0.32816	<.0001	0.32030	<.0001
Metabolism disorder NEC	27789	0.13233	0.0078	0.13012	0.0089
Morbid obesity	27801	0.00932	0.3779	0.00456	0.6664
Obesity hypovent synd	27803	-0.02953	0.3107	-0.03330	0.253
Sickle cell disease NOS	28260	0.61472	<.0001	0.60712	<.0001
Antin chemo indcd pancyt	28411	0.39212	<.0001	0.36961	<.0001
Other pancytopenia	28419	0.17159	<.0001	0.16941	<.0001
Neutropenia NOS	28800	0.19529	<.0001	0.19467	<.0001
Drug induced neutropenia	28803	0.29116	<.0001	0.29394	<.0001
Prim hypercoagulable st	28981	0.15977	<.0001	0.15749	<.0001
Senile dementia uncomp	2900	0.07334	<.0001	0.08098	<.0001
Senile delusion	29020	0.1114	0.0105	0.11073	0.011
Vascular dementia,uncomp	29040	0.10829	<.0001	0.11062	<.0001
Drug withdrawal	2920	0.13901	0.0014	0.13186	0.0024
Dementia w/o behav dist	29410	0.10461	<.0001	0.10741	<.0001
Dementia w behavior dist	29411	0.12167	<.0001	0.13003	<.0001
Demen NOS w/o behv dstrb	29420	0.15134	<.0001	0.15265	<.0001
Mental disor NEC oth dis	2948	0.16473	<.0001	0.16480	<.0001
Schizophrenia NOS-unspec	29590	0.16904	<.0001	0.16688	<.0001
Depress psychosis-unspec	29620	0.08783	<.0001	0.08581	<.0001
Recurr depr psychos-unsp	29630	0.04595	0.0459	0.04318	0.0608
Recur depr psych-severe	29633	0.04953	0.0214	0.05826	0.0068
Bipolar disorder NOS	29680	0.03951	0.0718	0.03852	0.0792
Bipolar disorder NEC	29689	0.0765	0.1406	0.07663	0.14
Episodic mood disord NOS	29690	-0.0061	0.8254	-0.00805	0.7711
Alcoh dep NEC/NOS-unspec	30390	0.02262	0.4481	0.01772	0.5525
Alcoh dep NEC/NOS-remiss	30393	-0.0592	0.1194	-0.06103	0.1081
Opioid dependence-unspec	30400	0.23963	<.0001	0.23251	<.0001
Opioid dependence-contin	30401	0.10216	0.0083	0.09609	0.0131
Drug depend NOS-unspec	30490	0.09283	0.0412	0.09262	0.0415
Cereb degeneration NOS	3319	0.10725	<.0001	0.11542	<.0001
Grand mal status	3453	-0.00454	0.8984	-0.00611	0.8635
Psymotr epil w/o int epi	34540	-0.05696	0.1739	-0.05466	0.1919
Anoxic brain damage	3481	0.2873	<.0001	0.28681	<.0001
Idio periph neurpthy NOS	3569	0.03128	0.0003	0.03480	<.0001
Neuropathy in diabetes	3572	0.0258	0.0042	0.01952	0.0303
Critical illness myopthy	35981	-0.10948	0.0009	-0.10703	0.0011
Prolif diab retinopathy	36202	-0.056	<.0001	-0.04794	<.0001

ICD-9 Description	ICD-9 Code	Baseline SMR		SES-adjusted SMR	
		Coefficient	P-value	Coefficient	P-value
Mod nonprolf db retinoph	36205	-0.10539	0.0017	-0.09839	0.0034
Diabetic macular edema	36207	-0.16216	<.0001	-0.15551	<.0001
Hyp ht dis NOS w ht fail	40291	-0.01224	0.5579	-0.00822	0.6944
Pulm embol/infarct NEC	41519	0.02084	0.2221	0.02418	0.1565
Prim pulm hypertension	4160	0.05884	0.0002	0.07312	<.0001
Chr pulmon heart dis NEC	4168	0.1898	<.0001	0.18235	<.0001
Prim cardiomyopathy NEC	4254	0.23084	<.0001	0.22949	<.0001
Crbl emblsm w infrct	43411	0.18777	<.0001	0.18506	<.0001
Crbl art ocl NOS w infrc	43491	0.12749	<.0001	0.13064	<.0001
Aortic atherosclerosis	4400	0.03595	0.0233	0.03158	0.0465
Athscl extrm ntv art NOS	44020	0.02718	0.0013	0.03302	<.0001
Ath ext ntv at w claudct	44021	0.02956	0.0173	0.03543	0.0044
Ath ext ntv at w rst pn	44022	0.0837	<.0001	0.08269	<.0001
Ath ext ntv art ulcrtion	44023	0.05416	<.0001	0.05839	<.0001
Dsct of thoracic aorta	44101	0.11966	0.0452	0.11933	0.0462
Lower extremity aneurysm	4423	0.02375	0.4642	0.02257	0.487
Periph vascular dis NEC	44389	0.02878	0.0596	0.03332	0.0294
Periph vascular dis NOS	4439	0.16444	<.0001	0.16631	<.0001
Deep phlebitis-leg NEC	45119	-0.04641	0.1151	-0.03405	0.2481
Oth inf vena cava thromb	4532	0.30687	<.0001	0.29469	<.0001
Ac DVT/emb prox low ext	45341	0.08701	<.0001	0.07657	0.0001
Ch DVT/embl low ext NOS	45350	0.05663	0.1025	0.05742	0.0979
Ch DVT/embl prox low ext	45351	0.03822	0.3528	0.03670	0.3723
Ch emblsm subclav veins	45375	0.16767	<.0001	0.16457	0.0001
Ac DVT/embl up ext	45382	0.07744	0.0026	0.07820	0.0023
Ac emblsm axillary veins	45384	0.07944	0.049	0.07311	0.0702
Ac embl internl jug vein	45386	0.08068	0.0006	0.07453	0.0016
Ac embl thorac vein NEC	45387	0.07384	0.0288	0.07472	0.0269
Esoph varice oth dis NOS	45621	0.18859	<.0001	0.18789	<.0001
Obs chr bronc w(ac) exac	49121	0.13193	<.0001	0.12911	<.0001
Obs chr bronc w ac bronc	49122	-0.0088	0.5824	-0.00995	0.5339
Emphysema NEC	4928	0.07809	<.0001	0.08582	<.0001
Chronic obst asthma NOS	49320	0.01834	0.1388	0.01747	0.1583
Ch obst asth w (ac) exac	49322	0.01286	0.4885	0.01140	0.5388
Bronchiectas w/o ac exac	4940	0.03515	0.3221	0.04016	0.2583
Chr airway obstruct NEC	496	0.16266	<.0001	0.16095	<.0001
Food/vomit pneumonitis	5070	0.1607	<.0001	0.15828	<.0001
Postinflam pulm fibrosis	515	0.15118	<.0001	0.15382	<.0001
Lung involv in oth dis	5178	0.15956	0.0088	0.15551	0.0108
Ac resp flr fol trma/srg	51851	0.02845	0.355	0.02576	0.4026
Ot pul insuf fol trm/srg	51852	-0.06297	0.3178	-0.05118	0.4168
Other pulmonary insuff	51882	0.09857	<.0001	0.10648	<.0001
Chronic respiratory fail	51883	0.11434	<.0001	0.11153	<.0001
Acute & chronc resp fail	51884	0.12628	<.0001	0.11971	<.0001
Gastrostomy comp - mech	53642	0.15365	<.0001	0.15654	<.0001
Regional enteritis NOS	5559	0.12126	0.0002	0.11992	0.0002
Ulceratve colitis unspcf	5569	0.02044	0.5561	0.02618	0.4509
Chr vasc insuff intest	5571	0.13302	<.0001	0.12928	<.0001
Fecal impaction	56032	0.04821	0.1281	0.04974	0.1165
Intestinal obstruct NOS	5609	0.08494	<.0001	0.08695	<.0001
Alcohol cirrhosis liver	5712	0.15572	<.0001	0.15281	<.0001
Cirrhosis of liver NOS	5715	0.41697	<.0001	0.41478	<.0001
Hepatic encephalopathy	5722	0.31225	<.0001	0.30759	<.0001
Portal hypertension	5723	0.22903	<.0001	0.22448	<.0001
Oth sequela, chr liv dis	5728	0.2376	<.0001	0.23753	<.0001
Chronic pancreatitis	5771	0.17966	<.0001	0.17399	<.0001
Pressure ulcer, low back	70703	0.22465	<.0001	0.22107	<.0001
Pressure ulcer, hip	70704	0.24053	<.0001	0.24067	<.0001

ICD-9 Description	ICD-9 Code	Baseline SMR		SES-adjusted SMR	
		Coefficient	P-value	Coefficient	P-value
Pressure ulcer, buttock	70705	0.09838	<.0001	0.10478	<.0001
Ulcer of lower limb NOS	70710	0.09412	<.0001	0.09780	<.0001
Ulcer other part of foot	70715	0.08756	<.0001	0.08939	<.0001
Ulcer oth part low limb	70719	0.16587	<.0001	0.16417	<.0001
Chronic skin ulcer NEC	7078	0.14188	<.0001	0.14378	<.0001
Syst lupus erythematosus	7100	0.19554	<.0001	0.19217	<.0001
Systemic sclerosis	7101	0.39484	<.0001	0.39577	<.0001
Pyogen arthritis-unspec	71100	-0.04327	0.3753	-0.03074	0.5285
Pyogen arthritis-l/leg	71106	0.02859	0.4542	0.02339	0.5399
Rheumatoid arthritis	7140	0.0896	<.0001	0.08839	<.0001
Inflamm polyarthrop NOS	7149	-0.02268	0.6699	-0.01212	0.8198
Sacroiliitis NEC	7202	0.04558	0.2878	0.05254	0.221
Ac osteomyelitis-unspec	73000	-0.04987	0.131	-0.04126	0.2117
Ac osteomyelitis-ankle	73007	-0.08917	<.0001	-0.08530	<.0001
Ac osteomyelitis NEC	73008	-0.03235	0.307	-0.02967	0.3489
Osteomyelitis NOS-hand	73024	0.24478	<.0001	0.25059	<.0001
Osteomyelitis NOS-ankle	73027	-0.12149	<.0001	-0.12727	<.0001
Path fx vertebrae	73313	0.22531	<.0001	0.22783	<.0001
Aseptic necrosis femur	73342	0.10754	0.0188	0.10703	0.0194
Asept necrosis bone NEC	73349	0.15539	0.006	0.15596	0.0058
Coma	78001	0.21242	<.0001	0.21663	<.0001
Fracture of pubis-closed	8082	0.11422	0.0001	0.11024	0.0002
Pelvic fracture NOS-clos	8088	0.05103	0.1367	0.06459	0.0593
Fx femur intrcaps NEC-cl	82009	-0.00952	0.7647	-0.01431	0.6523
Fx neck of femur NOS-cl	8208	0.04397	0.0051	0.05341	0.0007
Fx femur NOS-closed	82100	-0.02136	0.4055	-0.01357	0.5972
Amput below knee, unilat	8970	-0.09002	<.0001	-0.08001	<.0001
Amputat bk, unilat-compl	8971	-0.01234	0.7926	-0.00414	0.9299
Amput above knee, unilat	8972	-0.11732	<.0001	-0.11178	<.0001
Amputat leg, unilat NOS	8974	-0.08497	0.064	-0.07749	0.0912
React-indwell urin cath	99664	0.05432	0.0555	0.05003	0.0778
Compl heart transplant	99683	0.09947	0.1582	0.10317	0.1429
Asymp hiv infectn status	V08	0.46221	<.0001	0.45689	<.0001
Heart transplant status	V421	0.19932	0.0002	0.19111	0.0003
Liver transplant status	V427	0.03733	0.2656	0.03314	0.3237
Trnspl status-pancreas	V4283	0.1358	0.0026	0.12049	0.0076
Gastrostomy status	V441	0.02576	0.2534	0.02395	0.288
Ileostomy status	V442	-0.07135	0.0349	-0.07559	0.0254
Colostomy status	V443	0.01882	0.4186	0.01801	0.4392
Urinostomy status NEC	V446	0.27221	<.0001	0.26452	<.0001
Respirator depend status	V4611	0.08244	<.0001	0.08209	<.0001
Status amput othr toe(s)	V4972	-0.02421	0.1067	-0.02797	0.0622
Status amput below knee	V4975	0.14259	<.0001	0.13869	<.0001
Status amput above knee	V4976	0.09281	<.0001	0.09153	<.0001
Atten to gastrostomy	V551	-0.05311	0.0197	-0.04863	0.0326
Long-term use of insulin	V5867	0.0585	<.0001	0.05185	<.0001
BMI 40.0-44.9, adult	V8541	-0.03968	0.0375	-0.04271	0.0252
Less than 6 months of Medicare eligible claims in the previous calendar year	—	0.53332	<.0001	0.44731	<.0001

Patient-level SDS: Compared with men, women were less likely to die (OR=0.92; p<0.01). Patients of Asian/PI, Native American and Other/unknown race, respectively, all had lower odds of mortality compared to the reference group of white patients (OR=0.72, p<0.01; OR= 0.87, p<0.01; OR=0.78, p<0.01). Mortality in Black patients was not significantly

different from the reference group. We did find that Hispanic patients had lower odds of mortality (OR=0.73, p<0.01), consistent with observations in previous studies

Patient-level SES: Patients employed prior to ESRD incidence, and patients with unknown employment status (OR=1.13, p<0.01) had higher odds of mortality (OR=1.05; p<0.01) compared to unemployed patients. Note that for employment categories, the "Other/Unknown" category represents a diverse patient group with regard to SES, such as students, homemakers and those who are retired. Compared with Medicare-only patients, patients with both Medicare and Medicaid (OR=1.01; p=.004) and patients with Medicare as secondary/Medicare HMO (OR=1.31; p<0.01) had higher odds of mortality. The result for dually eligible patients having higher mortality is consistent with the hypothesis that this insurance category, on average, represents an at-risk group, but further examination is needed for the higher odds of mortality for patients with Medicare as secondary payer or HMO. It is possible that these patients represent a larger portion of incident ESRD patients, which has a known higher mortality in the first year of ESRD.

Area-level SES: Areas with high measures of deprivation are likely to have higher mortality as demonstrated in the literature for the general population as well as for the ESRD population. In general, we observed small effects on odds of mortality, in the expected direction, for most of the individual indicators of area deprivation, with several achieving statistical significance. This included a low percentage of the population with a high school diploma. The percentage of single parent households with children <18 years however had a slightly negative impact on odds of mortality. But this could be attributed to being a generally a younger population that qualifies for social assistance and Medicaid. Overall the results provide nominal support for the postulated relationships between indicators of area- level deprivation and mortality. Further analysis would need to be conducted to determine any differences in impact when combining these factors into a composite measure of area-level deprivation. But this will be subject to data availability.

The figure below shows the correlation between facility SMRs with and without adjustment for patient and area-level SES.

Figure 1. Correlation between SMR with and without SES adjustment, 2010-2013

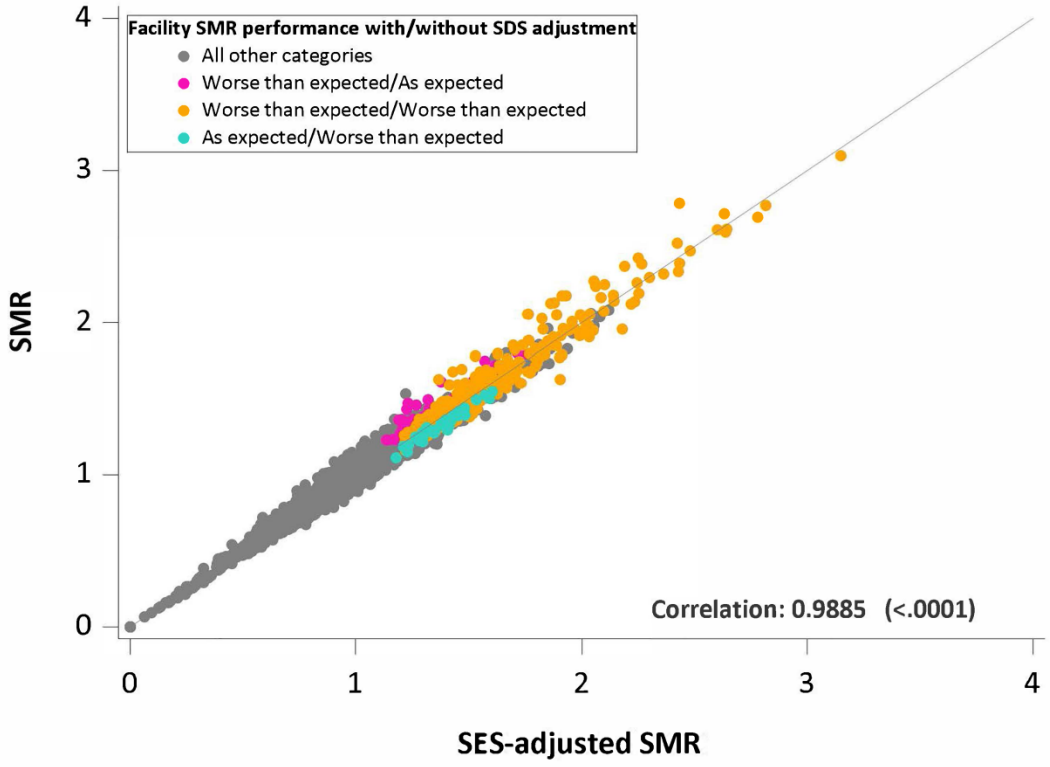


Table 5. Flagging rates, by model with and without all SES adjustors: 2010-2013

	With SES			
Without SDS *(current model)*	Better than Expected	As Expected	Worse than Expected	**Total**
Better than Expected	400	57	0	**457 (7.7%)**
As Expected	52	4938	33	**5023 (84.7%)**
Worse than Expected	0	57	393	**450 (7.6%)**
Total	**452 (7.6%)**	**5052 (85.2%)**	**426 (7.2%)**	—

After adjustment for patient and area-level SES, 199 facilities (3.4%) changed performance categories. Ninety (1.5%) facilities were down-graded, and 109 (1.8%) were upgraded.

These analyses indicate that some patient-level SES variables affect expected death rates, while most patient and area-level SES indicators have at most minimal effect. Furthermore, SMRs with and without adjustment for patient SES and area SES are highly correlated (0.9885, $p<0.0001$), and adjustment for SES shifts facility performance only slightly. This suggests SES does not contribute much to the flagging profiles for facility performance.

Risk adjustment for SES factors would probably reduce the likelihood of penalizing facilities serving a disproportionately larger disadvantaged patient population, resulting in lower quality performance scores and incentive payment reductions for the facility. At the same time, risk adjustment for SES may improve access to care for disadvantaged patients, by guarding against the potential providers may be otherwise less willing to take on these patients because of their higher comorbidity burden. This in effect comes with the risk of effectively holding providers to different (more relaxed) standards for expected patient outcomes, and relatedly may reduce access to the highest quality care for disadvantaged patients. Not adjusting for these sociodemographic and SES factors minimizes the likelihood of reinforcing disparities and counters the notion that different standards in care are acceptable in these populations. In the absence of definitive evidence demonstrating that socioeconomic risk adjustment does not result in differential access to care, we believe that the most appropriate decision is not to risk adjust for socioeconomic factors. Our primary goal should be to implement quality measures that result in the highest quality of patient care and equitable access for all patients to that care.

In the final SMR model we continue to include race, ethnicity, and sex (SDS factors) for risk adjustment based on results from the literature, discussed in section 2b4.3. Patient level SES factors are not included in the final risk adjusted model. Given the very small impact of area-level SES factors we decided not to include these as risk adjustments in the final model. While other studies have shown the association between these patient and area-level SES factors and mortality, further work is needed to demonstrate that differences based on these factors are not related to facility care, in order to prevent disparities in care.

2b4.5. 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 2b4.9

See 2b4.3.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

In this model, the C-Index=0.724 which suggests good predictive ability of the risk model.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

N/A

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

See Figure 2 in 2b4.10.

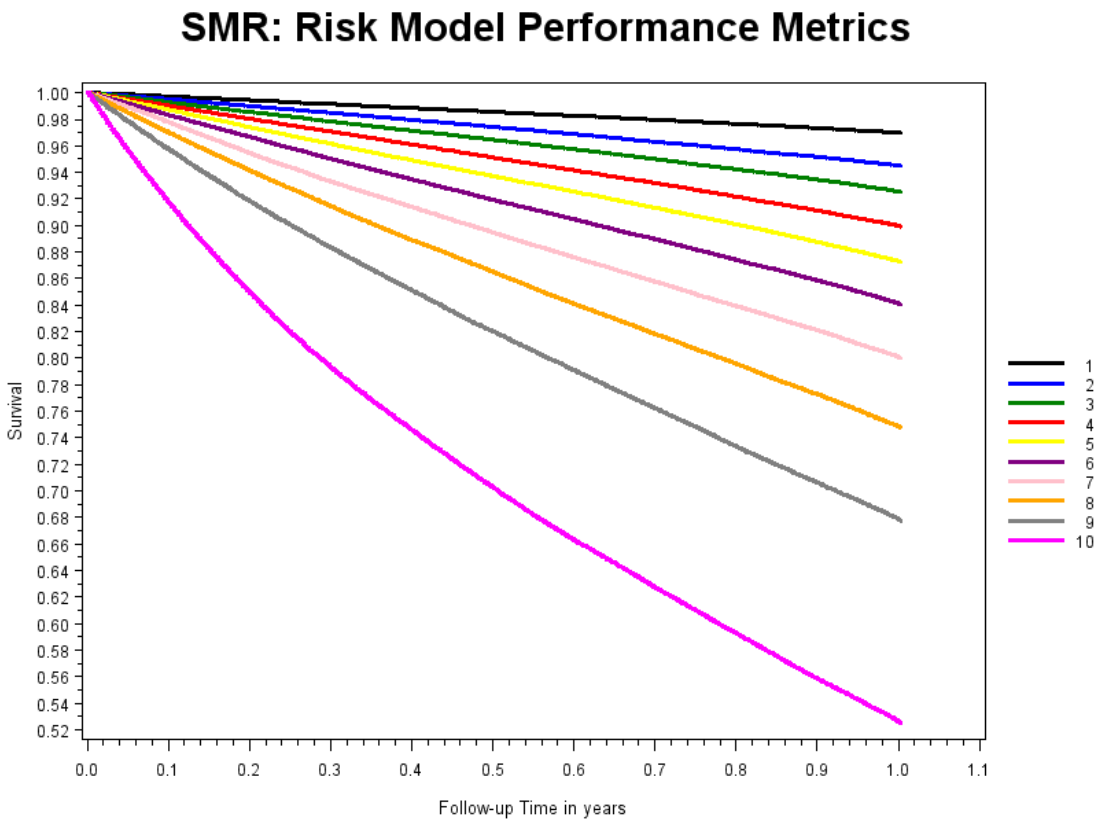
2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. 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*)

Figure 2 is the decile plot showing estimates of cumulative rates by years. The plot shows 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 the best survival rates). The absolute differences between the groups is also large with survival at one year ranging from 96% for those patients predicted to have the lowest mortality rates (group 1) down to 60% for those predicted to have the lowest rates of survival (group 10).

Figure2. Decile plot for SMR



2b4.11. 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

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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 p-value for a given facility is a measure of the strength of the evidence against the hypothesis that the mortality rate for this facility is identical to that seen nationally overall, having adjusted for the patient mix. Thus, the p-value is the probability that the facility's SMR would deviate from 1.00 (national rate) by at least as much as the facility's observed SMR. In practice, the p-value is computed using a Poisson approximation under which the distribution of the number of deaths in the facility is Poisson with a mean value equal to E, the expected number of deaths as computed from the Cox model. Accordingly, if the observed number, O, is greater than E, then p-value = 2 * Pr(X>=O) where X has a Poisson distribution with mean E. Similarly, if O<E, the p-value = 2 * Pr(X <=O) where X has a Poisson distribution with mean E.

2b5.2. 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*)

Table 6. Number and percentage of facilities by classification of the 2013 SMR. Categories stratified by facility size.

Number of patients	Better than expected	As expected	Worse than expected
<=45	0.48% (26)	21.09% (1141)	0.54% (29)
45-85	1.09% (59)	37.93% (2052)	1.50% (81)
>=86	2.03% (110)	33.48% (1811)	1.87% (101)

Table 7. Number and percentage of facilities by classification of the 2010-2013 SMR. Categories stratified by facility size.

Number of patients	Better than expected	As expected	Worse than expected
<=135	0.69% (41)	19.05% (1131)	1.18% (70)
136-305	2.21% (131)	34.38% (2041)	2.49% (148)
>=306	4.80 % (285)	31.28% (1857)	3.91% (232)

2b5.3. 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?*)

Facilities are flagged if they have outcomes that are extreme when compared to the variation in national death rates adjusted for patient case-mix.

For both the one-year SMR and four-year SMR, a majority of facilities had mortality that was "As Expected." Overall, for the 2013 SMR, approximately 3.6% of facilities had SMR that was "Better than expected," while 3.9% of all facilities had SMR that was "Worse than expected." Across all facilities, for the 2010-2013 SMR, approximately 7.7% of facilities had a SMR that was "Better than expected," while 7.6% of facilities had a SMR that was "Worse than expected."

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

2b7.2. 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*)

2b7.3. 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)

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)

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)*

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.**

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.**

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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

N/A

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Dialysis Facility Compare http://www.medicare.gov/dialysisfacilitycompare/

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting: Dialysis Facility Compare (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 who are eligible for the measure, and have at least 3 expected deaths during 2010-2013. For the most recent DFC report, that was 5916 facilities.

Patients included: All patients who meet the requirements to be included in the measure.

4a.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

4a.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

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Mortality rates have decreased over time as evidenced by the coefficients for calendar year from the SMR model. The mortality rate for 2011 was 2.6% lower compared to 2010 (p-value<0.0001), and the rates for 2012 and 2013 were lower compared to 2010 at 12.4% and 13.0%, respectively (p-value <0.0001).

2011: Coefficient = -0.026, P-value = <0.0001

2012: Coefficient = -0.124, P-value = <0.0001
2013: Coefficient = -0.130, P-value = <0.0001

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
In the past, a concern has been raised about patient selection relating to ensuring access to care for sicker patients. CMS considers ensuring patient access to dialysis care to be particularly important, and we continue to seek ways to ensure that access is unabated as part of the measures we develop, specifically outcome measures like SMR. The SMR measure incorporates a risk adjustment methodology that levels the playing field for facilities with different patient case-mixes in order to dis-incentivize cherry-picking of healthier patients over sicker patients at higher risk of mortality. Given the extensive list of adjustments for patients' prevalent comorbidities, which reflect a substantial modification to the SMR, we think this additional risk adjustment strategy would discourage avoidance of treating sicker and more complex patients that may also have limited life expectancy, and that may be more likely to die while receiving care at their facility.

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.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
1463 : Standardized Hospitalization Ratio for Dialysis Facilities
2496 : Standardized Readmission Ratio (SRR) for dialysis facilities

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The specifications are not completely harmonized. Each measure assesses different outcomes as reflected in certain differences across the measure specifications. SMR, and SHR and SRR are harmonized to the population they measure (Medicare-covered ESRD

patients), methods (SMR and SHR) and certain risk adjustment factors specific to the ESRD population. SMR and SHR adjust for the same comorbidity risk factors, a similar set of patient characteristics, and use fixed effects in their modeling approach. The differences between SMR and SHR and SRR reflect adjustment for factors specific to the outcome of each respective measure. Both SMR and SHR adjust for a set of prevalent comorbidities (observed in a prior year), however the complete set of comorbidities for SMR differs from SRR. SRR, a measure of hospital utilization adjusts for planned readmissions; and for discharging hospital, acknowledging that for readmission, hospitals also bear accountability for properly coordinating care with the dialysis facility. These risk adjustments in SRR account for those characteristics specifically associated with readmission, and do not apply to SMR. Only SMR adjusts for state death rates, race, and ethnicity to account for these respective differences related to mortality outcomes and that are deemed outside of a facility's control.

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.)
N/A

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:** 0369_Appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, Sophia.Chan@cms.hhs.gov, 410-786-5050-
Co.3 **Measure Developer if different from Measure Steward:** University of Michigan Kidney Epidemiology and Cost Center
Co.4 **Point of Contact:** Casey, Parrotte, parrotte@med.umich.edu

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.
The following is a list of TEP members who participated in the End-Stage Renal Disease Evaluation of Potential Prevalent Comorbidity Adjustments in the Standardized Hospitalization Ratio (SHR) and the Standardized Mortality Ratio (SMR) TEP. In this advisory role, the primary duty of the TEP was to review any existing measures in terms of comorbidities included as adjusters, and determine if there was sufficient evidence to support the inclusion of specific proposed comorbidities as measure adjusters, and relatedly, suggest measure specifications.

Caroline Steward, APRN, CCRN, CNN
Advanced Practice Nurse (Hemodialysis)
Capital Health System
Trenton, NJ

Dana Miskulin, MD, MS
Staff Nephrologist
Tufts Medical Center
Boston, MA
Associate Professor of Medicine

Outcomes Monitoring Program, Dialysis Clinic Inc.
Nashville, TN

David Gilbertson, PhD
Co-Director of Epidemiology and Biostatistics
Chronic Disease Research Group
Minneapolis, MN

Eduardo Lacson Jr, MD, MPH
Nephrologist
American Society of Nephrology
Lexington, MA

Jennifer Flythe, MD, MPH
Research Fellow
University of North Carolina at Chapel Hill
Assistant Professor of Medicine
Chapel Hill, NC

Lorien Dalrymple, MD, MPH
Associate Professor
University of California, Davis
Division of Nephrology
Sacramento, CA

Mark Mitsnefes, MD, MS
Professor of Pediatrics
Cincinnati Children's Hospital Medical Center
Program Director
University of Cincinnati
Cincinnati, OH

Roberta Wager, MSN, RN
Renal Care Coordinator
Fresenius Medical Care
Member of Forum of ESRD Networks Beneficiary Council
Forum of ESRD Networks
Boerne, TX

Danielle Ward
Member of Forum of ESRD Networks Beneficiary Council
Forum of ESRD Networks
Board Member
Network 6
Wake Forest, NC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 1995
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, 2017

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

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

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1463
Measure Title: Standardized Hospitalization Ratio for Dialysis Facilities
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: Standardized hospitalization ratio for dialysis facility patients. This measure is calculated as a ratio but can also be expressed as a rate.
Developer Rationale: Hospitalization rates are an important indicator of patient morbidity and quality of life. On average, dialysis patients are admitted to the hospital nearly twice a year and spend an average of 11.2 days in the hospital per year [1]. Hospitalizations account for approximately 40 percent of total Medicare expenditures for ESRD patients [1]. Measures of the frequency of hospitalization have the potential to help efforts to control escalating medical costs, and to play an important role in identifying potential problems and helping facilities provide cost-effective health care.

[1] United States Renal Data System. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015

Numerator Statement: Number of inpatient hospital admissions among eligible patients at the facility during the reporting period.
Denominator Statement: Number of hospital admissions that would be expected among eligible patients at the facility during the reporting period, given the patient mix at the facility.
Denominator Exclusions: None.

Measure Type: Outcome
Data Source: Administrative claims, Electronic Clinical Data
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Aug 16, 2011 **Most Recent Endorsement Date:** Aug 16, 2011

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

The developer reports the following:

- This measure calculates the standardized hospitalization ratio for dialysis facility patients. This measure is calculated as a ratio but can also be expressed as a rate.
- As a rationale for measuring this health outcome, the developer provides evidence that states:
  - Hospitalization rates remain very high in US chronic dialysis patients relative to the general population, despite a nearly 20% decline from 2005-2013.
  - According to the 2015 USRDS Annual Report, approximately ½ of all dialysis patient hospitalizations continue to be caused by cardiovascular or infectious causes over that time period.
  - Programs developed to impact dialysis provider practices have been shown to improve intermediate outcomes (reduced catheter vascular access, small solute adequacy, anemia management) and mortality, modality options, infection prevention, and dialysis organization culture. These practice improvements have been linked to reduced hospitalizations in this population.

Guidance from the Evidence Algorithm :
Measure assesses performance on a health outcome (Box 1)→ Relationship established between measured health outcome and at least one healthcare action (Box 2) → PASS

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence?*

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states the standardized hospitalization admission rates vary widely across facilities. For 2014, the SHR Admissions varied from 0.07 to 2.92. The mean value was 0.99 and the Standard Deviation (or error) was 0.27.
- The data used to calculate these rates is limited to those facilities with at least 5 patient years at risk
- The developer reports following distribution of the SHR, 2011-2014:

Year	Facilities	Mean SHR	Standard Error	10th percentile	90th percentile
2011	5386	.99	.28	.66	1.33
2012	5568	.99	.28	.66	1.34
2013	5700	.99	.27	.68	1.33
2014	5857	.99	.27	.68	1.32

Disparities
The developer provides the following information:

- Race and ethnicity have been shown to be predictors of hospitalization. Using data from 2013, the developer observed the following:
  - White and black patients are hospitalized at similar rates (both SHRs = 1.01).
  - Native American and Asian/Pacific Islander patients are hospitalized at lower rates than would be expected (SHR = 0.90 and 0.84, respectively).
  - Hispanic patients had slightly lower than expected hospitalization rates (SHR = 0.98), while non-Hispanic and patients of unknown ethnicity were hospitalized at the same rate (both SHRs = 1.00).
- While there are differences across the race and ethnicity groups, the results suggest no clear disparities in outcomes and that it would not be appropriate to adjust for these factors.

Questions for the Committee:

- *Would it be easier to understand the opportunity for improvement if the measure were expressed as a rate?*
- *Is there a gap in care that warrants a national performance measure?*
- *I Is it appropriate for the developer to not adjust for race and ethnicity based on the evidence provided?*
- *There has been very little (if any) change in performance over time, even with additional facilities reporting – is this lack of variation indicative of quality problems?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **While there has been virtually no change in overall SHR for 4 years 2011-2014, rates very widely across facilities in the US. Lack of change in SHR may reflect the degree of illness in the population - or may reflect the absence of sufficient QI efforts to keep patients healthy at home.

No troublesome disparities across race or ethnicity

Use of a ratio in this context rather than a rate is fine from my perspective

**This measure calculates as a ratio the standardized hospitalization ratio for dialysis facility patients. The rationale for measuring this health outcome includes: Hospitalization rates still remain very high in US chronic dialysis patients relative to the general population, and despite a nearly 20% decline from 2005-2013. Approximately ½ of all dialysis patient hospitalizations continue to be caused by cardiovascular or infectious causes over that time period. Programs developed to impact dialysis provider practices such as reduced catheter vascular access, small solute adequacy, anemia management have improved outcomes. These practice improvements have been linked to reduced hospitalizations in this population.

**The evidence supports the relationship between facility performance and hospitalization. The developers though have purposely excluded some adjustors which are reflective either of natural disease progression or dialysis unit underperformance. The inability to proportion between those two causes is problematic.

The denominator is stated as the predicted number of hospitalizations at that facility. Though prior approaches including the SMR include the denominator as the national average. Could this be clarified?

**Reliability testing looks good

**Yes. Rationale remains unchanged. Hospitalization may reflect opportunities for improvement at several points in the care process.

**No new evidence to discuss

**evidence provided of high rates of hospitalization in dialysis pts and how practice improvements can decrease hospitalization

1b. Performance Gap

Comments: **The evidence of a performance gap is thus equivocal (see above)

Preliminary rating: Moderate

**Standardized hospitalization admission rates vary widely across facilities. Race and ethnicity have been shown to be predictors of hospitalization. Data was provided on population subgroups. However , the results suggest no clear disparities in outcomes and that it would not be appropriate to adjust for these factors

**Assuming the current claims based SHR is being used, then a gap has been demonstrated.

**There is a wide spread in outcomes – unclear how much variability is due to chance and how much is due to practices.

**High rates and variation suggest performance gap on face. Disparities based on race or ethnicity not demonstrated.

**Agree with staff comments

**average SHR 0.99 but 10%ile 0.66 and 90%ile 1.32 so some gap exists; no clear disparities in evidence provided

1c. High Priority (previously referred to as High Impact)

Comments: **Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data

Specifications:

- The following updates have been made since the last submission
  - The model now adjusts for each incident comorbidity separately rather than using a comorbidity index.
  - The indicators for diabetes were modified by consolidating the individual indicators.
  - Adjustments for 210 prevalent comorbidities (identified through Medicare claims) were included.
- The numerator of this measure is: *Number of inpatient hospital admissions among eligible patients at the facility during the reporting period.*
- The denominator of this measure is: *Number of hospital admissions that would be expected among eligible patients at the facility during the reporting period, given the patient mix at the facility.*
- The ICD-9 and ICD-10 codes have been included in the Data Dictionary Code Table.
- The calculation algorithm is stated in S.18 and appears straightforward.
- This outcome measure is risk adjusted, using a statistical risk model.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*
- *How will the changes to the specifications affect ranking of facilities?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- To assess reliability, the developer assessed the degree to which the Standardized Hospital Ratio for Admissions was consistent year to year using data on hospitalizations among ESRD patients. Year to year variability in the SHR values was assessed across the years 2006, 2007 and 2008 based on the 4338 dialysis centers for which an SHR is reported in the 2010 Dialysis Facility Reports (DFRs).
- The correlation between SHR admissions across adjacent years (2006 versus 2007 and 2007 vs 2008) was approximately 0.67 indicating that centers with large or small SHR tended to have larger or smaller SHR on the following year. Similarly, there was persistence in SHRs that were significant from year to year.
- The developer stated the measure is based on complete data and is not subject to judgment or rater variability. Hence the measures of inter-rater variability are not relevant here.
- The Committee questioned the need for a 3-year time period, and the developer indicated that 1 year was acceptable.

Describe any updates to testing: The reliability of the SHR was assessed using data among Medicare ESRD dialysis patients during 2010-2013.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- Since the SHR is not a simple average, the developer estimated the inter-unit reliability (IUR) using a bootstrap approach, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by an one-way analysis of variance (ANOVA).
- The developer states 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.

Results of reliability testing

Table 1: IUR for one-year SHR, Overall and by Facility Size, 2010-2013

	2010		2011		2012		2013	
Facility Size (Number of patients)	**IUR**	**N**	**IUR**	**N**	**IUR**	**N**	**IUR**	**N**
All	0.72	5407	0.71	5583	0.70	5709	0.70	5864
Small (<=50)	0.54	1864	0.51	1921	0.48	1977	0.46	2028
Medium (51–87)	0.65	1702	0.63	1785	0.58	1825	0.57	1930
Large (>=88)	0.81	1841	0.81	1877	0.81	1907	0.82	1906

- IURs for the one-year SHRs have a range of 0.70-0.72 across the years 2010, 2011, 2012 and 2013, which developer interprets as indicating that over two-thirds of the variation in the one-year SHR can be attributed to the between-facility differences and less than one-third to within-facility variation.
- When stratified by facility size, the developer found that larger facilities have greater IUR.

Guidance from the Reliability Algorithm :
Specifications precise unambiguous and complete (Box 1)→ Empirical reliability testing conducted (Box 2)→ Testing conducted at computed measure score level (Box 4)→ Method described and appropriate (Box 5) → Level of certainty or confidence that measure scores are reliable (Box 6) →MODERATE (rationale that reliability improves as the sample sizes increase, medium and small facilities have lower reliability estimates)

Questions for the Committee:

1. *Is the test sample adequate to generalize for widespread implementation?*
2. *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
3. *The reliability testing provided by the developer has been updated but has not significantly affected the results. Does the Committee agree there is no need for repeat discussion and vote on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Validity of the Standardized Hospital Ratio for Admissions was assessed using data on hospitalizations as well as other quality measures among ESRD patients over a three year period of 2006-2008 by examining its covariability with other measures of quality as well as by examining the relationship of the overall hospitalization measure with measures that were more directly focused on specific causes. The developer provided the following results:
  - The SHR Admissions measure is correlated with the Standardized Mortality Ratio (SMR) over the three year cohort (r=0.40) and in individual years r was approximately equal to 0.33, both correlations being highly significant.
  - SHR Admissions is negatively correlated in each of the three year with percent of patients in the facility with AV Fistula (r=-0.27, -0.23, -0.21). Thus higher values of SHR are associated with lower usage of AV Fistulas.
  - SHR admissions is positively correlated with catheter use (r=0.24, 0.23, 0.22), indicating that higher values of SHR are associated with increased use of catheters. These correlations are all highly significant (p<0.001).
  - The SHR Admissions is also found to be negatively correlated (r=-0.10, p<0.0001) with the percent of patients with URR>65, again in the direction expected.
- Additionally, hospitalization measures were reviewed by a Technical Expert Panel (TEP) in 2007 and overall measures based on admissions and on days were recommended for inclusion in the Dialysis Facility reports. The developer also stated that there is a very strong case for face validity of the SHR admissions measure hospitalization because it is a major cost factor in the management of ESRD patients.

Describe any updates to validity testing: Empirical validity testing of the measure score updated with 2010-2013 data and new face validity was conducted with a TEP in 2015.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- The developer assessed the validity of the measure through various comparisons of this measure with other quality measures in use, using Spearman correlations.
- In 2015, a TEP was held specifically to consider prevalent comorbidity adjustments for inclusion in the measure. The TEP's recommendations are reflected in the risk adjustment methodology.

Validity testing results:

- The SHR Admissions measure is correlated with the Standardized Mortality Ratio (SMR) for each individual year from 2010-2013, where Spearman's correlation coefficient ranged from 0.27 to 0.30, with all four correlations being highly significant (p<0.0001).
- For each year from 2011-2013, the SHR was correlated with the Standardized Readmission Ratio (SRR) (Spearman's rho=0.54, 0.50, 0.48; p<0.0001).
- SHR Admissions is negatively correlated in each of the four years with percent of patients in the facility with AV Fistula (Spearman's rho= -0.12, -0.15, -0.12, -0.13). Thus higher values of SHR are associated with lower usage of AV Fistulas.
- SHR admissions is positively correlated in each of the four years with percent of patients with catheter >= 90 days (Spearman's rho=0.21, 0.21, 0.18, 0.16), indicating that higher values of SHR are associated with increased use of catheters. These correlations are all highly significant (p<0.001).

- For each year of 2010 through 2013, the SHR Admissions is found to be negatively correlated with the percent of hemodialysis patients with Kt/V>=1.2, again in the direction expected (Spearman's rho= -0.11, -0.13, -0.10,-0.11; p<0.0001). Lower SHRs are associated with a higher percentage of patients receiving adequate dialysis dose.
- The SHR correlates with outcomes, processes of care, and causes of hospitalization that are commonly thought to be potentially related to poor quality of care. Higher hospitalization was associated with higher facility mortality rates; and similarly with higher readmissions. The developer found higher values of SHR are associated with lower usage of AV Fistulas, higher catheter use, and suboptimal dialysis adequacy.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *The validity testing provided by the developer has been updated but results have not significantly changed. Does the Committee agree there is no need for repeat discussion and vote on Validity?*

2b3-2b7. Threats to Validity

2b3. Exclusions: No exclusions.

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary:

The developer provided the following information:

- The risk adjustment is based on a Cox or relative risk model. The adjustment is made for patient age, sex, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, a set of prevalent comorbidities, and calendar year.
- In this model, covariates are taken to act multiplicatively on the admission rate and the adjustment model is fitted with facility defining strata so as to provide valid estimates even if the distribution of adjustment variables differs across facilities. All analyses are done using SAS.
- In general, adjustment factors for the SHR were selected based on several considerations:
- The developer began with a large set of patient characteristics, including demographics, comorbidities at ESRD incidence, a set of prevalent comorbidities, and other characteristics.
- Factors considered appropriate were then investigated with statistical models, including interactions between sets of adjusters, to determine if they were related to hospitalizations.
- Factors related to the SHR were also evaluated for face validity before being included.
- Finally, SDS/SES factors were evaluated based on appropriateness (whether related to disparities in care), empirical association with the outcome, and as supported in published literature.
- In 2007, a Technical Expert Panel was convened; the TEP provided advice on various aspects of the SHR, including adjustment factors.
- The 2007 Hospitalization TEP felt that facility characteristics are generally not appropriate for use as adjusters, but should be evaluated for their potential as proxies for patient characteristics. They also recommended that facility market characteristics, such as local hospital utilization rates, should not be considered as risk adjusters.
- In response to great interest among dialysis care providers and other stakeholders in adjusting for more current (prevalent) comorbidities to reflect the current health status of dialysis patients, and specifically inclusion of conditions associated with hospitalization, CMS contracted with UM-KECC to convene a Technical Expert Panel (TEP) in September 2015 to consider the addition of prevalent comorbidity risk adjustment. The TEP evaluated a list of prevalent comorbidities derived through the following process:
    - The ESRD Hierarchical Condition Categories (ESRD-HCCs) were used as a starting point to identify ICD-9 diagnosis codes related to dialysis care.
    - Those individual ICD-9 conditions that comprised the respective ESRD HCCs, with a prevalence of at least 0.1% in the patient population, were then selected for analysis to determine their statistical relationship to

mortality and/or hospitalization. This step resulted in 555 diagnoses comorbidities (out of over 3000 ICD-9 diagnosis codes in the ESRD-HCCs).
    - Next, an adaptive lasso variable selection method was applied to these 555 diagnoses to identify those with a statistically significant relationship to mortality and/or hospitalization ($p<0.05$). This process identified 242 diagnoses.
    - The TEP members then scored each of these diagnoses. This scoring exercise aimed at identifying a set of prevalent comorbidities not likely the result of facility care and therefore potentially appropriate as risk adjusters for SHR and SMR. The TEP established that comorbidities scored as "unlikely" or "very unlikely the result of facility care" by at least half of TEP members (simple majority) were judged as appropriate for inclusion as risk-adjusters. This process resulted in 210 conditions as risk adjustors.
    - The TEP further recommended that:
        - Comorbidities for inclusion as risk-adjusters in a particular year should be present in Medicare claims in the preceding calendar year
        - Determination of a prevalent comorbidity required at least two outpatient claims or one inpatient claim. The set of prevalent comorbidities recommended by the TEP for inclusion as risk-adjusters is presented listed below.
- Consideration of SDS/SES risk factors:
    - The likelihood of hospitalization is related to socioeconomic disadvantage through differences in health status, insurance coverage, and access to quality primary care.
    - Individual and market or area-level measures of deprivation have been shown to contribute independently to preventable hospitalizations.
    - Health care outcomes and utilization are associated with area-level income and residential segregation, but particularly so for racial minorities. This suggests the interplay of patient level (race) and area level SES factors related to lower income, neighborhood poverty, segregation, levels of educational attainment, and unemployment levels that jointly influence key health outcomes related to morbidity.
    - Within the dialysis population area-level SES are associated with poor outcomes; while patient level factors such as race are predictive of differences in certain clinical outcomes by race.
    - Insurance status is also related to health outcomes but this has not been studied extensively within the dialysis population as it relates to hospitalization, though the association has been documented in studies of the general dual Medicare and Medicaid population. Dual eligibles typically have greater comorbidity burden, face access to care barriers which in turn drive higher hospital utilization.
    - Maintaining employment is a challenge for dialysis patients which in turn can influence well-being and may have a proximal impact on outcomes such as hospitalization (Curtin et al, AJKD 1996).
- Table 2a demonstrates model coefficients and Table 2b demonstrates prevalent comorbidity coefficients for data years 2010–2013. Most of the coefficient estimates for the prevalent comorbidities are positive and statistically significant, but several do not obtain statistical significance.
- Table 3a and Table 3b show the parameter estimates for patient- and area-level SDS/SES variables based on a Cox model for hospital admissions that included these variables along with the original covariates adjusted for in SHR.
- Evaluating Adjustments for SDS/SES:
    - Figure 1 displays the comparison of SHRs adjusted and not adjusted for race by facility percentage of black patients (deciles) for 2013.
- Patient-level SDS:
    - Compared with males, females were more likely to experience a hospital admission (OR=1.06; $p<0.01$).However the interaction of female sex and age demonstrated the highest odds were observed in the age 15 – 24, 25-44, and 45-59 age groups, with a decreasing gradient, and the 45-59 age group showing the most diminished impact. There was no significant difference in the oldest female-age-specific group. These results suggest the possibility of an unidentified biologic effect or, alternatively, confounding by an unmeasured association for younger females. Figure 3 demonelative effects of coefficients related to sex in the 2013 SHR model.
    - Hispanics were less likely to be admitted to the hospital (OR=0.92; $p<0.01$) than non- Hispanics.
    - Compared with white patients, Asian/PI (OR=0.81, $p<0.01$), Native American (OR=0.97, $p<0.01$) and black (OR=0.94, $p<0.01$) patients were less likely to be admitted to the hospital.
- Patient-level SES:
    - Compared with Medicare-only patients, patients with both Medicare and Medicaid (OR=1.08; $p<0.01$) and

patients with Medicare as secondary/Medicare HMO (OR=2.66, p<0.01) were more likely to be hospitalized.
    - Patients who were employed prior to ESRD incidence were more likely to be admitted to the hospital (OR=1.05; p<0.01) than unemployed patients.
- Area-level SES:
    - Overall, measures of area-level deprivation had very low impact on the odds of hospitalization.
    - Among statistically significant impacts were measures of low median family income (OR=0.998, p=0.0188), the percentage of families below the poverty level (OR=1.001, p=0.002), the percentage of individuals without a high school diploma (OR=1.002, p<0.01), and the area-level unemployment rate (OR=1.002, p<0.01). In general the magnitude of the effects of the individual indicators was very small. In addition to the very small coefficients, a few were not in the expected direction suggesting potential collinearity with other SES or SDS factors in the model.
- After adjustment for SDS/SES, 278 facilities (4.5%) changed performance categories. 105 (1.7%) facilities were down-graded, and 173 (2.8%) were upgraded.
- These analyses indicate that select patient-level variables for SDS/SES affect expected hospitalization rates, while area-level indicators had either minimal or no effect on expected hospital admissions. Furthermore, SHRs with and without adjustment for SDS/SES are highly correlated (0.9109) but adjustment for SDS/SES shifts facility performance only slightly. This suggests SDS/SES does not contribute much to the flagging profiles for facility performance.
- In the final SHR model, the developer continues to include sex for risk adjustment. Race, ethnicity and patient level SES factors are not included in the final risk adjusted model. While adjustment for these factors would account for different outcomes by race and ethnicity and SES factors and guard against barriers in access to care, adjustment would also introduce the potential unintended consequence of allowing access to lower quality of care. Given the very small impact of area-level SES factors, the developer decided not to include these as risk adjustments in the final model.
- The developer is currently examining other measures of SES and SDS to assess impact on expected hospitalization and whether it would be appropriate to adjust for these factors.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
    - *Is there a conceptual relationship between the SDS factor(s) and the measure focus?*
    - *Does empirical analysis (as provided by the measure developer) show that the SDS factor(s) has a significant and unique effect on the outcome in question?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's decision, based on their analysis, to not include SES factors(race, ethnicity and patient level factors) in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To adjust for over-dispersion of the data, the developer computed the p-value for their estimates using the empirical null distribution, a approach that takes account of the natural random variation among facilities that is not accounted for in the model. Their algorithm consists of the following steps:
    - Fit an over-dispersed Poisson model for the number of hospital admissions
    - Then obtain a z-score for each facility by dividing the natural log of its SHR by the standard error from the general linear model described above.
    - These z-scores are then grouped into quartiles based on the number of patient years at risk for Medicare patients in each facility.
    - Using estimates of location and scale based on the normal curve fitted to the center of the z-scores for the SHR, the developer derives the mean and variance of a normal empirical null distribution for each quartile. This empirical null distribution is then used to calculate the p-value for a facility's SHR.
- Overall, most facilities are flagged as expected (94.03%), while approximately 1% are better than expected, and approximately 5% are flagged as worse than expected.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Can stakeholders make judgements about quality of care when 94% of cafilities are "as expected"?*

2b6. Comparability of data sources/methods: Not needed.

2b7. Missing Data: An analysis of missing data analysis was not provided on this measure.

Preliminary rating for validity: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Guidance from Validity Algorithm:

Specifications consistent with evidence (Box 1)→Potential threats to validity assessed (Box 2) →Empirical validity testing of measure as specified (Box 3) →Testing performed with measure score (Box 6) → Method described and appropriate (Box 7) →Level of certainty or confidence that measure score is a valid indicator of quality (Box 8) →Moderate

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: None

2a2. Reliability Testing

Comments:

**Agree that the reliability as measured by developed is high. Method used is reasonable.

I agree there is no need for repeat discussion and would vote on reliability

Preliminary ranking: High

**Reliability was tested with an adequate scope. The developer assessed the degree to which the Standardized Hospital Ratio for Admissions was consistent year to year using data on hospitalizations among ESRD patients. The developer stated the measure is based on complete data and is not subject to judgment or rater variability. Hence the measures of inter-rater variability are not relevant here.

**1. The USRDS produced by the University of Michigan KECC has demonstrated a reduction in ESRD related hospitalization overall. However the SHR was demonstrated by the developer as not changing for facilities stratified by size over time. It would be helpful for the developer to explain that.

2. It is not clear why the lookback varies in terms of number of years between this measure and similar measures like the SMR measure. What is the rationale for that?

3.

**Reliability testing looks good

**Yes

**No concerns

**overall IUR in 0.7 range but for small facilities 0.46-0.54 so far less robust for this group and not sure it can be considered reliable for smaller units

2b2. Validity Testing

Comments:

**Validity testing is reasonable and high degree of validity demonstrated.

No need for discussion on validity

Preliminary ranking: High

**Validity of the Standardized Hospital Ratio for Admissions was assessed using data on hospitalizations as well as other quality measures among ESRD patients over a three year period of 2006-2008 by examining its covariability with other measures of quality as well as by examining the relationship of the overall hospitalization measure with measures that were more directly focused on specific causes. Validity was tested with an adequate scope with results indicating correlation to URR, AV fistula rate, standardized mortality rate

**1. As with the SMR, the TEP convened was asked to narrowly opine on the comorbidities used for adjustment, and was not privy to seeing the actual model or other details. Therefore face validity for the measure from the TEP is lacking.

2. Given that the SMR is correlated with the SHR and the SHR is correlated with the SMR and measure follow a similar statistical adjustment, what are the implications of the using these for validation if they are so closely related in methodology?
3. If the denominator is indeed the national average, given that hospitalization rates have geographic variances, then there should be a test of SDS differences by geography
**Validity: Results correlate well with other outcomes in the expected direction
**yes. Validated and re-validated
**No concerns
**SHR analyzed in terms of correlation to other measures of potential quality (use of AVF, adequate URR, etc) -- statistics are significant but many of the correlation coefficients are not that robust

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:

**Risk adjustment reasonable
SES consideration: SES clearly IS a factor determining SHR at the patient level.
At the "area" level, there is no clear evidence that SES changes SHR - - so I agree we should not include SES as risk adjusters to the facility-level SHR.
Stakeholders can make judgments about care quality when "outliers" represent 6% of facilities... but the utility fo such identification is limited.
Preliminary rating: Moderate
**The risk adjustment is based on a Cox or relative risk model. The adjustment is made for patient age, sex, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, a set of prevalent comorbidities, and calendar year. In general, adjustment factors for the SHR were selected based on several considerations. The developer began with a large set of patient characteristics, including demographics, comorbidities at ESRDincidence, a set of prevalent comorbidities, and other characteristics. In 2007, a Technical Expert Panel was convened; the TEP provided advice on various aspects of the SHR, including adjustment factors. It was found that select patient-level variables for SDS/SES affect expected hospitalization rates, while area-level indicators had either minimal or no effect on expected hospital admissions. Furthermore, SHRs with and without adjustment for SDS/SES are highly correlated (0.9109) but adjustment for SDS/SES shifts facility performance only slightly. Given the very small impact of area-level SES factors, the developer decided not to include these as risk adjustments in the final model.
**1. Since this measure for hospitalization includes repeat hospitalization and CMS has already instituted a standardized readmission ratio this measure has the potential to penalize dialysis units twice. CMS should implement one not both of these metrics. In addition there has been a debate around the SRR regarding the lag period following hospitalization that should be excluded as not under the dialysis facilities control. This has not been addressed for this measure.
2. This measure uses the 2728 data which has been demonstrated to have validity issues.
3. It is not clear by the look back period for the claims used varies from those used in other measures such as SMR
**Detailed analysis results with and without various SDS/SES. In the limited time available for review, I could not say what the result of all that means, but it is an impressive bode of work.
Meaningful differences - there are only a small fraction judged better or worse than expected. Despite pouring over the many pages, I did not find the definition of Better or Worse than Expected.
**Appropriate adjustments
**No concerns
**94% perform as expected -- 1% better and 5% worse -- not sure if difference meaningful given that so many do OK

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the 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.

- All data elements are in defined fields in electronic form and generated or collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- o *Are the required data elements routinely generated and used during care delivery?*
- o *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- o *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: None

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details:

- This measure is publically reported nationally in Dialysis Facility Compare (DFC).

Improvement results:

The developer states hospitalization rates have decreased over time as evidenced by the coefficients for calendar year from the SHR model. The hospitalization rate for 2011 decreased by 3% compared to 2010 (p-value <0.0001). Subsequent years had a larger decrease in the hospitalization rate compared to 2010 at 12.7% lower for 2012 and about 16.2% lower for 2013 (p-value<0.0001 for both).

Year	Coefficient	P-value
2011	-0.03	<0.0001
2012	-0.127	<0.0001
2013	-0.162	<0.0001

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms:

In the past, a concern has been raised about patient selection relating to ensuring access to care for sicker patients.

The SHR measure incorporates a risk adjustment methodology that levels the playing field for facilities with different patient case-mixes in order to dis-incentivize patient cherry-picking. Given the adjustments for patients' prevalent comorbidities, which reflect a substantial modification to SHR, we think this additional risk adjustment strategy would discourage avoidance of treating sicker and more complex patients that are more likely to experience a hospital admission.

Feedback : No feedback provided on QPS. During the 2015-2016 MAP review, MAP supported updates to this measure for inclusion in the End-Stage Renal Disease Quality Reporting Incentive program, with the condition that NQF review and endorse the measure updates and examine SDS factors as part of the review.

***Questions for the Committee**:*

- o *Would the measure results be more usable to stakeholders if expressed as a rate rather than a ratio?*
- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments:

**Usability is less clear - - - with < 5% population falling into "greater than expected" hospitalizations, with the overall absence of change in hospitalization over years... this metric may have very limited use.

Preliminary rating: Low

**This measure is publically reported nationally in Dialysis Facility Compare (DFC).

The SHR measure incorporates a risk adjustment methodology that levels the playing field for facilities with different patient case-mixes in order to disincentivize patient cherry-picking. Given the adjustments for patients' prevalent comorbidities, which reflect a substantial modification to SHR, this additional risk adjustment strategy would discourage avoidance of treating sicker and more complex patients that are more likely to experience a hospital admission.

**1. This measure should be reported as a rate.

2. Given that dialysis units do not have access to the data required to calculate the statistically adjusted measure, cycle times for continuous improvement are long. CMS should provide this data using the six month lagged claims file on a monthly basis.

**Used now for public reporting – they contend this makes the duialysis facilities accountable Not sure I agree.

They make a good case that they have mitigated the unintended consequence of cherry-picking patients.

**reported nationally in DFC.

**Currently used. Data shows a decrease in rate of hospitalization over time. Good measure of quality improvement efforts.

**used in DFC

Criterion 5: Related and Competing Measures

Related or competing measures

- o 0369 : Standardized Mortality Ratio for Dialysis Facilities
- o 2496 : Standardized Readmission Ratio (SRR) for dialysis facilities

Harmonization

- o The developer states that these measures are not completely harmonized. Each measure assesses different outcomes as reflected in certain differences across the measure specifications. SHR, SMR and SRR are harmonized to the population they measure (Medicare-covered ESRD patients), methods (SMR and SHR) and certain risk adjustment factors specific to the ESRD population. SHR and SMR adjust for all the same comorbidity

risk factors, a similar set of patient characteristics, and use fixed effects in their modeling approach. The differences between SHR, SMR and SRR reflect adjustment for factors specific to the outcome of each respective measure. Both SHR and SMR adjust for a set of prevalent comorbidities (observed in a prior year), however the complete set of comorbidities differs for SRR. SRR excludes planned readmissions; and adjusts for discharging hospital, acknowledging that for readmission, hospitals also bear accountability for properly coordinating care with the dialysis facility. These risk adjustments in SRR account for those characteristics specifically associated with readmission, and do not apply to SHR or SMR. SHR adjusts for sex to account for sex-age specific effects associated with higher hospitalization. Only SMR adjusts for state death rates, race, and ethnicity to account for these respective differences related to mortality outcomes and that are deemed outside of a facility's control.

Pre-meeting public and member comments

Comment by Daniel E. Weiner, MD, MS
Organization Dialysis Clinic, Inc. (DCI)
Comment #5671: I appreciate the opportunity to comment on NQF 0369 and NQF 1463, the Risk-Adjusted SMR and SHR. These are important outcome measures and the use of risk adjustment for comorbid conditions based on claims data is an important advance. The adjustment methodology has important validity issues.

Model selection needs to incorporate background knowledge about the relationship of a variable to the outcome of interest. Unfortunately, adjustment for prevalent comorbid conditions proposed in these metrics relied almost entirely on automatic variable selection techniques, resulting in a model that may be robust only for the data on which it was generated and that will rapidly lose validity as coders and codes change. In defending the methodology, the developer stated that the TEP agreed with the inclusion of this set of prevalent comorbidities. In discussing with TEP members, this is misleading, with members noting the same concerns as raised in this letter about the final models.

Examples include:

1. Cancer is good. The constellation of prostate, thyroid, and kidney cancer together has twice the protective effect against death that gangrene has for harm. This of course is ridiculous; however, the coefficients generated for these comorbid coefficients reflect multicollinearity among variables; coding habits; survival, indication and lead time biases; and, critically, lack of incorporation of existing knowledge into the predictive modeling approach.

2. Peripheral vascular disease codes for important conditions like gangrene, ulcers and osteomyelitis are messy, with numerous positive and negative coefficients that are likely to deviate from the truth with each passing year as coding habits change, providing a classic example of the pitfalls of multicollinearity in predictive models.

3. Codes for diabetic eye disease are highly protective. Why? Likely because these codes indicate that a dialysis patient has seen an ophthalmologist, which is likely an indicator of care coordination. Inclusion of these 3 variables will harm ESCOs for example, where an eye exam is a process measure. This makes no physiologic sense.

The examples above illustrate where, although statistically correct at the time of model development, the adjustment process is destined to lose robustness rapidly with time.

In evaluating these proposed measures, I hope NQF calls attention to the details of the adjustment model and suggests a refined approach moving forward that incorporates both the advanced statistical techniques that were used in the proposed model along with existing knowledge on the relationships of clinical conditions with outcomes and awareness of the biases inherent in the use of these administrative data to develop future comorbidity adjustment models that will remain robust for their intended purpose.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5683: KCP believes hospitalization is an important outcome to measure, but has concerns about the specifications, reliability, validity (risk model), and harmonization issues. Many of our comments have been articulated in the context of those we make on the SMR, but owing to the NQF's electronic portal for measure-by-measure comments, we repeat them for the SHR.

SPECIFICATIONS. KCP has strongly advocated for the use of prevalent co-morbidities in the SHR's risk model, and commends the developer for moving to incorporate prevalent co-morbidities in the specifications. We continue to be concerned about the validity of the Medical Evidence Form (CMS 2728) as a data source for incident co-morbidities, however, and urge that the Committee recommend that CMS assess this matter.

The SHR specifications for the time period also state "at least one year." Again, as a principle, KCP believes specifications should be unambiguous. We believe the time period should be an exact period.

As we discuss in the reliability section, KCP has significant concerns about the reliability of the 1-year SHR for small and medium facilities. The SHR specifications do not address a minimum sample size by excluding facilities of "x" or fewer patients, as we are aware other measures do.

Documentation indicates the minimum data requirement for the SHR is 5 patient-years at risk, which differs from the STrR, which uses 10 patient-years at risk. No justification or empirical analyses are offered to justify the selected threshold or the difference.

Finally, the SHR specifications indicate the measure can be expressed as a rate, but is calculated as a ratio. KCP prefers normalized rates or year-over-year improvement in rates instead of a standardized ratio. We believe comprehension, transparency, and utility to all stakeholders is superior with a scientifically valid rate methodology.

RELIABILITY. We again note a reliability statistic of 0.70 is often considered as "good" reliability, though we recognize the characterization also depends on the analytic method. Again, based on the results from the reliability testing, we have significant concerns about the reliability of the 1-year SHR for small and medium facilities (IUR range of 0.46-0.65, depending on the year). The SHR specifications do not address a minimum sample size by excluding facilities of "x" or fewer patients, as we are aware other measures do. As noted earlier, KCP thus believes the specifications must specifically require a minimum sample as identified through the developer's empirical testing.

(cont.)

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5684: VALIDITY. KCP has strongly advocated for the use of prevalent co-morbidities in the SHR's risk model, and commends the developer for moving to incorporate prevalent co-morbidities in the specifications. We continue to be concerned about the validity of the 2728 as a data source for incident co-morbidities, however, and urge that the Committee recommend that CMS assess this matter.
In previous comments to CMS, KCP noted that many of the prevalent co-morbidities in the final model had p-values significantly greater than 0.05—e.g., paralytic ileus (p=0.5007), episodic mood disorder NOS (p=0.8254). CMS responded that these were included because: "Most of the coefficient estimates for the

prevalent co-morbidities are positive and statistically significant, but several do not obtain statistical significance. The very large number of clinical factors in the model expectedly generates multi-collinearity among co-variates, likely resulting in some unexpected results in direction of coefficient sign and levels of statistical significance. Inclusion of this set of prevalent co-morbidities reflects the consensus of the TEP [Technical Expert Panel] that adjustment for all of these prevalent co-morbidities, in addition to incident co-morbidities, is important to reflect the initial and current health condition of the patient in risk adjustment."

We do not believe this approach is sufficient. Our conversations with TEP members indicate they did not advocate for model building in a vacuum without accounting for the meaning of the coded co-morbid conditions, but rather for including as many co-morbid conditions as possible. This is a very different interpretation than is offered by the developer's explanation and more appropriate when dealing with administrative coding habits that are not static over time. It may require, for example, grouping certain individual codes together to develop an appropriate overarching description of true co-morbidity burden.

(cont.)

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5685: VALIDITY (cont.). KCP is concerned the strategy adopted for the SHR (and SMR) results in a model that will not be generalizable. Currently, for example, having thyroid cancer is protective to the same magnitude that atrial fibrillation is harmful. This makes no sense, and we posit is a function of collinearity and coding idiosyncracy. Similarly, in the current model osteomyelitis NOS-ankle is associated with a lower risk of death, while ulcer of lower limb NOS is harmful. In actual medical practice, osteomyelitis is far worse than an ulcer of the lower limb. In the current model, lower extremity amputation is protective while 'status amput below knee' is harmful. Again, KCP supports prevalent co-morbidity adjustment, but we are concerned that the proposed collection of adjusters will be less robust with each year that passes from initial model development.

KCP also notes that the validity testing yielded an overall c-statistic for the SHR of 0.65. We are concerned the model will not adequately discriminate performance—particularly that smaller units might look worse than reality. We believe a minimum c-statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

Information on the risk model states that determination of a prevalent co-morbidity requires at least two outpatient claims or one inpatient claim, but no justification or empirical analyses are offered to support this algorithm over other approaches. As noted for the SMR, we are aware this approach has been validated for diabetes, but we are not that it has been validated for the large number of other co-morbidities or is broadly generalizable.

Finally, the risk model includes ambiguous language. The submission indicates patient characteristics included in the stage 1 model include "nursing home status in previous year." It is unclear if this means patients moving into a nursing home for the first time during the measurement year would not be adjusted for "nursing home status." Specifically, it is unclear as to whether the look-back is one year prior to the given event (inclusive of the data year) or if this verbiage means the look-back is in the previous calendar year (not inclusive of the data year). KCP believes such ambiguity should be addressed and that the current reporting year be included, not just the previous one.

(cont.)

Comment by Lisa McGonigal, MD, MPH

Organization Kidney Care Partners

Comment #5686: HARMONIZATION ISSUES. The risk models for the groupings used for patient age and duration of ESRD differ among the SMR, SHR, and STrR. For example, the age groups for the SMR is n=3, but for the SHR and STrR the age groupings are the same, but n=6. Similarly, the number of groups for ESRD duration for the SMR (n=4) differs from that for the SHR (n=6). No justification or empirical analyses are offered to justify these differences.

There also are significant inconsistencies in how facility size is defined when assessing reliability for the SMR, SHR, and STrR. Specifically, for the SMR, the definitions were <=45, 46-85, >=86 for the 1-year reliability analyses, but were <=135, 136-305, and >=306 for the 4-year analyses. For the SHR, <=50, 51-87, and >=88 were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as <=46, 47-78, and >=79, respectively. We understand reliability for a given measure depends on sample size, but find the varying demarcations analytically troubling. We posit a more appropriate analytic approach would be to analyze reliability using consistent "bins" of size (i.e., small, medium, and large are consistently defined) and identify the facility size at which reliability for that particular measure can be confidently inferred—and then reflect the minimum size in the actual specifications.

Comment by Joseph Vassalotti

Organization National Kidney Foundation

Comment #5697: The National Kidney Foundation supports this measure and agrees with the MAP recommendations. However, as we state in our comments on the Standardized Readmissions Ratio, we believe this measure should be stratified by cause of admission to include those that are actionable by the dialysis care team. The admissions that may be most actionable include congestive heart failure, volume overload and vascular access blood stream infections. Stratification would give clinicians information for quality improvement focus and help patients better assess the quality of care they can expect to receive in the dialysis unit. An alternative approach to stratification by cause of admission would be to exclude admissions that are not likely to be actionable by dialysis care.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1463
Measure Title: Standardized Hospitalization 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: 4/15/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
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

☒ Health outcome: Hospitalization

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

2011 Submission

Hospitalization rates are an important indicator of patient morbidity and quality of life. On average, dialysis patients are admitted to the hospital twice a year and hospitalizations account for approximately 36 percent of total Medicare expenditures for dialysis patients (U.S. Renal Data System, 2007). Measures of the frequency of hospitalization help efforts to control escalating medical costs, and play an important role in providing cost-effective health care.

2016 Submission

There are numerous dialysis care processes that can influence the likelihood of a patient's being hospitalized. These processes include:

(1) Inadequate processes related to fluid management/removal-. Inadequate control of total body fluid balance and fluid removal can result in fluid overload and congestive heart failure, increasing the possibility of the need for hospitalization.

(2) Inadequate infection prevention. Inadequate infection prevention processes, including suboptimal management of vascular access, can lead to bacteremia or septicemia, increasing the possibility of the need for hospitalization.

(3) Inadequate dialysis. Failure to maintain processes to ensure adequate dialysis can lead to low Kt/v, increasing the possibility of the need for hospitalization.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

2011 Submission

This was not a question on the 2011 Submission Form.

2016 Submission

Hospitalization rates remain very high in US chronic dialysis patients relative to the general population, despite a nearly 20% decline from 2005-2013. This trend in lower hospitalization is in contrast to the relatively stable hospitalization rates for the US general population over the same time period, suggesting that dialysis providers have been somewhat successful in reducing unnecessary hospitalizations through quality of care improvements.

According to the 2015 USRDS Annual Report, approximately ½ of all dialysis patient hospitalizations continue to be caused by cardiovascular or infectious causes over that time period [1]. Recent research points to many additional opportunities to further reduce unnecessary hospitalization in this population.

Programs developed to impact dialysis provider practices have been shown to improve intermediate outcomes (reduced catheter vascular access, small solute adequacy, anemia management) and mortality, modality options, infection prevention, and dialysis organization culture [2-19]. These practice improvements have been linked to reduced hospitalizations in this population. For example, one study examined dialysis provider interventions targeting incident patients in order to improve outcomes for these patients that are at particularly high risk for poor outcomes that can lead to higher morbidity and mortality [2]. The results suggested improved clinical outcomes in terms of the percentage of incident patients having a preferred vascular access type. In turn this has the potential to reduce hospitalization risk, along with mortality; other work on vascular access type also supports the link between access type and hospitalization, specifically due to chronic catheter use [3].

1] United States Renal Data System. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015.

[2] Wilson SM, Robertson JA, Chen G, Goel P, Benner DA, Krishnan M, Mayne TJ, Nissenson AR. The IMPACT (Incident Management of Patients, Actions Centered on Treatment) Program: A Quality Improvement Approach for Caring for Patients Initiating Long-term Hemodialysis. Am J Kidney Dis 60(3): 435-443, 2012

BACKGROUND: Patients beginning dialysis therapy are at risk of death and illness. The IMPACT (Incident Management of Patients, Actions Centered on Treatment) quality improvement program was developed to improve incident hemodialysis patient outcomes through standardized care.

STUDY DESIGN: Quality improvement report.

SETTING & PARTICIPANTS: Patients who started hemodialysis therapy between September 2007 and December 2008 at DaVita facilities using the IMPACT program (n = 1,212) constituted the intervention group. Propensity score-matched patients who initiated hemodialysis therapy in the same interval at DaVita facilities not using the IMPACT program (n = 2,424) made up the control group.

QUALITY IMPROVEMENT PLAN: IMPACT intervention included a structured intake process and monitoring reports; patient enrollment in a 90-day patient education program and 90-day patient management pathway.

OUTCOMES: Mean dialysis adequacy (Kt/V), hemoglobin and albumin levels, percentage of patients using preferred vascular access (arteriovenous fistula or graft), and mortality at each quarter.

RESULTS: Compared with the non-IMPACT group, the IMPACT group was associated with a higher proportion of patients dialyzing with a preferred access at 90 days (0.50 [95% CI, 0.47-0.53] vs 0.47 [95% CI, 0.45-0.49]; P = 0.1) and 360 days (0.63 [95% CI, 0.61-0.66] vs 0.48 [95% CI, 0.46-0.50]; P < 0.001) and a lower mortality rate at 90 days (24.8 [95% CI, 19.0-30.7] vs 31.9 [95% CI, 27.1-36.6] deaths/100 patient-years; P = 0.08) and 360 days (17.8 [95% CI, 15.2-20.4] vs 25.1 [95% CI, 20.7-25.2] deaths/100 patient-years; P = 0.01).

LIMITATIONS: The study does not determine the care processes responsible for the improved outcomes.
CONCLUSIONS: Intense management of incident dialysis patients with the IMPACT quality improvement program was associated with significantly decreased first-year mortality. Focused attention to the care of incident patients is an important part of a dialysis program.

[3] Vassalotti JA, Jennings WC, Beathard GA, Neumann M, Caponi S, Fox CH, Spergel LM and the Fistula First Breakthrough Initiative Community Education Committee. Fistula First Breakthrough Initiative: Targeting Catheter Last in Fistula First. Seminars Dialysis 25(3):303-310, 2012

Abstract: An arteriovenous fistula (AVF) is the optimal vascular access for hemodialysis (HD), because it is associated with prolonged survival, fewer infections, lower hospitalization rates, and reduced costs. The AVF First breakthrough initiative (FFBI) has made dramatic progress, effectively promoting the increase in the national AVF prevalence since the program's inception from 32% in May 2003 to nearly 60% in 2011. Central venous catheter (CVC) use has stabilized and recently decreased slightly for prevalent patients (treated more than 90 days), while CVC usage in the first 90 days remains unacceptably high at nearly 80%. This high prevalence of CVC utilization suggests important specific improvement goals for FFBI. In addition to the current 66% AVF goal, the initiative should include specific CVC usage target(s), based on the KDOQI goal of less than 10% in patients undergoing HD for more than 90 days, and a substantially improved initial target from the current CVC proportion. These specific CVC targets would be disseminated through the ESRD networks to individual dialysis facilities, further emphasizing CVC avoidance in the transition from advanced CKD to chronic kidney failure, while continuing to decrease CVC by prompt conversion of CVC-based hemodialysis patients to permanent vascular access, utilizing an AVF whenever feasible.

[4] Ng LJ, Chen F, Pisoni RL, Krishnan M, Mapes D, Keen M, Bradbury BD. Hospitalization risks related to vascular access type among incident US hemodialysis patients. Nephrol Dial Transplant. 26(11):3659-66, 2011

BACKGROUND: The excess morbidity and mortality related to catheter utilization at and immediately following dialysis initiation may simply be a proxy for poor prognosis. We examined hospitalization burden related to vascular access (VA) type among incident patients who received some predialysis care.

METHODS: We identified a random sample of incident US Dialysis Outcomes and Practice Patterns Study hemodialysis patients (1996-2004) who reported predialysis nephrologist care. VA utilization was assessed at baseline and throughout the first 6 months on dialysis. Poisson regression was used to estimate the risk of all-cause and cause-specific hospitalizations during the first 6 months.

RESULTS: Among 2635 incident patients, 60% were dialyzing with a catheter, 22% with a graft and 18% with a fistula at baseline. Compared to fistulae, baseline catheter use was associated with an increased risk of all-cause hospitalization [adjusted relative risk (RR) = 1.30, 95% confidence interval (CI): 1.09-1.54] and graft use was not (RR = 1.07, 95% CI: 0.89-1.28). Allowing for VA changes over time, the risk of catheter versus fistula use was more pronounced (RR = 1.72, 95% CI: 1.42-2.08) and increased slightly for graft use (RR = 1.15, 95% CI: 0.94-1.41). Baseline catheter use was most strongly related to infection-related (RR = 1.47, 95% CI: 0.92-2.36) and VA-related hospitalizations (RR = 1.49, 95% CI: 1.06-2.11). These effects were further strengthened when VA use was allowed to vary over time (RR = 2.31, 95% CI: 1.48-3.61 and RR = 3.10, 95% CI: 1.95-4.91, respectively). A similar pattern was noted for VA-related hospitalizations with graft use. Discussion. Among potentially healthier incident patients, hospitalization risk, particularly infection and VA-related, was highest for patients dialyzing with a catheter at initiation and throughout follow-up, providing further support to clinical practice recommendations to minimize catheter placement.

[5] Block GA, Kilpatrick RD, Lowe KA, Wang W, Danese MD. CKD-Mineral and Bone Disorder and Risk of Death and Cardiovascular Hospitalization in Patients on Hemodialysis. CJASN 8:2132-2140, 2013.

BACKGROUND AND OBJECTIVES: Parathyroid hormone, calcium, and phosphate have been independently associated with cardiovascular event risk. Because these parameters may be on the same causal pathway and have been proposed as quality measures, an integrated approach to estimating event risks is needed.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Prevalent dialysis patients were followed from August 31, 2005 to December 31, 2006. A two-stage modeling approach was used. First, the 16-month probabilities of death and composite end point of death or cardiovascular hospitalization were estimated and adjusted for potential confounders. Second, patients were categorized into 1 of 36 possible phenotypes using average parathyroid hormone, calcium, and phosphate values over a 4-month baseline period. Associations among

phenotypes and outcomes were estimated and adjusted for the underlying event risk estimated from the first model stage.

RESULTS: Of 26,221 patients, 98.5% of patients were in 22 groups with at least 100 patients and 20% of patients were in the reference group defined using guideline-based reference ranges for parathyroid hormone, calcium, and phosphate. Within the 22 most common phenotypes, 20% of patients were in groups with significantly ($P<0.05$) higher risk of death and 54% of patients were in groups with significantly higher risk of the composite end point relative to the in-target reference group. Increased risks ranged from 15% to 47% for death and from 8% to 55% for the composite. More than 40% of all patients were in the three largest groups with elevated composite end point risk (high parathyroid hormone, target calcium, and high phosphate; target high parathyroid hormone, target calcium, and high phosphate; and target high parathyroid hormone, target calcium, and target phosphate).

CONCLUSION: After adjusting for baseline risk, phenotypes defined by categories of parathyroid hormone, calcium, and phosphate identify patients at higher risk of death and cardiovascular hospitalization. Identifying common high-risk phenotypes may inform clinical interventions and policies related to quality of care.

[6] Pun PH, Horton JR, Middleton JP. Dialysate calcium concentration and the risk of sudden cardiac arrest in hemodialysis patients. CJASN 8:797-803, 2013.

BACKGROUND AND OBJECTIVES: The optimal dialysate calcium concentration to maintain normal mineralization and reduce risk of cardiovascular events in hemodialysis patients is debated. Guidelines suggest that dialysate Ca concentration should be lowered to avoid vascular calcification, but cardiac arrhythmias may be more likely to occur at lower dialysate Ca. Concurrent use of QT-prolonging medications may also exacerbate arrhythmic risk. This study examined the influence of serum Ca, dialysate Ca, and QT interval-prolonging medications on the risk of sudden cardiac arrest in a cohort of hemodialysis patients.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: This case-control study among 43,200 hemodialysis patients occurred between 2002 and 2005; 510 patients who experienced a witnessed sudden cardiac arrest were compared with 1560 matched controls. This study examined covariate-adjusted sudden cardiac arrest risk associations with serum Ca, dialysate Ca, serum dialysate Ca gradient, and prescription of QT-prolonging medications using logistic regression techniques.

RESULTS: Patients assigned to low Ca dialysate<2.5 mEq/L were more likely to be exposed to larger serum dialysate Ca gradient and had a greater fall in BP during dialysis treatment. After accounting for covariates and baseline differences, low Ca dialysate<2.5 mEq/L (odds ratio=2.00, 95% confidence interval=1.40-2.90), higher corrected serum Ca (odds ratio=1.10, 95% confidence interval=1.00-1.30), and increasing serum dialysate Ca gradient (odds ratio=1.40, 95% confidence interval=1.10-1.80) were associated with increased risk of sudden cardiac arrest, whereas there were no significant risk associations with QT-prolonging medications.

CONCLUSIONS: Increased risk of sudden cardiac arrest associated with low Ca dialysate and large serum dialysate Ca gradients should be considered in determining the optimal dialysate Ca prescription.

[7] Ishani A, Liu J, Wetmore JB, Lowe KA, Do T, Bradbury BD, Block GA, Collins AJ. Clinical outcomes after parathyroidectomy in a nationwide cohort of patients on hemodialysis. Clin J Am Soc Nephrol. 10(1):90-7, 2015.

BACKGROUND AND OBJECTIVES: Patients receiving dialysis undergo parathyroidectomy to improve laboratory parameters in resistant hyperparathyroidism with the assumption that clinical outcomes will also improve. However, no randomized clinical trial data demonstrate the benefits of parathyroidectomy. This study aimed to evaluate clinical outcomes up to 1 year after parathyroidectomy in a nationwide sample of patients receiving hemodialysis.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Using data from the US Renal Data System, this study identified prevalent hemodialysis patients aged ≥18 years with Medicare as primary payers who underwent parathyroidectomy from 2007 to 2009. Baseline characteristics and comorbid conditions were assessed in the year preceding parathyroidectomy; clinical events were identified in the year preceding and the year after parathyroidectomy. After parathyroidectomy, patients were censored at death, loss of Medicare coverage, kidney transplant, change in dialysis modality, or 365 days. This study estimated cause-specific event rates for both periods and rate ratios comparing event rates in the postparathyroidectomy versus preparathyroidectomy periods.

RESULTS: Of 4435 patients who underwent parathyroidectomy, 2.0% died during the parathyroidectomy hospitalization and the 30 days after discharge. During the 30 days after discharge, 23.8% of patients were rehospitalized; 29.3% of these patients required intensive care. In the year after parathyroidectomy, hospitalizations were higher by 39%, hospital days by 58%, intensive care unit admissions by 69%, and emergency room/observation visits requiring hypocalcemia treatment by 20-fold compared with the preceding year. Cause-specific hospitalizations were higher for acute myocardial infarction (rate ratio, 1.98; 95% confidence interval, 1.60 to 2.46) and dysrhythmia (rate ratio 1.4; 95% confidence interval1.16 to 1.78); fracture rates did not differ (rate ratio 0.82; 95% confidence interval 0.6 to 1.1).

CONCLUSIONS: Parathyroidectomy is associated with significant morbidity in the 30 days after hospital discharge and in the year after the procedure. Awareness of clinical events will assist in developing evidence-based risk/benefit determinations for the indication for parathyroidectomy.

[8] Tentori F, McCullough K, Kilpatrick RD, Bradbury BD, Robinson BM, Kerr PG, Pisoni RL. High rates of death and hospitalization follow bone fracture among hemodialysis patients. Kidney Int. 85(1):166-73, 2014.

Abstract: Altered bone structure and function contribute to the high rates of fractures in dialysis patients compared to the general population. Fracture events may increase the risk of subsequent adverse clinical outcomes. Here we assessed the incidence of post-fracture morbidity and mortality in an international cohort of 34,579 in-center hemodialysis patients in the Dialysis Outcomes and Practice Patterns Study (DOPPS). We estimated country-specific rates of fractures requiring a hospital admission and associated length of stay in the hospital. Incidence rates of death and of a composite event of death/rehospitalization were estimated for 1 year after fracture. Overall, 3% of participants experienced a fracture. Fracture incidence varied across countries, from 12 events/1000 patient-years (PY) in Japan to 45/1000 PY in Belgium. In all countries, fracture rates were higher in the hemodialysis group compared to those reported for the general population. Median length of stay ranged from 7 to 37 days in the United States and Japan, respectively. In most countries, postfracture mortality rates exceeded 500/1000 PY and death/rehospitalization rates exceeded 1500/1000 PY. Fracture patients had higher unadjusted rates of death (3.7-fold) and death/rehospitalization (4.0-fold) compared to the overall DOPPS population. Mortality and hospitalization rates were highest in the first month after the fracture and declined thereafter. Thus, the high frequency of fractures and increased adverse outcomes following a fracture pose a significant health burden for dialysis patients. Fracture prevention strategies should be identified and applied broadly in nephrology practices.

[9] Weinhandl ED, Arneson TJ, St Peter WL. Clinical outcomes associated with receipt of integrated pharmacy services by hemodialysis patients: a quality improvement report. Am J Kidney Dis. Sep;62(3):557-67, 2013.

Reducing medication-related problems and improving medication adherence in hemodialysis patients may improve clinical outcomes. In 2005, a large US dialysis organization created an integrated pharmacy program for its patients. We aimed to compare the outcomes of hemodialysis patients enrolled in this program and matched control patients.

STUDY DESIGN: Quality improvement report.

SETTING & PARTICIPANTS: Hemodialysis patients with concurrent Medicare and Medicaid eligibility who chose to receive program services and propensity score-matched controls; the propensity score was an estimated function of demographic characteristics, comorbid conditions, medication exposure, serum concentrations, and vascular access method.

QUALITY IMPROVEMENT PLAN: Program services included medication delivery, refill management, medication list reviews, telephonic medication therapy management, and prior authorization assistance.
OUTCOMES: Relative rates of death and hospitalization.

MEASUREMENTS: Survival estimates calculated with the Kaplan-Meier method; mortality hazards compared with Cox regression; hospitalization rates compared with Poisson regression.

RESULTS: In outcome models, there were 8,864 patients receiving integrated pharmacy services and 43,013 matched controls. In intention-to-treat and as-treated analyses, mortality HRs for patients receiving integrated pharmacy services versus matched controls were 0.92 (95% CI, 0.86-0.97) and 0.79 (95% CI, 0.74-0.84), respectively. Corresponding relative rates of hospital admissions were 0.98 (95% CI, 0.95-1.01) and 0.93 (95% CI, 0.90-0.96), respectively, and of hospital days, 0.94 (95% CI, 0.90-0.98) and 0.86 (95% CI, 0.82-0.90), respectively. Cumulative incidences of disenrollment from the pharmacy program were 23.4% at 12 months and 37.0% at 24 months.

LIMITATIONS: Patients were not randomly assigned to receive integrated pharmacy services; as-treated analyses may be biased because of informative censoring by disenrollment from the pharmacy program; data regarding use of integrated pharmacy services were lacking.

CONCLUSIONS: Receipt of integrated pharmacy services was associated with lower rates of death and hospitalization in hemodialysis patients with concurrent Medicare and Medicaid eligibility. Studies are needed to measure pharmacy program use and assess detailed clinical and economic outcomes.

[10]. Weinhandl ED, Gilbertson DT, Collins AJ. Mortality, Hospitalization, and Technique Failure in Daily Home Hemodialysis and Matched Peritoneal Dialysis Patients: A¬†Matched Cohort Study. Am J Kidney Dis. 67(1):98-110, 2016.

BACKGROUND: Use of home dialysis is growing in the United States, but few direct comparisons of major clinical outcomes on daily home hemodialysis (HHD) versus peritoneal dialysis (PD) exist.

STUDY DESIGN: Matched cohort study.

SETTING & PARTICIPANTS: We matched 4,201 new HHD patients in 2007 to 2010 with 4,201 new PD patients from the US Renal Data System database.

PREDICTOR: Daily HHD versus PD.

OUTCOMES: Relative mortality, hospitalization, and technique failure.

RESULTS: Mean time from end-stage renal disease onset to home dialysis therapy initiation was 44.6 months for HHD and 44.3 months for PD patients. In intention-to-treat analysis, HHD was associated with 20% lower risk for all-cause mortality (HR, 0.80; 95% CI, 0.73-0.87), 8% lower risk for all-cause hospitalization (HR, 0.92; 95% CI, 0.89-0.95), and 37% lower risk for technique failure (HR, 0.63; 95% CI, 0.58-0.68), all relative to PD. In the subset of 1,368 patients who initiated home dialysis therapy within 6 months of end-stage renal disease onset, HHD was associated with similar risk for all-cause mortality (HR, 0.95; 95% CI, 0.80-1.13), similar risk for all-cause hospitalization (HR, 0.96; 95% CI, 0.88-1.05), and 30% lower risk for technique failure (HR, 0.70; 95% CI, 0.60-0.82). Regarding hospitalization, risk comparisons favored HHD for cardiovascular disease and dialysis access infection and PD for bloodstream infection.

LIMITATIONS: Matching unlikely to reduce confounding attributable to unmeasured factors, including residual kidney function; lack of data regarding dialysis frequency, duration, and dose in daily HHD patients and frequency and solution in PD patients; diagnosis codes used to classify admissions.

CONCLUSIONS: These data suggest that relative to PD, daily HHD is associated with decreased mortality, hospitalization, and technique failure. However, risks for mortality and hospitalization were similar with these modalities in new dialysis patients. The interaction between modality and end-stage renal disease duration at home dialysis therapy initiation should be investigated further.

[11] Rosenblum A, Wang W, Ball LK, Latham C, Maddux FW, Lacson E. Hemodialysis catheter care strategies: A cluster-randomized quality improvement initiative. Am J Kidney Dis. 63(2):259-267, 2014.

BACKGROUND: The prevalence of central venous catheters (CVCs) for hemodialysis remains high and, despite infection-control protocols, predisposes to bloodstream infections (BSIs).

STUDY DESIGN: Stratified, cluster-randomized, quality improvement initiative.

SETTING & PARTICIPANTS: All in-center patients with a CVC within 211 facility pairs matched by region, facility size, and rate of positive blood cultures (January to March 2011) at Fresenius Medical Care, North America.

QUALITY IMPROVEMENT PLAN: Incorporate the use of 2% chlorhexidine with 70% alcohol swab sticks for exit-site care and 70% alcohol pads to perform "scrub the hubs" in dialysis-related CVC care procedures compared to usual care.

OUTCOME: The primary outcome was positive blood cultures for estimating BSI rates.

MEASUREMENTS: Comparison of 3-month baseline period from April 1 to June 30 and follow-up period from August 1 to October 30, 2011.

RESULTS: Baseline BSI rates were similar (0.85 vs 0.86/1,000 CVC-days), but follow-up rates differed at 0.81/1,000 CVC-days in intervention facilities versus 1.04/1,000 CVC-days in controls ($P = 0.02$). Intravenous antibiotic starts during the follow-up period also were lower, at 2.53/1,000 CVC-days versus 3.15/1,000 CVC-days in controls ($P < 0.001$). Cluster-adjusted Poisson regression confirmed 21%-22% reductions in both ($P < 0.001$). Extended follow-up for 3 successive quarters demonstrated a sustained reduction of bacteremia rates for patients in intervention facilities, at 0.50/1,000 CVC-days (41% reduction; $P < 0.001$). Hospitalizations due to sepsis during 1-year extended follow-up were 0.19/1,000 CVC-days (0.069/CVC-year) versus 0.26/1,000 CVC-days (0.095/CVC-year) in controls (~27% difference; $P < 0.05$).

LIMITATIONS: Inability to capture results from blood cultures sent to external laboratories, underestimation of sepsis-specific hospitalizations, and potential crossover adoption of the intervention protocol in control facilities.

CONCLUSIONS: Adoption of the new catheter care procedure (consistent with Centers for Disease Control and Prevention recommendations) resulted in a 20% lower rate of BSIs and intravenous antibiotic starts, which were sustained over time and associated with a lower rate of hospitalizations due to sepsis.

[12] Patel PR, Kallen AJ. Bloodstream infection prevention in ESRD: Forging a pathway for success. Am J Kidney Dis. 63(2):180-182, 2014

Introduction: There should be little doubt regarding the importance of infections in the hemodialysis patient population. For years, the US Renal Data System has reported increasing hospitalization rates for all infectious diagnoses and for bacteremia/sepsis in patients treated with hemodialysis.[1] In 2011, the Centers for Disease

Control and Prevention (CDC) reported that although the burden of central line–associated bloodstream infections (BSIs) in hospitalized patients had declined nationally, the estimated burden of central line–associated BSIs in people treated with outpatient hemodialysis was substantial, possibly reaching 37,000 in 2008.[2] Soon after, the US Department of Health and Human Services released their National Action Plan to Prevent Healthcare-Associated Infections (HAIs) for End Stage Renal Disease (ESRD) Facilities.[3] The Action Plan, which was developed by the Federal Steering Committee for the Prevention of HAIs in ESRD Facilities with dialysis community stakeholder input, highlighted BSIs as a top priority for national prevention efforts.

[13] Gilbertson DT, Guo H, Arneson TJ, Collins AJ. The association of pneumococcal vaccination with hospitalization and mortality in hemodialysis patients. Nephrol Dial Transplant. Sept;26(9):2934-9, 2011.

BACKGROUND: Few studies have examined the effectiveness of pneumococcal vaccination (alone or with influenza vaccination) in improving hemodialysis patient outcomes. We aimed to describe vaccination rates between 2003-2005 and to study the effects on outcomes.

METHODS: For 118,533 prevalent patients who initiated hemodialysis ≥90 days before 1 November 2003, had Medicare Part A and Part B and were aged ≥18 years, and alive through 31 October 2005, Cox proportional hazards models were used to assess pneumococcal vaccination effects on subsequent hospitalization and mortality, adjusting for demographics and comorbidity.

RESULTS: The 21% of patients who received vaccinations were older; a higher proportion were white, with diabetes as cause of end-stage renal disease and more comorbidity. Pneumococcal vaccination was associated with a statistically significant decreased mortality hazard [hazard ratio (HR) 0.94, 95% confidence interval (CI) 0.90-0.98], cardiac death (HR 0.91, 95% CI 0.85-0.97) and hospitalization for bacteremia/viremia/septicemia (HR 0.95, 95% CI 0.91-1.00). The mortality hazard was 0.73 (95% CI 0.68-0.78) for patients who received pneumococcal and influenza vaccinations.
CONCLUSIONS: The small but significant association between pneumococcal vaccination and lower mortality risk was seen despite factors associated with poor outcomes in patients most likely to be vaccinated. Pneumococcal and influenza vaccines may have beneficial synergistic effects. Hemodialysis patients may benefit from revaccination more frequently than the recommended 5-year intervals.

[14] Dalrymple LS, Mu Y, Nguyen DV, Romano PS, Chertow GM, Grimes B, Kaysen GA, Johansen KL. CJASN 10:2170-2180, 2015.

BACKGROUND AND OBJECTIVES: Infection-related hospitalizations have increased dramatically over the last 10 years in patients receiving in-center hemodialysis. Patient and dialysis facility characteristics associated with the rate of infection-related hospitalization were examined, with consideration of the region of care, rural-urban residence, and socioeconomic status.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: The US Renal Data System linked to the American Community Survey and Rural-Urban Commuting Area codes was used to examine factors associated with hospitalization for infection among Medicare beneficiaries starting in-center hemodialysis between 2005 and 2008. A Poisson mixed effects model was used to examine the associations among patient and dialysis facility characteristics and the rate of infection-related hospitalization.

RESULTS: Among 135,545 Medicare beneficiaries, 38,475 (28%) had at least one infection-related hospitalization. The overall rate of infection-related hospitalization was 40.2 per 100 person-years. Age ≥85 years old, cancer, chronic obstructive pulmonary disease, inability to ambulate or transfer, drug dependence, residence in a care facility, serum albumin <3.5 g/dl at dialysis initiation, and dialysis initiation with an access other than a fistula were associated with a ≥20% increase in the rate of infection-related hospitalization. Patients residing in isolated small rural compared with urban areas had lower rates of hospitalization for infection (rate ratio, 0.91; 95% confidence interval, 0.86 to 0.97), and rates of hospitalization for infection varied

across the ESRD networks. Measures of socioeconomic status (at the zip code level), total facility staffing, and the composition of staff (percentage of nurses) were not associated with the rate of hospitalization for infection.

CONCLUSIONS: Patient and facility factors associated with higher rates of infection-related hospitalization were identified. The findings from this study can be used to identify patients at higher risk for infection and inform the design of infection prevention strategies.

[15] Gilbertson DT, Wetmore JB. Infections Requiring Hospitalization in Patients on Hemodialysis CJASN 10:2101-2103, 2015.

Introduction: Although the past decade has witnessed significant improvements in survival or patients receiving hemodialysis (HD) (1), hospitalization rates, particularly for infection, have not improved commensurately. Notable lack of progress is evident regarding hospitalizations for bacteremia/septicemia and pulmonary infections, such as pneumonia and influenza (2). For bacteremia/septicemia, first–year (incident) admission rates showed a 39% relative increase between 2003 and 2010 from 12.9% to 18.0%. Similarly, admission rates for prevalent patients increased 36% from 8.6% to 11.6%. Pneumonia/influenza hospitalization rates also did not improve between 2003 and 2010; although first–year admission rates decreased slightly (from 10.2% to 9.0%), rates for prevalent patients increased from 8.3% to 9.0%.

[16] Arneson TJ, Liu J, Qiu Y, Gilbertson DT, Foley RN, Collins AJ. Hospital treatment for fluid overload in the Medicare hemodialysis population. Clin J Am Soc Nephrol.(6):1054-63, 2010.

BACKGROUND AND OBJECTIVES: Fluid overload in hemodialysis patients sometimes requires emergent dialysis, but the magnitude of this care has not been characterized. This study aimed to estimate the magnitude of fluid overload treatment episodes for the Medicare hemodialysis population in hospital settings, including emergency departments.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Point-prevalent hemodialysis patients were identified from the Centers for Medicare and Medicaid Renal Management Information System and Standard Analytical Files. Fluid overload treatment episodes were defined by claims for care in inpatient, hospital observation, or emergency department settings with primary discharge diagnoses of fluid overload, heart failure, or pulmonary edema, and dialysis performed on the day of or after admission. Exclusion criteria included stays >5 days. Cost was defined as total Medicare allowable costs for identified episodes. Associations between patient characteristics and episode occurrence and cost were analyzed.

RESULTS: For 25,291 patients (14.3%), 41,699 care episodes occurred over a mean follow-up time of 2 years: 86% inpatient, 9% emergency department, and 5% hospital observation. Heart failure was the primary diagnosis in 83% of episodes, fluid overload in 11%, and pulmonary edema in 6%. Characteristics associated with more frequent events included age <45 years, female sex, African-American race, causes of ESRD other than diabetes, dialysis duration of 1 to 3 years, fewer dialysis sessions per week at baseline, hospitalizations during baseline, and most comorbid conditions. Average cost was $6,372 per episode; total costs were approximately $266 million.

CONCLUSIONS: Among U.S. hemodialysis patients, fluid overload treatment is common and expensive. Further study is necessary to identify prevention opportunities.

[17] Erickson KF, Winkelmayer WC, Chertow GM, Bhattacharya J. Physician visits and 30-day hospital readmissions in patients receiving hemodialysis. J Am Soc Nephrol 25:2079-2087, 2014.

Abstract: A focus of health care reform has been on reducing 30-day hospital readmissions. Patients with ESRD are at high risk for hospital readmission. It is unknown whether more monitoring by outpatient providers can reduce hospital readmissions in patients receiving hemodialysis. In nationally representative cohorts of patients

in the United States receiving in-center hemodialysis between 2004 and 2009, we used a quasi-experimental (instrumental variable) approach to assess the relationship between frequency of visits to patients receiving hemodialysis following hospital discharge and the probability of rehospitalization. We then used a multivariable regression model and published hospitalization data to estimate the cost savings and number of hospitalizations that could be prevented annually with additional provider visits to patients in the month following hospitalization. In the main cohort (n=26,613), one additional provider visit in the month following hospital discharge was estimated to reduce the absolute probability of 30-day hospital readmission by 3.5% (95% confidence interval, 1.6% to 5.3%). The reduction in 30-day hospital readmission ranged from 0.5% to 4.9% in an additional four cohorts tested, depending on population density around facilities, facility profit status, and patient Medicaid eligibility. At current Medicare reimbursement rates, the effort to visit patients one additional time in the month following hospital discharge could lead to 31,370 fewer hospitalizations per year, and $240 million per year saved. In conclusion, more frequent physician visits following hospital discharge are estimated to reduce rehospitalizations in patients undergoing hemodialysis. Incentives for closer outpatient monitoring following hospital discharge could lead to substantial cost savings.

[18] Kliger AS. Maintaining safety in the dialysis facility. CJASN 10:688-695, 2015.

Abstract: Errors in dialysis care can cause harm and death. While dialysis machines are rarely a major cause of morbidity, human factors at the machine interface and suboptimal communication among caregivers are common sources of error. Major causes of potentially reversible adverse outcomes include medication errors, infections, hyperkalemia, access-related errors, and patient falls. Root cause analysis of adverse events and "near misses" can illuminate care processes and show system changes to improve safety. Human factors engineering and simulation exercises have strong potential to define common clinical team purpose, and improve processes of care. Patient observations and their participation in error reduction increase the effectiveness of patient safety efforts.

[19] Nissenson AR. Improving outcomes for ESRD patients: Shifting the quality paradigm. CJASN 9:430-434, 2014.

Abstract: The availability of life-saving dialysis therapy has been one of the great successes of medicine in the past four decades. Over this time period, despite treatment of hundreds of thousands of patients, the overall quality of life for patients with ESRD has not substantially improved. A narrow focus by clinicians and regulators on basic indicators of care, like dialysis adequacy and anemia, has consumed time and resources but not resulted in significantly improved survival; also, frequent hospitalizations and dissatisfaction with the care experience continue to be seen. A new quality paradigm is needed to help guide clinicians, providers, and regulators to ensure that patients' lives are improved by the technically complex and costly therapy that they are receiving. This paradigm can be envisioned as a quality pyramid: the foundation is the basic indicators (outstanding performance on these indicators is necessary but not sufficient to drive the primary outcomes). Overall, these basics are being well managed currently, but there remains an excessive focus on them, largely because of publically reported data and regulatory requirements. With a strong foundation, it is now time to focus on the more complex intermediate clinical outcomes-fluid management, infection control, diabetes management, medication management, and end-of-life care among others. Successfully addressing these intermediate outcomes will drive improvements in the primary outcomes, better survival, fewer hospitalizations, better patient experience with the treatment, and ultimately, improved quality of life. By articulating this view of quality in the ESRD program (pushing up the quality pyramid), the discussion about quality is reframed, and also, clinicians can better target their facilities in the direction of regulatory oversight and requirements about quality. Clinicians owe it to their patients, as the ESRD program celebrates its 40th anniversary, to rekindle the aspirations of the creators of the program, whose primary goal was to improve the lives of the patients afflicted with this devastating condition.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A

_____

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

_____

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
1463_Evidence_form.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Hospitalization rates are an important indicator of patient morbidity and quality of life. On average, dialysis patients are admitted to the hospital nearly twice a year and spend an average of 11.2 days in the hospital per year [1]. Hospitalizations account for approximately 40 percent of total Medicare expenditures for ESRD patients [1]. Measures of the frequency of hospitalization have the potential to help efforts to control escalating medical costs, and to play an important role in identifying potential problems and

helping facilities provide cost-effective health care.

[1] United States Renal Data System. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Standardized hospitalization admission rates vary widely across facilities. For example, for 2014, the SHR Admissions varied from 0.07 to 2.92. The mean value was 0.99 and the SD was 0.27. The data used to calculate these rates is limited to those facilities with at least 5 patient years at risk (reflecting how the measure is currently calculated on DFC).

Distribution of the SHR, 2011-2014:

2011: Facilities = 5386, Mean SHR = .99, Standard Error = .28, 10th = .66, 25th = .80, 50th = .96, 75th = 1.14, 90th = 1.33

2012: Facilities = 5568, Mean SHR = .99, Standard Error = .28, 10th = .66, 25th = .81, 50th = .97, 75th = 1.15, 90th = 1.34

2013: Facilities = 5700, Mean SHR = .99, Standard Error = .27, 10th = .68, 25th = .81, 50th = .97, 75th = 1.15, 90th = 1.33

2014: Facilities = 5857, Mean SHR = .99, Standard Error = .27, 10th = .68, 25th = .82, 50th = .97, 75th = 1.14, 90th = 1.32

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Race and ethnicity have been shown to be predictors of hospitalization. Using data from 2013, it is observed that white and black patients are hospitalized at similar rates (both SHRs = 1.01). Native American and Asian/Pacific Islander patients are hospitalized at lower rates than would be expected (SHR = 0.90 and 0.84, respectively). Also, Hispanic patients had slightly lower than expected hospitalization rates (SHR = 0.98), while non-Hispanic and patients of unknown ethnicity were hospitalized at the same rate (both SHRs = 1.00). While there are differences across the race and ethnicity groups, the results suggest no clear disparities in outcomes and that it would not be appropriate to adjust for these factors.

Refer to Risk Adjustment section (2b4) for further analyses on race, ethnicity, sex and socioeconomic status.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, High resource use, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Hospitalization rates are an important indicator of patient morbidity and quality of life. On average, dialysis patients are admitted to the hospital twice a year and spend an average of 11.2 days in the hospital per year [1]. Hospitalizations account for approximately 40 percent of total Medicare expenditures for ESRD patients [1]. Measures of the frequency of hospitalization have the potential to help efforts to control escalating medical costs, and to play an important role in identifying potential problems and helping facilities provide cost-effective health care.

At the end of 2013 there were 661,648 patients being dialyzed, of which 117,162 were new (incident) ESRD patients [1]. In 2013, total Medicare costs for the ESRD program were $30.9 billion, a 1.6% increase from 2012 [1]. Correspondingly, hospitalization costs for ESRD patients are very high with Medicare costs of over $10.3 billion in 2013.

Hospitalization measures have been in use in the Dialysis Facility Reports (formerly Unit-Specific Reports) since 1995. The Dialysis Facility Reports are used by the dialysis facilities and ESRD Networks for quality improvement, and by ESRD state surveyors for monitoring and surveillance. In particular, the SHR for Admissions is used by ESRD state surveyors in conjunction with other standard criteria for prioritizing and selecting facilities to survey and has been found to be predictive of citations in the past (ESRD State Outcomes List). The SHR is also a public reporting measure on the Centers for Medicare and Medicaid Services (CMS) Dialysis Facility Compare website.

As noted above, hospitalization among dialysis patients is common and accounts for a large fraction of Medicare expenditures for ESRD beneficiaries. The Agency for Healthcare Research and Quality (AHRQ) Prevention Quality Indicators (PQIs) has identified several diagnoses where timely and effective ambulatory care can significantly reduce hospitalization. These diagnoses represent hospitalizations that might be prevented with effective ambulatory care including but not limited to dialysis facilities. We identified the PQIs most common for ESRD patients and compared the frequency of those diagnoses for the ESRD population to that of the general Medicare population in the fee-for-service system. Based on clinical input we identified several other diagnoses common among dialysis patients that may be preventable through the delivery of appropriate dialysis care [2]. Our analysis showed that compared to the general Medicare population, ESRD patients were hospitalized at higher rates for the following potentially preventable conditions as defined by AHRQ PQIs: diabetes with long term complications (16 times the rate of the general Medicare population), lower extremity amputation (22 times), and diabetes with short term complications (22 times). Applying the ESRD-specific potentially preventable conditions, ESRD patients were hospitalized at a higher rate for the following: complications of device/implant/graft (ESRD-related only) (13 times), septicemia (except in labor) (7 times) and fluid and electrolyte disorder (8 times). Since for most dialysis patients the dialysis facility is the principal source of ambulatory care and may even be considered by some as their medical home, it is reasonable to expect that high quality care by the dialysis facility could reduce the very high rate of hospitalizations among dialysis patients. Further, the facility-level correlation between the hospitalization rate for potentially preventable hospitalizations and that for all hospitalizations (the SHR) was found in this study to be high (0.84 for facilities with more than 20 patient years). This result provides further evidence that facilities have opportunities to reduce hospitalizations through appropriate dialysis care [2].

A 2015 Technical Expert Panel closely reviewed comorbidities related to hospitalization and provided an assessment of each and the likelihood whether they were related to facility care. This assessment process and the results are further described in the risk adjustment section below.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

[1] United States Renal Data System. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015.
[2] Wheeler J, Hirth R, Meyer K, Messana JM. Exploring preventable hospitalizations of dialysis patients. J Am Soc Nephrol 22, 2011.
[3] Erickson KF, Winkelmayer WC, Chertow GM, Bhattacharya J. Physician visits and 30-day hospital readmissions in patients receiving hemodialysis. J Am Soc Nephrol 25, 2014 (published online before print).
[4] Arora P, Kausz AT, Obrador GT, Ruthazer R, Khan S, Jenuleson CS, Meyer KB, Pereira BJ. Hospital utilization among chronic dialysis patients. J Am Soc Nephrol 11: 740 –746, 2000.
[5] Piraino B. Staphylococcus aureus infections in dialysis patients: focus on prevention. ASAIO J 46(6): S13-S17, 2000.
[6] Dalrymple LS, Johansen KL, Romano PS, Chertow GM, Mu Y, Ishida JH, Grimes B, Kaysen GA, Nguyen DV. Comparison of hospitalization rates among for-profit and nonprofit dialysis facilities. Clin J Am Soc Nephrol 9, 2014 (published online before print).

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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:** 1463_Data_Dictionary_Code_Table.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
This form is being used for endorsement maintenance. Updates include:

- The model now adjusts for each incident comorbidity separately rather than using a comorbidity index.
- We have also modified the indicators for diabetes by consolidating the individual indicators.
- We have included adjustments for 210 prevalent comorbidities (identified through Medicare claims)

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of inpatient hospital admissions among eligible patients at the facility during the reporting period.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
At least one year.

S.6. **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, 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.
The numerator is calculated through use of Medicare claims data. When a claim is made for an inpatient hospitalization, the patient is identified and attributed to a dialysis facility following rules discussed below in the denominator details. The numerator is the count of all such hospitalizations over the reporting period.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of hospital admissions that would be expected among eligible patients at the facility during the reporting period, given the patient mix at the facility.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Assignment of Patients to Facilities
UM-KECC's treatment history file provides a complete history of the status, location, and dialysis treatment modality of an ESRD patient from the date of the first ESRD service until the patient dies or the data collection cutoff date is reached. For each patient, a new record is created each time he/she changes facility or treatment modality. Each record represents a time period associated with a specific modality and dialysis facility. SIMS/CROWNWeb is the primary basis for placing patients at dialysis facilities, and dialysis claims are used as an additional source. Information regarding first ESRD service date, death and transplant is obtained from additional sources including the CMS Medical Evidence Form (Form CMS-2728), transplant data from the Organ Procurement and Transplant Network (OPTN), the Death Notification Form (Form CMS-2746) and the Social Security Death Master File.

As patients can receive dialysis treatment at more than one facility in a given year, we assign each patient day to a facility (or no facility, in some cases) based on a set of conventions described below, which largely align with those for the Standardized Mortality Ratio (SMR). We detail patient inclusion criteria, facility assignment and how to count days at risk, all of which are required for the risk adjustment model.

General Inclusion Criteria for Dialysis Patients
Though a patient's follow-up in the database can be incomplete during the first 90 days of ESRD therapy, we only include a patient's follow-up in the tabulations after that patient has received chronic renal replacement therapy for at least 90 days. Thus, hospitalizations, mortality and survival during the first 90 days of ESRD do not enter into the calculations. This minimum 90-day period also assures that most patients are eligible for Medicare, either as their primary or secondary insurer. It also excludes from analysis patients who die or recover renal function during the first 90 days of ESRD.

In order to exclude patients who only received temporary dialysis therapy, we assign patients to a facility only after they have been on dialysis there for the past 60 days. This 60 day period is used both for patients who started ESRD for the first time and for those who returned to dialysis after a transplant. That is, hospitalizations during the first 60 days of dialysis at a facility do not affect the SHR of that facility.

Identifying Facility Treatment Histories for Each Patient
For each patient, we identify the dialysis provider at each point in time. Starting with day 91 after onset of ESRD, we attribute patients to facilities according to the following rules. A patient is attributed to a facility once the patient has been treated there for the past 60 days. When a patient transfers from one facility to another, the patient continues to be attributed to the original facility for 60 days and then is attributed to the destination facility. In particular, a patient is attributed to his or her current facility on day 91 of ESRD if that facility had treated him or her for the past 60 days. If on day 91, the facility had not treated a patient for the past 60 days, we wait until the patient reaches day 60 of continuous treatment at that facility before attributing the patient to that facility. When a patient is not treated in a single facility for a span of 60 days (for instance, if there were two switches within 60 days of each other), we do not attribute that patient to any facility. Patients are removed from facilities three days prior to transplant in order to exclude the transplant hospitalization. Patients who withdrew from dialysis or recovered renal function remain assigned to their treatment facility for 60 days after withdrawal or recovery.

If a period of one year passes with neither paid dialysis claims nor SIMS information to indicate that a patient was receiving dialysis treatment, we consider the patient lost to follow-up and do not include that patient in the analysis. If dialysis claims or other evidence of dialysis reappears, the patient is entered into analysis after 60 days of continuous therapy at a single facility.

Days at Risk for Medicare Dialysis Patients
After patient treatment histories are defined as described above, periods of follow-up in time since ESRD onset are created for each patient. In order to adjust for duration of ESRD appropriately, we define 6 time intervals with cut points at 6 months, 1 year, 2 years, 3 years and 5 years. A new time period begins each time the patient is determined to be at a different facility, or at the start of each calendar year or when crossing any of the above cut points.

Since hospitalization data tend not to be as complete as mortality data, we include only patients whose Medicare billing records include all hospitalizations. To achieve this goal, we require that patients reach a certain level of Medicare-paid dialysis bills to be included in the hospitalization statistics, or that patients have Medicare-paid inpatient claims during the period. Specifically, months within a given dialysis patient-period are used for SHR calculation when they meet the criterion of being within two months after a month with either: (a) $900+ of Medicare-paid dialysis claims OR (b) at least one Medicare-paid inpatient claim. The intention of this criterion is to assure completeness of information on hospitalizations for all patients included in the analysis.

The number of days at risk in each of these patient-ESRD facility-year time periods is used to calculate the expected number of hospital admissions for the patient during that period. The SHR for a facility is the ratio of the total number of observed hospitalizations to the total number of expected hospitalizations during all time periods at the facility. Based on a risk adjustment model for the overall national hospitalization rates, we compute the expected number of hospitalizations that would occur for each month that each patient is attributed to a given facility. The sum of all such expectations for patients and months yields the overall number of hospital admissions that would be expected given the specific patient mix and this forms the denominator of the measure.

The denominator of the SHR 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).

References:

Cook, R. and Lawless, J. The Statistical Analysis of Recurrent Events. New York: Springer. 2007.
Cox, D.R. (1972) Regression Models and Life Tables (with Discussion). J. Royal statistical Society, Series B, 34, 187-220.
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.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
None.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
N/A

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The regression model used to compute a facility's "expected" number of hospitalizations for the SHR measure contains many factors thought to be associated with hospitalization rates. Specifically, the model adjusts for patient age, sex, diabetes as cause of ESRD, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, prevalent comorbidities, and calendar year. The stage 1 model allows the baseline hospitalization rates to vary between strata, which are defined by 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. In essence, it avoids a possible confounding between facility effects and patient covariates as can arise, for example, if patients with favorable values of the covariate tend to be treated at facilities with better treatment policies and outcomes. Thus, for example, if patients with diabetes as a cause of ESRD tended to be treated at better facilities, one would underestimate the effect of diabetes unless the model is adjusted for facility. In this model, facility adjustment is done by stratification.

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.
•Sex: We determine each patient's sex from his/her Medical Evidence Form (CMS-2728).
•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 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.
•Prevalent comorbidities: We identify a patient's prevalent comorbidities based on claims from the previous calendar year. The comorbidities adjusted for include those listed in data dictionary/code table (excel file).
•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, sex and duration and cause of ESRD are also included:
•Diabetes as cause of ESRD*Duration of ESRD
•Diabetes as cause of ESRD*Sex
•Diabetes as cause of ESRD*Age
•Age*Sex

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)
See flowchart in appendix.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patients with missing data are not excluded from the model. For the purposes of calculation, missing values for BMI are replaced with mean values for patients of similar age and identical race, sex, and cause of ESRD. Missing values for cause of ESRD are replaced with the other/unknown category. No patients were missing age, sex, or date of first ESRD treatment. Indicator variables identifying patients with missing values for cause of ESRD, comorbidities at incidence (missing CMS-2728), and BMI are also included as covariates in the model. For 2010-2013, 3% of the patients included in the SHR model calculation were missing BMI.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(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.

In calculating the SHR, Medicare inpatient claims that are adjacent or overlap with another claim are collapsed into one record. Specifically, if the admission date of an inpatient record is within one day of a following admission's discharge date, these adjacent inpatient records will be collapsed into one inpatient record that takes on the first admission's admission date and the following admission's discharge date. Similarly, if an inpatient record overlaps with another inpatient record, the two records are collapsed into one record where the earliest admission date between the two records becomes the new admission date and the latest discharge date between the two records becomes the new discharge date.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Dialysis Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
1463_Testing_form-635963152645467173.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 1463

Measure Title: Standardized Hospitalization Ratio for Dialysis Facilities

Date of Submission: 4/15/2016

Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *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.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at

start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/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), 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.

1.3. What are the dates of the data used in testing? Calendar years 2010 through 2013

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.26*)**	**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 <u>measured entities</u> 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 of the four years from 2010-2013 there were 5,406, 5,582, 5,708 and 5,863 facilities, respectively.

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*)

Medicare dialysis patients were included in the testing and analysis for each of the four years from 2010-2013 of which there were 377,675, 387,249, 396,167 and 403,337 patients, respectively.

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.

N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patient level:

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

**Assessed at the start of time at risk based on calendar year and facility assignment. Medicare coverage in the model was defined as:*
1. Medicare as primary and Medicaid
2. Medicare as primary and NO Medicaid
3. Medicare as secondary or Medicare HMO

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

Proxy/Area level: 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+) without high school diploma (%)

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*)

2011 Submission

Reliability of the Standardized Hospital Ratio for Admissions was assessed using data on hospitalizations among ESRD patients over a three year period of 2006-2008 for 4338 dialysis centers. Data for the hospitalization measures are derived from an extensive national ESRD patient database, which is largely derived from the Standard Information Management System (SIMS) database maintained by the 18 ESRD Networks, the CMS Annual Facility Survey (Form CMS-2744), Medicare dialysis and hospital payment records, the CMS Medical Evidence Form (Form CMS-2728), transplant data from the Organ Procurement and Transplant Network (OPTN), the Death Notification Form (Form CMS-2746), the Nursing Home Minimum Dataset, and the Social Security Death Master File. The database is comprehensive for Medicare patients. Information on hospitalizations is obtained from Medicare Inpatient Claims Standard Analysis Files (SAFs).

To assess reliability, we assessed the degree to which the measures were consistent year to year. If one looks at two adjacent time intervals, one should expect that a reliable measure will exhibit correlation over these periods since large changes in patterns affecting the measure should not occur for most centers over shorter periods. Year to year variability in the SHR values was assessed across the years 2006, 2007 and 2008 based on the 4338 dialysis centers for which an SHR is reported in the 2010 DFRs.

2016 Submission

The reliability of the SHR was assessed using data among Medicare ESRD dialysis patients during 2010-2013. 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 total variation of a measure that is attributable to the between-facility variation. The SHR, 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 SHR 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 *i*th facility has n_i subjects, randomly draw with replacement n_i subjects from those in the same facility, find their corresponding SHR_i and repeat the process B (say, 100) times. Thus, for the *i*th facility, we have bootstrapped SHRs of $T_{i1}^*,\ldots, T_{i200}^*$. Let S_i^* be the sample variance of this bootstrap sample. From this it can be seen that

$$s_{t,w}^2 = \frac{\sum_{i=1}^{N}[(n_i - 1)S_i^{*2}]}{\sum_{i=1}^{N}(n_i - 1)}$$

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

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

where

$$\bar{T} = \sum n_i T_i / \sum n_i$$

is the weighted mean of the observed SHR 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 the total variation of SHR and 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 estimated IUR, which is defined by

$$IUR = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{t,w}^2}$$

can be estimated with $(s_t^2 - s_{t,w}^2)/s_t^2$.

The SHR calculation only included facilities with at least 5 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*)

2011 Submission

The correlation between SHR admissions across adjacent years (2006 versus 2007 and 2007 vs 2008) was approximately 0.67 indicating that centers with large or small SHR tended to have larger or smaller SHR on the following year. These correlations were highly significant. Similarly, there was persistence in SHRs that were significant from year to year. For example, there were 4.3% of facilities that had significant evidence of a true SHR of at least 1.2 in 2006. Of those that were significantly larger than 1.2 in 2006, 1.8/4.3 = 42% were again significantly larger than 1.2 in 2007. Of those that were not significant in 2006, only 2.5% were found to be significantly larger than 1.2 in 2007.

The measure is based on complete data and is not subject to judgment or rater variability. Hence the measures of inter-rater variability are not relevant here.

2016 Submission

Overall, we found that IURs for the one-year SHRs have a range of 0.70-0.72 across the years 2010, 2011, 2012 and 2013, which indicates that over two-thirds of the variation in the one-year SHR can be attributed to the between-facility differences and less than one-third to within-facility variation.

Table 1: IUR for one-year SHR, Overall and by Facility Size, 2010-2013

	2010		2011		2012		2013	
Facility Size (Number of patients)	**IUR**	**N**	**IUR**	**N**	**IUR**	**N**	**IUR**	**N**
All	0.72	5407	0.71	5583	0.70	5709	0.70	5864
Small (<=50)	0.54	1864	0.51	1921	0.48	1977	0.46	2028
Medium (51–87)	0.65	1702	0.63	1785	0.58	1825	0.57	1930
Large (>=88)	0.81	1841	0.81	1877	0.81	1907	0.82	1906

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?*)

2011 Submission

This was not a question on the 2011 Submission Form.

2016 Submission

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

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

2011 Submission

Validity of the Standardized Hospital Ratio for Admissions was assessed using data on hospitalizations as well as other quality measures among ESRD patients over a three year period of 2006-2008. We examined the validity of the measure by examining its covariability with other measures of quality as well as by examining the relationship of the overall hospitalization measure with measures that were more directly focused on specific causes.

We have assessed the validity of the measure through various comparisons of this measure with other quality measures in use. Also, hospitalization measures were reviewed by a TEP in 2007 and overall measures based on admissions and on days were recommended for inclusion in the Dialysis Facility reports. In addition, hospitalization is a major cost factor in the management of ESRD patients as noted earlier, so there is here a very strong case for face validity of the SHR admissions measure.

2016 Submission

We have assessed the validity of the measure through various comparisons of this measure with other quality measures in use, using Spearman correlations.

The measure is also maintained on face validity. Hospitalization measures were reviewed by a TEP in 2007 and overall measures based on admissions and on days were recommended for inclusion in the Dialysis Facility Reports. In 2015, a TEP was held specifically to consider prevalent comorbidity adjustments for inclusion in the measure. The TEP's recommendations are reflected in the risk adjustment methodology. In addition, hospitalization is a major cost factor in

the management of ESRD patients as noted earlier, further establishing a very strong case for face validity of the SHR admissions measure.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

2011 Submission

The SHR Admissions measure is correlated with the Standardized Mortality Ratio (SMR) over the three year cohort (r=0.40) and in individual years r was approximately equal to 0.33, both correlations being highly significant. In addition, SHR Admissions is negatively correlated in each of the three year with percent of patients in the facility with AV Fistula (r=-0.27, -0.23, -0.21). Thus higher values of SHR are associated with lower usage of AV Fistulas. On the other hand, SHR admissions is positively correlated with catheter use (r=0.24, 0.23, 0.22), indicating that higher values of SHR are associated with increased use of catheters. These correlations are all highly significant ($p<0.001$). The SHR Admissions is also found to be negatively correlated (r=-0.10, $p<0.0001$) with the percent of patients with URR>65, again in the direction expected.

The SHR Admissions is an overall measure of hospital use and is comprised of many different causes or reasons for hospitalization. The TEP considered the possibility of devising cause specific SHRs, but recommended the use of overall SHR measures due to various reasons including the lack of clear research to indicate what causes should be selected as indicative of poor ESRD care and issues associated with inter-rater reliability in assessing cause of hospitalization. The TEP reached a strong consensus that the overall measures should give a reliable and valid measure that would typically be related to quality of care. We have some crude measures of cause of hospitalization which we have taken to assess the relationship between the overall measure and cause specific components. These measures are useful in assessing the overall SHR measures, but we caution that the cause specific hospitalizations have not been tested or validated at this time. The overall SHR Admissions is strongly correlated with the SHR for cause specific hospitalizations. The correlation with Septicemia is r=0.44, with Chronic Heart Failure is r=0.55 and with an overall measure including Septicemia and a collection of coronary causes is r=0.66. Thus the overall hospitalization rate also correlates strongly with causes that are commonly thought to be potentially related to poor quality of care.

2016 Submission

The SHR Admissions measure is correlated with the Standardized Mortality Ratio (SMR) for each individual year from 2010-2013, where Spearman's correlation coefficient ranged from 0.27 to 0.30, with all four correlations being highly significant ($p<0.0001$). Also for each year from 2011-2013, the SHR was correlated with the Standardized Readmission Ratio (SRR) (Spearman's rho=0.54, 0.50, 0.48; $p<0.0001$).

In addition, SHR Admissions is negatively correlated in each of the four years with percent of patients in the facility with AV Fistula (Spearman's rho= -0.12, -0.15, -0.12, -0.13). Thus higher values of SHR are associated with lower usage of AV Fistulas. Further, SHR admissions is positively correlated in each of the four years with percent of patients with catheter >= 90 days (Spearman's rho=0.21, 0.21, 0.18, 0.16), indicating that higher values of SHR are associated with increased use of catheters. These correlations are all highly significant ($p<0.001$). For each year of 2010 through 2013, the SHR Admissions is also found to be negatively correlated with the percent of hemodialysis patients with Kt/V>=1.2, again in the direction expected (Spearman's rho= -0.11, -0.13, -0.10,-0.11; $p<0.0001$). Lower SHRs are associated with a higher percentage of patients receiving adequate dialysis dose.

2b2.4. 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?*)

2011 Submission

This was not a question on the 2011 Submission Form.

2016 Submission

The SHR correlates with outcomes, processes of care, and causes of hospitalization that are commonly thought to be potentially related to poor quality of care. Higher hospitalization was associated with higher facility mortality rates; and similarly with higher readmissions. We found higher values of SHR are associated with lower usage of AV Fistulas, higher catheter use, and suboptimal dialysis adequacy.

The 2007 TEP considered the possibility of developing cause specific SHRs, but recommended the use of all-cause SHR measures due to various reasons including the lack of clear research to indicate what causes (i.e., reason for admission) should be selected as valid indicators of poor ESRD care, and issues associated with inter-rater reliability in assessing cause of hospitalization. The TEP reached a strong consensus that the all-cause measure would be reliable and valid and the measure would typically be related to quality of care. We have some crude measures of cause of hospitalization which we have used to assess the relationship between the all-cause measure and cause specific components. These measures are useful in assessing the overall SHR measures, but we caution that the cause specific hospitalizations have not been tested or validated at this time. All correlations are in the expected direction and highly significant, ($p<0.0001$). Thus these preliminary analyses show that the overall hospitalization rate also correlates with specific causes that are commonly thought to be potentially related to poor quality of care. In 2015, a TEP was held specifically to consider prevalent comorbidity adjustments for inclusion in this measure (and SMR). The TEP's recommendations are reflected in the risk adjustment methodology.

2b3. EXCLUSIONS ANALYSIS
NA ☒ no exclusions — *skip to section 2b4*

2b3.1. 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*)

N/A

2b3.2. 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*)

N/A

2b3.3. 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*)

N/A

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with** 229 **risk factors (**diabetes, sex, age, BMI at incidence, calendar year, nursing home status, 13 comorbidities at incidence and 210 prevalent comorbidities)

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

The risk adjustment is based on a Cox or relative risk model. The adjustment is made for patient age, sex, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, a set of prevalent comorbidities, and calendar year. In this model, covariates are taken to act multiplicatively on the admission rate and the adjustment model is fitted with facility defining strata so as to provide valid estimates even if the distribution of adjustment variables differs across facilities. Relevant references are Cox (1972), Kalbfleisch and Prentice (2002), Lawless and Nadeau (1995), Lin et al. (2000), Cook and Lawless (2007) and Liu, Schaubel and Kalbfleisch (2010). All analyses are done using SAS.

In general, adjustment factors for the SHR were selected based on several considerations. As noted above, we began with a large set of patient characteristics, including demographics, comorbidities at ESRD incidence, a set of prevalent comorbidities, and other characteristics. Factors considered appropriate were then investigated with statistical models, including interactions between sets of adjusters, to determine if they were related to hospitalizations. Factors related to the SHR were also evaluated for face validity before being included. Finally, SDS/SES factors were evaluated based on appropriateness (whether related to disparities in care), empirical association with the outcome, and as supported in published literature.

First, in 2007, a Technical Expert Panel was convened; the TEP provided advice on various aspects of the SHR, including adjustment factors. The 2007 Hospitalization TEP felt that facility characteristics are generally not appropriate for use as adjusters, but should be evaluated for their potential as proxies for patient characteristics. They also recommended that facility market characteristics, such as local hospital utilization rates, should not be considered as risk adjusters.

More recently, there has been great interest among dialysis care providers and other stakeholders in adjusting for more current (prevalent) comorbidities to reflect the current health status of dialysis patients, and specifically inclusion of conditions associated with hospitalization. In response CMS contracted with UM-KECC to convene a Technical Expert Panel (TEP) in September 2015 to consider the addition of prevalent comorbidity risk adjustment. The summary report for the TEP can be found here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TechnicalExpertPanels.html. The TEP was charged with evaluating the potential of including prevalent comorbidities in the SMR and SHR risk adjustment models. Specific objectives included: (1) review of the comorbidity adjustment (determined at ESRD incidence) in the current NQF endorsed SMR and SHR measures; and (2) consideration of what, if any, prevalent comorbidities would be appropriate to include in each measure. In developing its recommendations, the TEP was asked to apply the criteria for risk-adjusters developed by the National Quality Forum

(NQF): (1) Risk adjustment should be based on patient factors that influence the measured outcome and are present at the start of care; (2) Measures should not be adjusted for factors related to disparities in care or the quality of care; (3) Risk adjustment factors must be substantially related to the outcome being measured; (4) Risk adjustment factors should not reflect quality of care by the provider/facility being evaluated.

Reflecting these criteria, the TEP evaluated a list of prevalent comorbidities derived through the following process. First, the ESRD Hierarchical Condition Categories (ESRD-HCCs) were used as a starting point to identify ICD-9 diagnosis codes related to dialysis care. Those individual ICD-9 conditions that comprised the respective ESRD HCCs, with a prevalence of at least 0.1% in the patient population, were then selected for analysis to determine their statistical relationship to mortality and/or hospitalization. This step resulted in 555 diagnoses comorbidities (out of over 3000 ICD-9 diagnosis codes in the ESRD-HCCs). Next, an adaptive lasso variable selection method was applied to these 555 diagnoses to identify those with a statistically significant relationship to mortality and/or hospitalization ($p<0.05$). This process identified 242 diagnoses. The TEP members then scored each of these diagnoses as follows:

1. Very likely the result of dialysis facility care
2. Likely the result of dialysis facility care
3. May or may not be the result of dialysis facility care
4. Unlikely to be the result of dialysis facility care
5. Very likely not the result of dialysis facility care

This scoring exercise aimed at identifying a set of prevalent comorbidities not likely the result of facility care and therefore potentially appropriate as risk adjusters for SHR and SMR. The TEP established that comorbidities scored as "unlikely" or "very unlikely the result of facility care" by at least half of TEP members (simple majority) were judged as appropriate for inclusion as risk-adjusters. This process resulted in 210 conditions as risk adjustors. The TEP further recommended that: (1) comorbidities for inclusion as risk-adjusters in a particular year should be present in Medicare claims in the preceding calendar year; and (2) determination of a prevalent comorbidity required at least two outpatient claims or one inpatient claim. The set of prevalent comorbidities recommended by the TEP for inclusion as risk-adjusters is presented listed below.

Consideration of SDS/SES risk factors

The relationship among patient level SDS, socioeconomic disadvantage and health care utilization such as hospitalization is well-established in the general population and has received considerable attention over the years. (AHRQ Reports, 2011; 2012; 2013; 2014; 2015). The likelihood of hospitalization is related to socioeconomic disadvantage through differences in health status, insurance coverage, and access to quality primary care (Basu et al, 2012; Blustein et al, 1998). Further, individual and market or area-level measures of deprivation have been shown to contribute independently to preventable hospitalizations (Moy et al, 2013).

Health care outcomes and utilization are associated with area-level income and residential segregation, but particularly so for racial minorities (Williams, 2006 ; Williams and Collins, 2001). This suggests the interplay of patient level (race) and area level SES factors related to lower income, neighborhood poverty, segregation, levels of educational attainment, and unemployment levels that jointly influence key health outcomes related to morbidity (Williams 2006; Williams and Collins, 2001; AHRQ, 2008).

Within the dialysis population area-level SES are associated with poor outcomes (Almachraki et al 2016); while patient level factors such as race are predictive of differences in certain clinical outcomes by race. (Yan et al 2014; Whittle et al 1991). In a study of first year hemodialysis patients, patients of Hispanic ethnicity had lowest all-cause hospital length of stay compared to whites, while patients of black race had intermediate all-cause hospital admissions that was lower

relative to whites but higher than Hispanic patient, with differences observed across certain age groups (Yan et al, CJASN 2014). Moreover the study authors found that infection-related hospitalizations were significantly higher for black and Hispanic patients compared to non-Hispanic whites. These associations could indicate certain facility level practices related to effective infection control and prevention may unevenly impact patients of black race and Hispanic ethnicity (Yan et al CJASN 2014 p7).

Insurance status is also related to health outcomes but this has not been studied extensively within the dialysis population as it relates to hospitalization, though the association has been documented in studies of the general dual Medicare and Medicaid population. Dual eligibles typically have greater comorbidity burden, face access to care barriers which in turn drive higher hospital utilization (Jiang et al, 2010; Moon and Shin,2006; Wright et al., 2015).

Maintaining employment is a challenge for dialysis patients which in turn can influence well-being and may have a proximal impact on outcomes such as hospitalization (Curtin et al, AJKD 1996).

Given these observed linkages we tested these patient- and area-level SDS/SES variables based on the conceptual relationships as described above and demonstrated in the literature, as well as the availability of data for the analyses. Measures of area-level socioeconomic deprivation are included as individual components from the Area Deprivation Index (Singh, 2003).

References:

Agency for Healthcare Research and Quality, Rockville, MD. Internet Citation: Chapter 3: Creation of New Race-Ethnicity Codes and SES Indicators for Medicare Beneficiaries - Chapter 3. January 2008. Publication # 08-0029-EF. http://archive.ahrq.gov/research/findings/final-reports/medicareindicators/medicareindicators3.html

Agency for Healthcare Research and Quality (AHRQ). 2010 National Health Care Disparities Report. Washington, DC: AHRQ; 2011).

Agency for Healthcare Research and Quality (AHRQ). 2011 National Health Care Disparities Report. Washington, DC: AHRQ; 2012).

Agency for Healthcare Research and Quality (AHRQ). 2012 National Health Care Disparities Report. Washington, DC: AHRQ; Reports: 2013).

Agency for Healthcare Research and Quality (AHRQ). 2013 National Health Care Disparities Report. Washington, DC: AHRQ; Reports: 2014).

Agency for Healthcare Research and Quality (AHRQ). 2014 National Health Care Disparities Report. Washington, DC: AHRQ; 2015).

Almachraki F, Tuffli M, Lee P, Desmarais M, Shih HC, Nissenson A, and Krishnan M. Population Health Management. Volume 19, Number 1, 2016.

Basu, J., Thumula, V., and Mobley, L.R. (2012, July-September). Changes In Preventable Hospitalization Patterns Among Adults. A Small Area Analysis Of U.S. States. Journal of Ambulatory Care Management 35(3), pp.3280-3290

J Blustein, K Hanson and S Shea. Preventable Hospitalizations And Socioeconomic Status. Health Affairs 17, no.2 (1998):177-189).

Curtin R, Oberley E, Sacksteder P, and Friedman A. Differences Between Employed and Nonemployed Dialysis Patients. AJKD Vol 27:4. (April) 1996. 533-540.

J. F. Lawless and C. Nadeau. Some Simple Robust Methods for the Analysis of Recurrent Events Technometrics. Vol. 37, No. 2 (May, 1995), pp. 158-168

Cox DR: Regression models and life tables (with discussion). JRStat Soc [SerB]34: 187–220, 1972

Jiang H, Wier L, Potter DEB, Burgess J. AHRQ Statistical Brief #96 Potentially Preventable Hospitalizations among Medicare-Medicaid Dual Eligibles, September 2010.

Kalbfleisch JD, Prentice RL: The statistical analysis of failure time data, Hoboken, New Jersey, John Wiley & Sons, Inc., 2002

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.

Liu, D., Schaubel, D.E. and Kalbfleisch, J.D. (2012). Computationally efficient marginal models for clustered recurrent event data. Biometrics, 68, 637-647.

Moon S., Shin J. BMC Public Health. 2006 Apr 5;6:88. Health Care Utilization Among Medicare-Medicaid Dual Eligibles: A Count Data Analysis.

Moy E, Chang E, Barrett M. Potentially Preventable Hospitalizations — United States, 2001–2009. CDC Morbidity and Mortality Weekly Report (MMWR). Supplements November 22, 2013 / 62(03);139-143

Rodriguez R, Sen S, Mehta K, Moody-Ayers S, Bacchetti P, and O'Hare A. Geography Matters: Relationships among Urban Residential Segregation, Dialysis Facilities, and Patient Outcomes. Ann Intern Med. 2007; 146:493-501.

Singh, GK. Area Deprivation And Widening Inequalities In US Mortality, 1969–1998. Am J Public Health. 2003; 93(7):1137–1143.

Williams D. "Race, Socioeconomic Status, and Health: The Added Effects of Racism and Discrimination. Annals of the New York Academy of Sciences. Volume 896, Issue 1, Article first published online: 6 FEB 2006.

Williams D, and Collins C, Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. Public Health Reports / September–October 2001. Volume 116. 404-416.

Whittle JC, Whelton PK, Seidler AJ, Klag MJ. Does Racial Variation In Risk Factors Explain Black-White Differences In The Incidence Of Hypertensive End-Stage Renal Disease? Arch Intern Med. 1991 Jul;151(7):1359-64.

Wright B, Potter A, and Trivedi A. Federally Qualified Health Center Use Among Dual Eligibles: Rates Of Hospitalizations And Emergency Department Visits Health Affairs, 34, no.7 (2015):1147-1155.

Yan G, Norris K, Greene T, Yu A, Ma J, Yu W, and Cheung A. Race/Ethnicity, Age, and Risk of Hospital Admission and Length of Stay during the First Year of Maintenance Hemodialysis. Clin J Am Soc Nephrol 9: epub (June), 2014.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Table 2a. Model Coefficients, Data Years 2010–2013.

Covariate	Coefficient	P-value
Comorbidities at start of ESRD		
At least one of the comorbidities listed below	0.08624	<.0001
Atherosclerotic heart disease	0.04999	<.0001
Other cardiac disease	0.04395	<.0001
Diabetes*	-0.02026	<.0001
Congestive heart failure	0.04269	<.0001
Inability to ambulate	0.02042	<.0001
Chronic obstructive pulmonary disease	0.05646	<.0001
Inability to transfer	0.02401	<.0001
Malignant neoplasm, cancer	0.04102	<.0001
Peripheral vascular disease	0.04104	<.0001
Cerebrovascular disease, CVA, TIA	0.01904	<.0001
Tobacco use (current smoker)	0.08539	<.0001
Alcohol dependence	0.01285	0.036
Drug dependence	0.17361	<.0001
No Medical Evidence (CMS-2728) Form	0.15316	<.0001
Cause of ESRD		
Diabetes	0.03848	<.0001
Missing	-0.03547	<.0001
Sex: Female	0.07156	<.0001
Age		
0-14	0.48884	<.0001
15-24	0.13135	<.0001
25-44	-0.0678	<.0001
45-59	-0.065	<.0001
60-74	Reference	

Covariate	Coefficient	P-value
75+	0.10178	<.0001
BMI		
Log BMI	-0.15032	<.0001
BMI missing	0.01656	0.0002
Calendar year		
2010	Reference	
2011	-0.02546	<.0001
2012	-0.12676	<.0001
2013	-0.16265	<.0001
In nursing home the previous year	0.20788	<.0001
Diabetes as cause of ESRD X time on ESRD interaction term		
91 days-6 months	Reference	
6 months-1 year	0.03417	<.0001
1-2 years	0.01166	0.0737
2-3 years	0.00139	0.8356
3-5 years	-0.01549	0.0147
5+ years	-0.06398	<.0001
Cause of ESRD: diabetes X sex: female interaction term	-0.02622	<.0001
Age X diabetes as cause of ESRD interaction term		
0-14	-0.93749	<.0001
15-24	0.16727	<.0001
25-44	0.15502	<.0001
45-59	0.05013	<.0001
60-74	Reference	
75+	-0.03426	<.0001
Age X female sex interaction term		
0-14	-0.13038	0.0002
15-24	0.24562	<.0001

Covariate	Coefficient	P-value
25-44	0.12877	<.0001
45-59	0.03139	<.0001
60-74	Reference	
75+	-0.00664	0.0685

*The diabetes indicator includes all diabetes comorbidities on CMS-2728 and diabetes as cause of ESRD

Table 2b. Prevalent Comorbidity Coefficients, Data Years 2010–2013.

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Sarcoidosis	135	0.0624	<.0001
Malign neopl prostate	185	-0.03133	<.0001
Malign neopl thyroid	193	-0.04837	0.0087
Oth severe malnutrition	262	0.0382	<.0001
Chr airway obstruct NEC	496	0.1908	<.0001
Postinflam pulm fibrosis	515	0.11769	<.0001
Malignant neopl rectum	1541	0.1335	<.0001
Mal neo liver, primary	1550	0.12225	<.0001
Mal neo upper lobe lung	1623	0.08088	<.0001
Mal neo bronch/lung NOS	1629	0.13617	<.0001
Malig neo bladder NOS	1889	0.10792	<.0001
Malig neopl kidney	1890	0.02548	0.0004
Secondary malig neo lung	1970	0.17282	<.0001
Second malig neo liver	1977	0.38071	<.0001
Secondary malig neo bone	1985	0.29043	<.0001
Malignant neoplasm NOS	1991	0.13518	<.0001
Protein-cal malnutr NOS	2639	0.10345	<.0001
Dis urea cycle metabol	2706	0.06036	0.0002
Senile dementia uncomp	2900	-0.02563	0.0001
Drug withdrawal	2920	0.26748	<.0001
Mental disor NEC oth dis	2948	0.04058	<.0001
Cereb degeneration NOS	3319	0.08582	<.0001
Aut neuropthy in oth dis	3371	0.02621	<.0001
Grand mal status	3453	0.01548	0.1722
Anoxic brain damage	3481	-0.03408	0.0008
Cerebral edema	3485	0.09181	<.0001
Idio periph neurpthy NOS	3569	0.09859	<.0001
Neuropathy in diabetes	3572	0.04133	<.0001
Intermed coronary synd	4111	0.2052	<.0001
Angina pectoris NEC/NOS	4139	0.12568	<.0001
Prim pulm hypertension	4160	-0.01251	0.0316
Chr pulmon heart dis NEC	4168	0.15189	<.0001
Prim cardiomyopathy NEC	4254	0.16394	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Cardiomyopath in oth dis	4258	0.16331	<.0001
Atriovent block complete	4260	0.02671	0.0001
Parox ventric tachycard	4271	0.09607	<.0001
Parox tachycardia NOS	4272	0.06145	<.0001
Subdural hemorrhage	4321	0.03408	0.0004
Aortic atherosclerosis	4400	0.09852	<.0001
Lower extremity aneurysm	4423	0.10898	<.0001
Periph vascular dis NOS	4439	0.09731	<.0001
Stricture of artery	4471	0.00238	0.6534
Oth inf vena cava thromb	4532	0.2153	<.0001
Emphysema NEC	4928	0.05787	<.0001
Bronchiectas w/o ac exac	4940	0.06175	<.0001
Food/vomit pneumonitis	5070	0.05726	<.0001
Lung involv in oth dis	5178	0.17403	<.0001
Regional enteritis NOS	5559	0.17154	<.0001
Ulceratve colitis unspcf	5569	0.06821	<.0001
Chr vasc insuff intest	5571	0.15765	<.0001
Paralytic ileus	5601	0.10245	<.0001
Intestinal obstruct NOS	5609	0.10671	<.0001
Alcohol cirrhosis liver	5712	0.05621	<.0001
Cirrhosis of liver NOS	5715	0.20344	<.0001
Hepatic encephalopathy	5722	0.17945	<.0001
Portal hypertension	5723	0.20086	<.0001
Oth sequela, chr liv dis	5728	0.14523	<.0001
Chronic pancreatitis	5771	0.38153	<.0001
Chronic skin ulcer NEC	7078	0.07843	<.0001
Syst lupus erythematosus	7100	0.24781	<.0001
Systemic sclerosis	7101	0.12899	<.0001
Rheumatoid arthritis	7140	0.10921	<.0001
Inflamm polyarthrop NOS	7149	0.02641	0.1369
Sacroiliitis NEC	7202	0.16649	<.0001
Gangrene	7854	0.05466	<.0001
Cachexia	7994	0.14375	<.0001
Fracture of pubis-closed	8082	0.06248	<.0001
Pelvic fracture NOS-clos	8088	-0.01048	0.4819
Fx neck of femur NOS-cl	8208	-0.02685	<.0001
Amput below knee, unilat	8970	-0.10393	<.0001
Amputat bk, unilat-compl	8971	-0.10582	<.0001
Amput above knee, unilat	8972	-0.08573	<.0001
Amputat leg, unilat NOS	8974	-0.077	<.0001
Candidal esophagitis	11284	0.1985	<.0001
Oth lymp unsp xtrndl org	20280	0.14363	<.0001
Mult mye w/o achv rmson	20300	0.19204	<.0001
Ch lym leuk wo achv rmsn	20410	0.25565	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Essntial thrombocythemia	23871	0.10421	<.0001
Low grde myelody syn les	23872	0.14376	<.0001
Myelodysplastic synd NOS	23875	0.17806	<.0001
DMII wo cmp nt st uncntr	25000	0.11986	<.0001
DMII wo cmp uncntrld	25002	0.02111	<.0001
DMII keto nt st uncntrld	25010	0.03729	<.0001
DMII ketoacd uncontrold	25012	0.13424	<.0001
DMI ketoacd uncontrold	25013	0.25355	<.0001
DMII hprosmlr uncontrold	25022	0.12376	<.0001
DMII renl nt st uncntrld	25040	0.0746	<.0001
DMI renl nt st uncntrld	25041	0.04644	<.0001
DMII ophth nt st uncntrl	25050	0.00743	0.0064
DMI ophth uncntrld	25053	0.05823	<.0001
DMII neuro nt st uncntrl	25060	0.05824	<.0001
DMI neuro nt st uncntrld	25061	0.04909	<.0001
DMII neuro uncntrld	25062	0.07612	<.0001
DMI neuro uncntrld	25063	0.13715	<.0001
DMII circ nt st uncntrld	25070	-0.04017	<.0001
DMI circ nt st uncntrld	25071	-0.05298	<.0001
DMII circ uncntrld	25072	-0.02251	<.0001
DMII oth nt st uncntrld	25080	0.08205	<.0001
DMI oth nt st uncntrld	25081	0.02286	0.0002
DMII oth uncntrld	25082	0.03781	<.0001
DMI oth uncntrld	25083	0.00729	0.3939
Glucocorticoid deficient	25541	0.17576	<.0001
Amyloidosis NEC	27739	0.15827	<.0001
Metabolism disorder NEC	27789	0.21983	<.0001
Morbid obesity	27801	0.07927	<.0001
Obesity hypovent synd	27803	-0.05432	<.0001
Sickle cell disease NOS	28260	0.71791	<.0001
Antin chemo indcd pancyt	28411	0.10449	0.0005
Other pancytopenia	28419	0.1945	<.0001
Neutropenia NOS	28800	0.16551	<.0001
Drug induced neutropenia	28803	0.14431	<.0001
Prim hypercoagulable st	28981	0.18562	<.0001
Senile delusion	29020	-0.11382	<.0001
Vascular dementia,uncomp	29040	-0.00174	0.8249
Dementia w/o behav dist	29410	0.01212	0.0613
Dementia w behavior dist	29411	-0.02334	0.0177
Demen NOS w/o behv dstrb	29420	0.04516	<.0001
Schizophrenia NOS-unspec	29590	0.15532	<.0001
Depress psychosis-unspec	29620	0.17524	<.0001
Recurr depr psychos-unsp	29630	0.08526	<.0001
Recur depr psych-severe	29633	0.07789	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Bipolar disorder NOS	29680	0.19198	<.0001
Bipolar disorder NEC	29689	0.08524	<.0001
Episodic mood disord NOS	29690	0.07786	<.0001
Alcoh dep NEC/NOS-unspec	30390	0.16788	<.0001
Alcoh dep NEC/NOS-remiss	30393	0.07322	<.0001
Opioid dependence-unspec	30400	0.25245	<.0001
Opioid dependence-contin	30401	0.18003	<.0001
Drug depend NOS-unspec	30490	0.27902	<.0001
Psymotr epil w/o int epi	34540	-0.08114	<.0001
Epilep NOS w/o intr epil	34590	0.19176	<.0001
Critical illness myopthy	35981	-0.09196	<.0001
Prolif diab retinopathy	36202	-0.08631	<.0001
Mod nonprolf db retinoph	36205	-0.07697	<.0001
Diabetic macular edema	36207	-0.0601	<.0001
Hyp ht dis NOS w ht fail	40291	0.03839	<.0001
Subendo infarct, initial	41071	0.18348	<.0001
AMI NEC, unspecified	41080	0.03986	0.0367
AMI NOS, unspecified	41090	-0.03149	<.0001
Ac ischemic hrt dis NEC	41189	0.11644	<.0001
Pulm embol/infarct NEC	41519	0.13237	<.0001
Atrial fibrillation	42731	0.13302	<.0001
Atrial flutter	42732	0.08346	<.0001
Sinoatrial node dysfunct	42781	-0.00923	0.0206
Crbl emblsm w infrct	43411	0.01754	0.0772
Crbl art ocl NOS w infrc	43491	0.07113	<.0001
Athscl extrm ntv art NOS	44020	0.00141	0.6632
Ath ext ntv at w claudct	44021	0.04379	<.0001
Ath ext ntv at w rst pn	44022	0.09607	<.0001
Ath ext ntv art ulcrtion	44023	0.02268	<.0001
Dsct of thoracic aorta	44101	0.23712	<.0001
Periph vascular dis NEC	44389	0.01881	0.0012
Deep phlebitis-leg NEC	45119	0.00269	0.7906
Ac DVT/emb prox low ext	45341	0.12676	<.0001
Ch DVT/embl low ext NOS	45350	0.12558	<.0001
Ch DVT/embl prox low ext	45351	0.09937	<.0001
Ch emblsm subclav veins	45375	0.17741	<.0001
Ac DVT/embl up ext	45382	0.08862	<.0001
Ac emblsm axillary veins	45384	0.10835	<.0001
Ac embl internl jug vein	45386	0.16307	<.0001
Ac embl thorac vein NEC	45387	0.13445	<.0001
Esoph varice oth dis NOS	45621	0.19764	<.0001
Obs chr bronc w(ac) exac	49121	0.16393	<.0001
Obs chr bronc w ac bronc	49122	0.11419	<.0001
Chronic obst asthma NOS	49320	0.10527	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
Ch obst asth w (ac) exac	49322	0.10999	<.0001
Ac resp flr fol trma/srg	51851	-0.04255	0.0003
Ot pul insuf fol trm/srg	51852	-0.0827	0.0003
Other pulmonary insuff	51882	0.13098	<.0001
Chronic respiratory fail	51883	0.0293	<.0001
Acute & chronc resp fail	51884	0.02507	<.0001
Gastrostomy comp - mech	53642	0.10042	<.0001
Fecal impaction	56032	0.09744	<.0001
Pressure ulcer, low back	70703	0.0362	<.0001
Pressure ulcer, hip	70704	0.09173	<.0001
Pressure ulcer, buttock	70705	0.00396	0.4043
Ulcer of lower limb NOS	70710	0.01138	0.0098
Ulcer other part of foot	70715	0.04066	<.0001
Ulcer oth part low limb	70719	0.03358	<.0001
Pyogen arthritis-unspec	71100	0.03922	0.0151
Pyogen arthritis-l/leg	71106	0.11218	<.0001
Ac osteomyelitis-unspec	73000	-0.04005	0.0005
Ac osteomyelitis-ankle	73007	-0.03799	<.0001
Ac osteomyelitis NEC	73008	-0.01851	0.102
Osteomyelitis NOS-hand	73024	0.05835	0.0001
Osteomyelitis NOS-ankle	73027	-0.03107	<.0001
Path fx vertebrae	73313	0.1329	<.0001
Aseptic necrosis femur	73342	0.20291	<.0001
Asept necrosis bone NEC	73349	0.17431	<.0001
Coma	78001	0.02143	0.1083
Convulsions NEC	78039	0.10277	<.0001
Fx femur intrcaps NEC-cl	82009	0.03652	0.0079
Fx femur NOS-closed	82100	-0.05632	<.0001
React-indwell urin cath	99664	0.15093	<.0001
Compl heart transplant	99683	0.02305	0.3552
Asymp hiv infectn status	V08	0.37403	<.0001
Heart transplant status	V421	0.26702	<.0001
Liver transplant status	V427	0.16234	<.0001
Trnspl status-pancreas	V4283	0.14978	<.0001
Gastrostomy status	V441	0.02184	0.0173
Ileostomy status	V442	0.12312	<.0001
Colostomy status	V443	0.13378	<.0001
Urinostomy status NEC	V446	0.33981	<.0001
Respirator depend status	V4611	-0.02597	0.001
Status amput othr toe(s)	V4972	0.031	<.0001
Status amput below knee	V4975	0.02473	<.0001
Status amput above knee	V4976	0.01774	0.0036
Atten to gastrostomy	V551	-0.03053	0.0012
Long-term use of insulin	V5867	0.12534	<.0001

ICD-9 Description	ICD-9 Code	Coefficient	P-value
BMI 40.0-44.9, adult	V8541	0.03116	<.0001
Less than 6 months of Medicare eligible claims in the previous calendar year		0.73799	<.0001

Most of the coefficient estimates for the prevalent comorbidities are positive and statistically significant, but several do not obtain statistical significance. The very large number of clinical factors in the model expectedly generates substantial multicollinearity among the covariates, likely resulting in some unexpected results in the direction of the coefficient sign and levels of statistical significance. Inclusion of this set of prevalent comorbidities reflects the consensus of the TEP that adjustment for all of these prevalent comorbidities, in addition to incident comorbidities, is important to reflect the current health condition of the patient in risk adjustment.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

The tables below show the parameter estimates for patient- and area-level SDS/SES variables based on a Cox model for hospital admissions that included these variables along with the original covariates adjusted for in SHR.

Table 3a. Comparing coefficients between sensitivity models with and without SDS/SES adjustors, 2010-2013: Model coefficients

	Baseline SHR			SDS/SES-adjusted SHR	
Covariate	**Coefficient**	**P-value**		**Coefficient**	**P-value**
Medicare coverage*					
Medicare primary + Medicaid	NA	NA		0.07628	<.0001
Medicare primary + no Medicaid	NA	NA		Reference	-
Medicare secondary/HMO	NA	NA		0.97671	<.0001
Employment status 6 months prior to ESRD					
Unemployed	NA	NA		Reference	-
Employed	NA	NA		0.05164	<.0001
Other/Unknown	NA	NA		0.02001	<.0001
Race					
White	NA	NA		Reference	-
Native American/Alaskan Native	NA	NA		-0.03346	<.0001
Asian/Pacific Islander	NA	NA		-0.20491	<.0001
Black	NA	NA		-0.06702	<.0001
Other/Unknown	NA	NA		0.01116	0.1526
Ethnicity					
Hispanic	NA	NA		-0.08082	<.0001
Non-Hispanic	NA	NA		Reference	-
Unknown	NA	NA		-0.05751	<.0001
ADI element					
Home value (median)	NA	NA		0.00208	0.2466
Family income (median)	NA	NA		-0.00197	0.0188
Income disparity**	NA	NA		-0.00118	0.0428
Monthly mortgage (median)	NA	NA		0.00029	0.9517
< 9 years of education (%)	NA	NA		-0.00124	<.0001
No high school diploma (%)	NA	NA		0.00186	<.0001
Home ownership rate (%)	NA	NA		-0.00056	<.0001
Families below the poverty level (%)	NA	NA		0.00061	0.0019
Gross rent (median)	NA	NA		0.01567	0.0081

Covariate	Baseline SHR			SDS/SES-adjusted SHR	
	Coefficient	**P-value**		**Coefficient**	**P-value**
Single-parent households with children <18 (%)	NA	NA		-0.00017	0.2071
Unemployment rate	NA	NA		0.00157	0.0001
Comorbidities at start of ESRD					
At least one of the comorbidities listed below	0.08624	<.0001		0.07638	<.0001
Atherosclerotic heart disease	0.04999	<.0001		0.04289	<.0001
Other cardiac disease	0.04395	<.0001		0.03238	<.0001
Diabetes***	-0.02026	<.0001		-0.04055	<.0001
Congestive heart failure	0.04269	<.0001		0.03675	<.0001
Inability to ambulate	0.02042	<.0001		0.01702	<.0001
Chronic obstructive pulmonary disease	0.05646	<.0001		0.04056	<.0001
Inability to transfer	0.02401	<.0001		0.02181	0.0002
Malignant neoplasm, cancer	0.04102	<.0001		0.03391	<.0001
Peripheral vascular disease	0.04104	<.0001		0.02916	<.0001
Cerebrovascular disease, CVA, TIA	0.01904	<.0001		0.01454	<.0001
Tobacco use (current smoker)	0.08539	<.0001		0.08095	<.0001
Alcohol dependence	0.01285	0.036		0.01570	0.0105
Drug dependence	0.17361	<.0001		0.17165	<.0001
No Medical Evidence (CMS-2728) Form	0.15316	<.0001		0.17504	<.0001
Cause of ESRD					
Diabetes	0.03848	<.0001		0.03011	<.0001
Missing	-0.03547	<.0001		-0.04048	<.0001
Sex: Female	0.07156	<.0001		0.06285	<.0001
Age					
0-14	0.48884	<.0001		0.49754	<.0001
15-24	0.13135	<.0001		0.17018	<.0001
25-44	-0.0678	<.0001		-0.02533	<.0001
45-59	-0.065	<.0001		-0.03439	<.0001
60-74	Reference	-		Reference	-
75+	0.10178	<.0001		0.07273	<.0001
BMI					
Log BMI	-0.15032	<.0001		-0.16225	<.0001
BMI missing	0.01656	0.0002		0.01456	0.0064
Calendar year					
2010	Reference	-		Reference	-
2011	-0.02546	<.0001		-0.02546	<.0001
2012	-0.12676	<.0001		-0.12349	<.0001
2013	-0.16265	<.0001		-0.16155	<.0001
In nursing home the previous year	0.20788	<.0001		0.17739	<.0001
Diabetes as cause of ESRD X time on ESRD interaction term					
91 days-6 months	Reference	-		Reference	-
6 months-1 year	0.03417	<.0001		0.02973	<.0001
1-2 years	0.01166	0.0737		0.00827	0.2049
2-3 years	0.00139	0.8356		0.00004	0.9954
3-5 years	-0.01549	0.0147		-0.01139	0.073
5+ years	-0.06398	<.0001		-0.05036	<.0001
Cause of ESRD: diabetes X sex: female interaction term	-0.02622	<.0001		-0.02295	<.0001
Age X diabetes as cause of ESRD interaction term					
0-14	-0.93749	<.0001		-0.87713	0.0003
15-24	0.16727	<.0001		0.17698	<.0001
25-44	0.15502	<.0001		0.15213	<.0001
45-59	0.05013	<.0001		0.04798	<.0001
60-74	Reference	-		Reference	-
75+	-0.03426	<.0001		-0.03067	<.0001
Age X female sex interaction term					
0-14	-0.13038	0.0002		-0.11088	0.0019

Covariate	Baseline SHR			SDS/SES-adjusted SHR	
	Coefficient	P-value		Coefficient	P-value
15-24	0.24562	<.0001		0.24326	<.0001
25-44	0.12877	<.0001		0.12323	<.0001
45-59	0.03139	<.0001		0.02849	<.0001
60-74	Reference	-		Reference	-
75+	-0.00664	0.0685		-0.00662	0.0696

*Patients without Medicare coverage or with unknown coverage type were excluded from the model.
**Log(100)*(the ratio of the number of households with less than $10,000 in income to the number of households with $50,000 or more in income).
***The diabetes indicator includes all diabetes comorbidities on CMS-2728 and diabetes as cause of ESRD.

Table 3b. Comparing coefficients between sensitivity models with and without SDS/SES adjustors, 2010-2013: Prevalent comorbidity coefficients

ICD-9 Description	ICD-9 Code	Baseline SHR			SDS/SES-adjusted SHR	
		Coefficient	P-value		Coefficient	P-value
Protein-cal malnutr NOS	2639	0.10345	<.0001		0.09068	<.0001
Aut neuropthy in oth dis	3371	0.02621	<.0001		0.02174	<.0001
Epilep NOS w/o intr epil	34590	0.19176	<.0001		0.16817	<.0001
Cerebral edema	3485	0.09181	<.0001		0.07959	<.0001
Subendo infarct, initial	41071	0.18348	<.0001		0.14855	<.0001
AMI NEC, unspecified	41080	0.03986	0.0367		0.07768	<.0001
AMI NOS, unspecified	41090	-0.03149	<.0001		0.01671	0.0021
Intermed coronary synd	4111	0.2052	<.0001		0.20521	<.0001
Ac ischemic hrt dis NEC	41189	0.11644	<.0001		0.11839	<.0001
Angina pectoris NEC/NOS	4139	0.12568	<.0001		0.1392	<.0001
Cardiomyopath in oth dis	4258	0.16331	<.0001		0.16447	<.0001
Atriovent block complete	4260	0.02671	0.0001		0.03722	<.0001
Parox ventric tachycard	4271	0.09607	<.0001		0.09379	<.0001
Parox tachycardia NOS	4272	0.06145	<.0001		0.07383	<.0001
Atrial fibrillation	42731	0.13302	<.0001		0.13334	<.0001
Atrial flutter	42732	0.08346	<.0001		0.07437	<.0001
Sinoatrial node dysfunct	42781	-0.00923	0.0206		0.01865	<.0001
Subdural hemorrhage	4321	0.03408	0.0004		0.04615	<.0001
Stricture of artery	4471	0.00238	0.6534		0.02688	<.0001
Paralytic ileus	5601	0.10245	<.0001		0.09073	<.0001
Convulsions NEC	78039	0.10277	<.0001		0.11375	<.0001
Gangrene	7854	0.05466	<.0001		0.04253	<.0001
Cachexia	7994	0.14375	<.0001		0.13784	<.0001
Candidal esophagitis	11284	0.1985	<.0001		0.18944	<.0001
Sarcoidosis	135	0.0624	<.0001		0.05333	<.0001
Malignant neopl rectum	1541	0.1335	<.0001		0.1436	<.0001
Mal neo liver, primary	1550	0.12225	<.0001		0.12933	<.0001
Mal neo upper lobe lung	1623	0.08088	<.0001		0.07581	<.0001
Mal neo bronch/lung NOS	1629	0.13617	<.0001		0.15539	<.0001
Malign neopl prostate	185	-0.03133	<.0001		0.00491	0.4173
Malig neo bladder NOS	1889	0.10792	<.0001		0.12933	<.0001
Malig neopl kidney	1890	0.02548	0.0004		0.04364	<.0001
Malign neopl thyroid	193	-0.04837	0.0087		-0.02906	0.1153
Secondary malig neo lung	1970	0.17282	<.0001		0.15946	<.0001
Second malig neo liver	1977	0.38071	<.0001		0.3608	<.0001
Secondary malig neo bone	1985	0.29043	<.0001		0.29427	<.0001
Malignant neoplasm NOS	1991	0.13518	<.0001		0.14138	<.0001
Oth lymp unsp xtrndl org	20280	0.14363	<.0001		0.1379	<.0001
Mult mye w/o achv rmson	20300	0.19204	<.0001		0.19396	<.0001
Ch lym leuk wo achv rmsn	20410	0.25565	<.0001		0.23055	<.0001
Essntial thrombocythemia	23871	0.10421	<.0001		0.09762	<.0001

		Baseline SHR			SDS/SES-adjusted SHR	
ICD-9 Description	ICD-9 Code	Coefficient	P-value		Coefficient	P-value
Low grde myelody syn les	23872	0.14376	<.0001		0.16016	<.0001
Myelodysplastic synd NOS	23875	0.17806	<.0001		0.17918	<.0001
DMII wo cmp nt st uncntr	25000	0.11986	<.0001		0.15129	<.0001
DMII wo cmp uncntrld	25002	0.02111	<.0001		0.04779	<.0001
DMII keto nt st uncntrld	25010	0.03729	<.0001		0.08276	<.0001
DMII ketoacd uncontrold	25012	0.13424	<.0001		0.11517	<.0001
DMI ketoacd uncontrold	25013	0.25355	<.0001		0.20779	<.0001
DMII hprosmlr uncontrold	25022	0.12376	<.0001		0.10357	<.0001
DMII renl nt st uncntrld	25040	0.0746	<.0001		0.07666	<.0001
DMI renl nt st uncntrld	25041	0.04644	<.0001		0.052	<.0001
DMII ophth nt st uncntrl	25050	0.00743	0.0064		0.00591	0.0305
DMI ophth uncntrld	25053	0.05823	<.0001		0.04352	<.0001
DMII neuro nt st uncntrl	25060	0.05824	<.0001		0.06459	<.0001
DMI neuro nt st uncntrld	25061	0.04909	<.0001		0.05464	<.0001
DMII neuro uncntrld	25062	0.07612	<.0001		0.07231	<.0001
DMI neuro uncntrld	25063	0.13715	<.0001		0.12346	<.0001
DMII circ nt st uncntrld	25070	-0.04017	<.0001		-0.02883	<.0001
DMI circ nt st uncntrld	25071	-0.05298	<.0001		-0.03436	<.0001
DMII circ uncntrld	25072	-0.02251	<.0001		-0.01743	0.0015
DMII oth nt st uncntrld	25080	0.08205	<.0001		0.07395	<.0001
DMI oth nt st uncntrld	25081	0.02286	0.0002		0.02003	0.0012
DMII oth uncntrld	25082	0.03781	<.0001		0.03026	<.0001
DMI oth uncntrld	25083	0.00729	0.3939		0.00901	0.2922
Glucocorticoid deficient	25541	0.17576	<.0001		0.16647	<.0001
Oth severe malnutrition	262	0.0382	<.0001		0.02159	0.0003
Dis urea cycle metabol	2706	0.06036	0.0002		0.06852	<.0001
Amyloidosis NEC	27739	0.15827	<.0001		0.14513	<.0001
Metabolism disorder NEC	27789	0.21983	<.0001		0.21405	<.0001
Morbid obesity	27801	0.07927	<.0001		0.06141	<.0001
Obesity hypovent synd	27803	-0.05432	<.0001		-0.06425	<.0001
Sickle cell disease NOS	28260	0.71791	<.0001		0.69038	<.0001
Antin chemo indcd pancyt	28411	0.10449	0.0005		0.08143	0.007
Other pancytopenia	28419	0.1945	<.0001		0.18252	<.0001
Neutropenia NOS	28800	0.16551	<.0001		0.1658	<.0001
Drug induced neutropenia	28803	0.14431	<.0001		0.14311	<.0001
Prim hypercoagulable st	28981	0.18562	<.0001		0.17246	<.0001
Senile dementia uncomp	2900	-0.02563	0.0001		0.00253	0.708
Senile delusion	29020	-0.11382	<.0001		-0.0962	<.0001
Vascular dementia,uncomp	29040	-0.00174	0.8249		0.00329	0.6754
Drug withdrawal	2920	0.26748	<.0001		0.2474	<.0001
Dementia w/o behav dist	29410	0.01212	0.0613		0.02147	0.0009
Dementia w behavior dist	29411	-0.02334	0.0177		-0.00281	0.7757
Demen NOS w/o behv dstrb	29420	0.04516	<.0001		0.04207	<.0001
Mental disor NEC oth dis	2948	0.04058	<.0001		0.0466	<.0001
Schizophrenia NOS-unspec	29590	0.15532	<.0001		0.15092	<.0001
Depress psychosis-unspec	29620	0.17524	<.0001		0.1634	<.0001
Recurr depr psychos-unsp	29630	0.08526	<.0001		0.0741	<.0001
Recur depr psych-severe	29633	0.07789	<.0001		0.08623	<.0001
Bipolar disorder NOS	29680	0.19198	<.0001		0.16867	<.0001
Bipolar disorder NEC	29689	0.08524	<.0001		0.08315	<.0001
Episodic mood disord NOS	29690	0.07786	<.0001		0.0807	<.0001
Alcoh dep NEC/NOS-unspec	30390	0.16788	<.0001		0.15674	<.0001
Alcoh dep NEC/NOS-remiss	30393	0.07322	<.0001		0.05354	<.0001
Opioid dependence-unspec	30400	0.25245	<.0001		0.23688	<.0001
Opioid dependence-contin	30401	0.18003	<.0001		0.1673	<.0001
Drug depend NOS-unspec	30490	0.27902	<.0001		0.27214	<.0001
Cereb degeneration NOS	3319	0.08582	<.0001		0.11595	<.0001
Grand mal status	3453	0.01548	0.1722		0.01564	0.1675
Psymotr epil w/o int epi	34540	-0.08114	<.0001		-0.06901	<.0001

		Baseline SHR			SDS/SES-adjusted SHR	
ICD-9 Description	ICD-9 Code	Coefficient	P-value		Coefficient	P-value
Anoxic brain damage	3481	-0.03408	0.0008		-0.03967	0.0001
Idio periph neurpthy NOS	3569	0.09859	<.0001		0.10174	<.0001
Neuropathy in diabetes	3572	0.04133	<.0001		0.02274	<.0001
Critical illness myopthy	35981	-0.09196	<.0001		-0.08218	<.0001
Prolif diab retinopathy	36202	-0.08631	<.0001		-0.06471	<.0001
Mod nonprolf db retinoph	36205	-0.07697	<.0001		-0.0567	<.0001
Diabetic macular edema	36207	-0.0601	<.0001		-0.04416	<.0001
Hyp ht dis NOS w ht fail	40291	0.03839	<.0001		0.05711	<.0001
Pulm embol/infarct NEC	41519	0.13237	<.0001		0.13027	<.0001
Prim pulm hypertension	4160	-0.01251	0.0316		0.02908	<.0001
Chr pulmon heart dis NEC	4168	0.15189	<.0001		0.13335	<.0001
Prim cardiomyopathy NEC	4254	0.16394	<.0001		0.15779	<.0001
Crbl emblsm w infrct	43411	0.01754	0.0772		0.01317	0.1847
Crbl art ocl NOS w infrc	43491	0.07113	<.0001		0.07869	<.0001
Aortic atherosclerosis	4400	0.09852	<.0001		0.08793	<.0001
Athscl extrm ntv art NOS	44020	0.00141	0.6632		0.01909	<.0001
Ath ext ntv at w claudct	44021	0.04379	<.0001		0.06012	<.0001
Ath ext ntv at w rst pn	44022	0.09607	<.0001		0.09649	<.0001
Ath ext ntv art ulcrtion	44023	0.02268	<.0001		0.03187	<.0001
Dsct of thoracic aorta	44101	0.23712	<.0001		0.24884	<.0001
Lower extremity aneurysm	4423	0.10898	<.0001		0.10403	<.0001
Periph vascular dis NEC	44389	0.01881	0.0012		0.02819	<.0001
Periph vascular dis NOS	4439	0.09731	<.0001		0.10228	<.0001
Deep phlebitis-leg NEC	45119	0.00269	0.7906		0.03874	0.0001
Oth inf vena cava thromb	4532	0.2153	<.0001		0.20467	<.0001
Ac DVT/emb prox low ext	45341	0.12676	<.0001		0.10691	<.0001
Ch DVT/embl low ext NOS	45350	0.12558	<.0001		0.11544	<.0001
Ch DVT/embl prox low ext	45351	0.09937	<.0001		0.09291	<.0001
Ch emblsm subclav veins	45375	0.17741	<.0001		0.17209	<.0001
Ac DVT/embl up ext	45382	0.08862	<.0001		0.08867	<.0001
Ac emblsm axillary veins	45384	0.10835	<.0001		0.09897	<.0001
Ac embl internl jug vein	45386	0.16307	<.0001		0.15905	<.0001
Ac embl thorac vein NEC	45387	0.13445	<.0001		0.1339	<.0001
Esoph varice oth dis NOS	45621	0.19764	<.0001		0.17113	<.0001
Obs chr bronc w(ac) exac	49121	0.16393	<.0001		0.15724	<.0001
Obs chr bronc w ac bronc	49122	0.11419	<.0001		0.10931	<.0001
Emphysema NEC	4928	0.05787	<.0001		0.07762	<.0001
Chronic obst asthma NOS	49320	0.10527	<.0001		0.10032	<.0001
Ch obst asth w (ac) exac	49322	0.10999	<.0001		0.10446	<.0001
Bronchiectas w/o ac exac	4940	0.06175	<.0001		0.07671	<.0001
Chr airway obstruct NEC	496	0.1908	<.0001		0.18441	<.0001
Food/vomit pneumonitis	5070	0.05726	<.0001		0.04838	<.0001
Postinflam pulm fibrosis	515	0.11769	<.0001		0.12366	<.0001
Lung involv in oth dis	5178	0.17403	<.0001		0.15417	<.0001
Ac resp flr fol trma/srg	51851	-0.04255	0.0003		-0.05125	<.0001
Ot pul insuf fol trm/srg	51852	-0.0827	0.0003		-0.0681	0.0032
Other pulmonary insuff	51882	0.13098	<.0001		0.1543	<.0001
Chronic respiratory fail	51883	0.0293	<.0001		0.0179	0.0021
Acute & chronc resp fail	51884	0.02507	<.0001		0.00683	0.1906
Gastrostomy comp - mech	53642	0.10042	<.0001		0.11609	<.0001
Regional enteritis NOS	5559	0.17154	<.0001		0.14951	<.0001
Ulceratve colitis unspcf	5569	0.06821	<.0001		0.07949	<.0001
Chr vasc insuff intest	5571	0.15765	<.0001		0.14385	<.0001
Fecal impaction	56032	0.09744	<.0001		0.09478	<.0001
Intestinal obstruct NOS	5609	0.10671	<.0001		0.11453	<.0001
Alcohol cirrhosis liver	5712	0.05621	<.0001		0.05224	<.0001
Cirrhosis of liver NOS	5715	0.20344	<.0001		0.20181	<.0001
Hepatic encephalopathy	5722	0.17945	<.0001		0.16256	<.0001
Portal hypertension	5723	0.20086	<.0001		0.18288	<.0001

ICD-9 Description	ICD-9 Code	Baseline SHR			SDS/SES-adjusted SHR	
		Coefficient	P-value		Coefficient	P-value
Oth sequela, chr liv dis	5728	0.14523	<.0001		0.14782	<.0001
Chronic pancreatitis	5771	0.38153	<.0001		0.36579	<.0001
Pressure ulcer, low back	70703	0.0362	<.0001		0.02419	<.0001
Pressure ulcer, hip	70704	0.09173	<.0001		0.09029	<.0001
Pressure ulcer, buttock	70705	0.00396	0.4043		0.0221	<.0001
Ulcer of lower limb NOS	70710	0.01138	0.0098		0.02116	<.0001
Ulcer other part of foot	70715	0.04066	<.0001		0.04168	<.0001
Ulcer oth part low limb	70719	0.03358	<.0001		0.02956	<.0001
Chronic skin ulcer NEC	7078	0.07843	<.0001		0.08132	<.0001
Syst lupus erythematosus	7100	0.24781	<.0001		0.23436	<.0001
Systemic sclerosis	7101	0.12899	<.0001		0.13113	<.0001
Pyogen arthritis-unspec	71100	0.03922	0.0151		0.07424	<.0001
Pyogen arthritis-l/leg	71106	0.11218	<.0001		0.09919	<.0001
Rheumatoid arthritis	7140	0.10921	<.0001		0.10251	<.0001
Inflamm polyarthrop NOS	7149	0.02641	0.1369		0.05225	0.0033
Sacroiliitis NEC	7202	0.16649	<.0001		0.17183	<.0001
Ac osteomyelitis-unspec	73000	-0.04005	0.0005		-0.01211	0.2959
Ac osteomyelitis-ankle	73007	-0.03799	<.0001		-0.02268	0.0005
Ac osteomyelitis NEC	73008	-0.01851	0.102		-0.01646	0.1459
Osteomyelitis NOS-hand	73024	0.05835	0.0001		0.06307	<.0001
Osteomyelitis NOS-ankle	73027	-0.03107	<.0001		-0.04842	<.0001
Path fx vertebrae	73313	0.1329	<.0001		0.1435	<.0001
Aseptic necrosis femur	73342	0.20291	<.0001		0.1894	<.0001
Asept necrosis bone NEC	73349	0.17431	<.0001		0.17243	<.0001
Coma	78001	0.02143	0.1083		0.03361	0.012
Fracture of pubis-closed	8082	0.06248	<.0001		0.04974	<.0001
Pelvic fracture NOS-clos	8088	-0.01048	0.4819		0.02635	0.0755
Fx femur intrcaps NEC-cl	82009	0.03652	0.0079		0.01917	0.1618
Fx neck of femur NOS-cl	8208	-0.02685	<.0001		-0.0007617	0.9099
Fx femur NOS-closed	82100	-0.05632	<.0001		-0.03439	0.0012
Amput below knee, unilat	8970	-0.10393	<.0001		-0.07656	<.0001
Amputat bk, unilat-compl	8971	-0.10582	<.0001		-0.07636	<.0001
Amput above knee, unilat	8972	-0.08573	<.0001		-0.06596	<.0001
Amputat leg, unilat NOS	8974	-0.077	<.0001		-0.05693	0.0017
React-indwell urin cath	99664	0.15093	<.0001		0.12326	<.0001
Compl heart transplant	99683	0.02305	0.3552		0.0336	0.1755
Asymp hiv infectn status	V08	0.37403	<.0001		0.35665	<.0001
Heart transplant status	V421	0.26702	<.0001		0.23506	<.0001
Liver transplant status	V427	0.16234	<.0001		0.13283	<.0001
Trnspl status-pancreas	V4283	0.14978	<.0001		0.10397	<.0001
Gastrostomy status	V441	0.02184	0.0173		0.01005	0.2728
Ileostomy status	V442	0.12312	<.0001		0.1086	<.0001
Colostomy status	V443	0.13378	<.0001		0.12704	<.0001
Urinostomy status NEC	V446	0.33981	<.0001		0.31177	<.0001
Respirator depend status	V4611	-0.02597	0.001		-0.02041	0.0095
Status amput othr toe(s)	V4972	0.031	<.0001		0.02001	<.0001
Status amput below knee	V4975	0.02473	<.0001		0.01286	0.0032
Status amput above knee	V4976	0.01774	0.0036		0.01293	0.034
Atten to gastrostomy	V551	-0.03053	0.0012		-0.01125	0.2309
Long-term use of insulin	V5867	0.12534	<.0001		0.10276	<.0001
BMI 40.0-44.9, adult	V8541	0.03116	<.0001		0.01971	0.0009
Less than 6 months of Medicare eligible claims in the previous calendar year	—	0.73799	<.0001		0.5303	<.0001

Evaluating Adjustments for SDS/SES

Figure 1. Comparison of SHRs adjusted and not adjusted for race by facility percentage of black patients (deciles), 2013

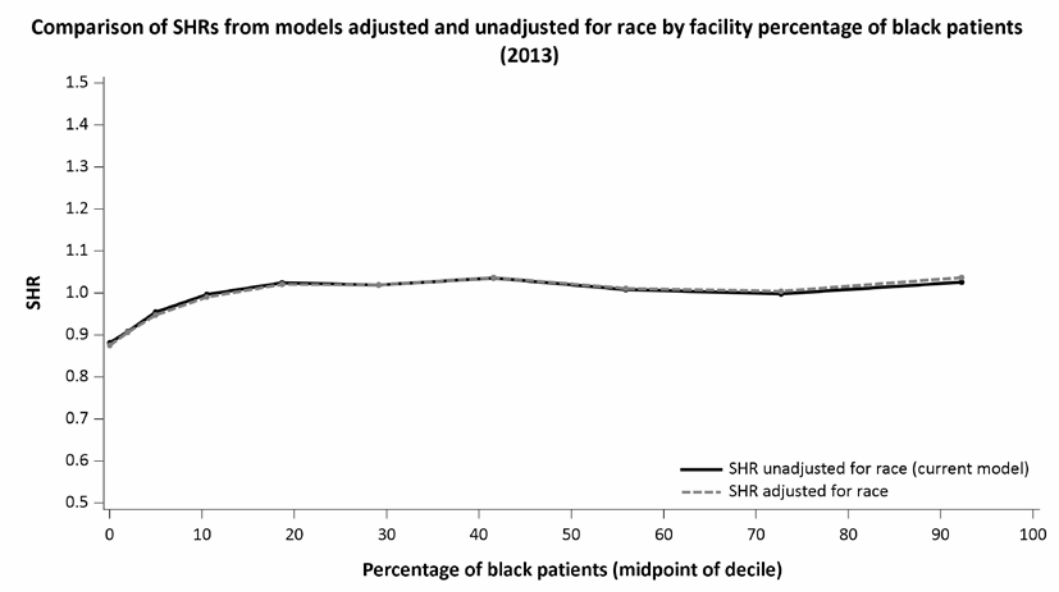


Figure 2. Comparison of SHRs adjusted and not adjusted for Hispanic ethnicity by facility percentage of Hispanic patients (deciles), 2013

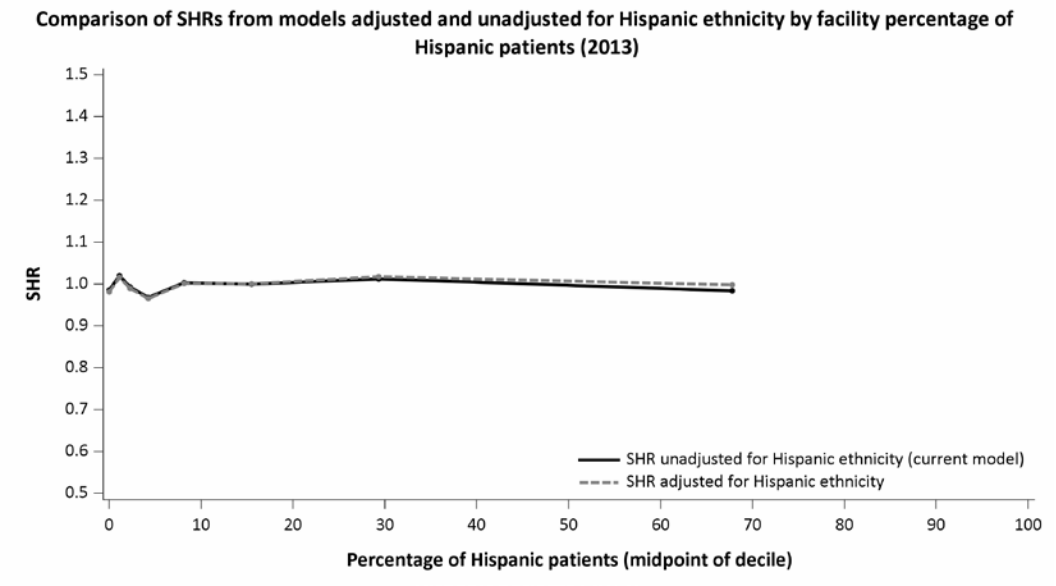
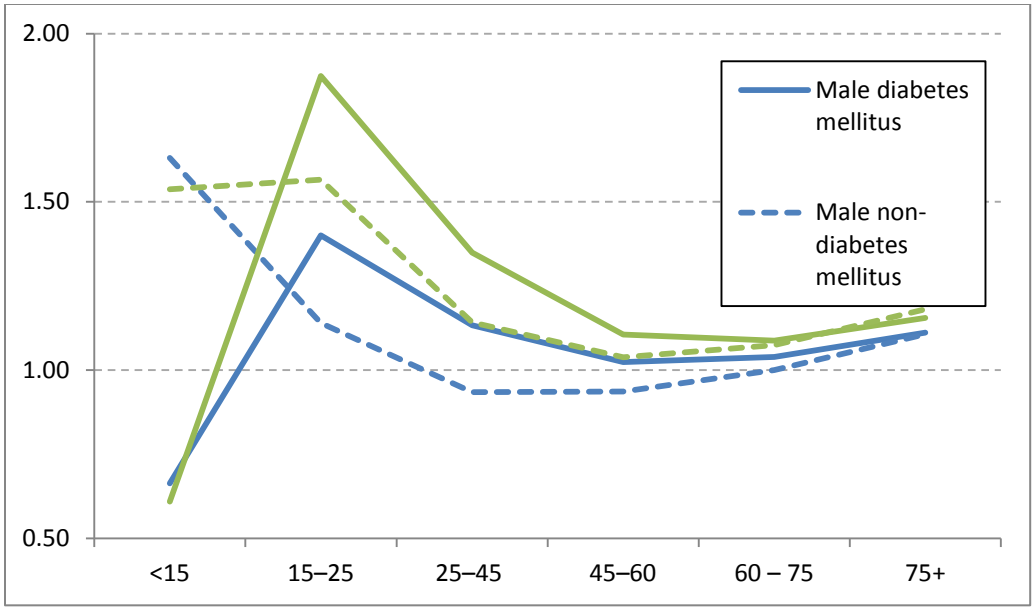


Figure 3. Relative effects of coefficients related to sex in the 2013 SHR model



Patient-level SDS: Compared with males, females were more likely to experience a hospital admission (OR=1.06; p<0.01).However the interaction of female sex and age demonstrated the highest odds were observed in the age 15 – 24, 25-44, and 45-59 age groups, with a decreasing gradient, and the 45-59 age group showing the most diminished impact. There was no significant difference in the oldest female-age-specific group. These results suggest the possibility of an unidentified biologic effect or, alternatively, confounding by an unmeasured association for younger females. Hispanics were less likely to be admitted to the hospital (OR=0.92; p<0.01) than non- Hispanics. Compared with white patients, Asian/PI (OR=0.81, p<0.01), Native American (OR=0.97, p<0.01) and black (OR=0.94, p<0.01) patients were less likely to be admitted to the hospital. The results for ethnicity and race are consistent with prior studies within the dialysis setting.

Patient-level SES: Compared with Medicare-only patients, patients with both Medicare and Medicaid (OR=1.08; p<0.01) and patients with Medicare as secondary/Medicare HMO (OR=2.66, p<0.01) were more likely to be hospitalized. The result for dually eligible patients having higher odds of hospitalization is consistent with the hypothesis that this insurance category, on average, represents an at-risk group. Further examination is needed for the higher odds of hospitalization for patients with Medicare as secondary payer or HMO. It is possible that these patients represent a larger portion of incident ESRD patients, which have a known higher risk of complications in the first year of ESRD.

Patients who were employed prior to ESRD incidence were more likely to be admitted to the hospital (OR=1.05; p<0.01) than unemployed patients. Note that for employment categories, the “Other/Unknown” category also had higher odds of hospital admission. We note this represents diverse patient groups with regard to SES, such as students, homemakers and those who are retired. The higher odds of hospitalization may be associated with unmeasured risk characteristics of this diverse group but that will require further empirical examination based on data availability.

Area-level SES: Overall, measures of area-level deprivation had very low impact on the odds of hospitalization. Among statistically significant impacts were measures of low median family income (OR=0.998, p=0.0188), the percentage of families below the poverty level (OR=1.001, p=0.002), the percentage of individuals without a high school diploma (OR=1.002, p<0.01), and the area-level unemployment rate (OR=1.002, p<0.01). In general the magnitude of the effects of the individual indicators was very small. In addition to the very small coefficients, a few were not in the expected direction suggesting potential collinearity with other SES or SDS factors in the model.

Table 4. Flagging rates, by model with and without all SDS/SES adjustors: 2010-2013

Baseline SHR	Model With SDS/SES			Total
	Better than Expected	As Expected	Worse than Expected	
Better than Expected	166	21	3	**190 (3.1%)**
As Expected	45	5546	81	**5672 (91.0%)**
Worse than Expected	5	123	244	**372 (6.0%)**
Total	**216 (3.5%)**	**5690 (91.3%)**	**328 (5.3%)**	—

After adjustment for SDS/SES, 278 facilities (4.5%) changed performance categories. 105 (1.7%) facilities were down-graded, and 173 (2.8%) were upgraded.

These analyses indicate that select patient-level variables for SDS/SES affect expected hospitalization rates, while area-level indicators had either minimal or no effect on expected hospital admissions. Furthermore, SHRs with and without adjustment for SDS/SES are highly correlated (0.9109) but adjustment for SDS/SES shifts facility performance only slightly. This suggests SDS/SES does not contribute much to the flagging profiles for facility performance.

In the final SHR model we continue to include sex (SDS factor) for risk adjustment. Our analysis of medical evidence and claims data is generally supportive of the current approach to sex adjustment in the SHR. It is consistent with the

consensus opinion that adjustment for sex is appropriate, in that there is some evidence of physiological cause for higher hospitalization rates among females.

Table 3a above presents the manner in which the SHR adjusts for sex, given current judgment that physiology accounts for some, if not a substantial part, of observed differences in hospitalization by sex. The main adjustment reflects the observation that, adjusting for age and a set of comorbidities, females are more likely to be hospitalized. The interaction terms for age and sex in the model indicate that the effect of sex depends substantially on patient age. Females in the 15-45 age range face a greater risk of experiencing an admission, as compared to men of the same age with similar risk profiles. This does not appear to be a consequence of facility performance, however, because the disparity is not generally applicable to females, but only to a limited age group. It is therefore important to risk adjust for sex to ensure that women in facilities with larger numbers of women aged 15 to 45 are not inappropriately disadvantaged in terms of access to care.

Figure 3 shows the interaction of age and sex in the SHR model, for patients diagnosed with and without diabetes. The figure makes clear that for both male and female patients, independent of diagnoses of diabetes, hospitalization is strongly associated with young age. Further, the male-female difference is concentrated in the younger age categories. Beyond age 45, where the hospitalization rates are generally quite low, there is very little difference between males and females. The figure also demonstrates that high hospitalization rates for females reflects utilization by younger females, suggesting a physiologic effect rather than a systematic difference in care by sex.

Race, ethnicity and patient level SES factors are not included in the final risk adjusted model. While adjustment for these factors would account for different outcomes by race and ethnicity and SES factors and guard against barriers in access to care, adjustment would also introduce the potential unintended consequence of allowing access to lower quality of care. Additionally, race and Hispanic ethnicity were observed to indicate lower risk of hospitalization, including race, Hispanic ethnicity did not contribute more to the SHR compared to a model with most of the current set of adjustors; similarly for socioeconomic status (Figures 1-2 above). We are currently examining other measures of SES and SDS to assess impact on expected hospitalization and whether it would be appropriate to adjust for these factors.

Given the very small impact of area-level SES factors we decided not to include these as risk adjustments in the final model. While other studies have shown the association between these patient and area-level SDS/SES factors and hospitalization, further work is needed to demonstrate that differences based on these factors are not related to facility care, in order to prevent disparities in care. Patients in lower SES strata are typically in poorer health as they face greater resource limitation as a result of their limited access to primary care. Adjusting for SES would effectively further comprise the quality of care received as it would lower standards of care based on an assumption these patients will just generally always be sicker.

2b4.5. 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 2b4.9

Two-way interactions were examined and selected for the final model based on both the magnitude and statistical significance of the estimates.

2b4.6. 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 estimate of the c-statistic for the SHR is 0.65.

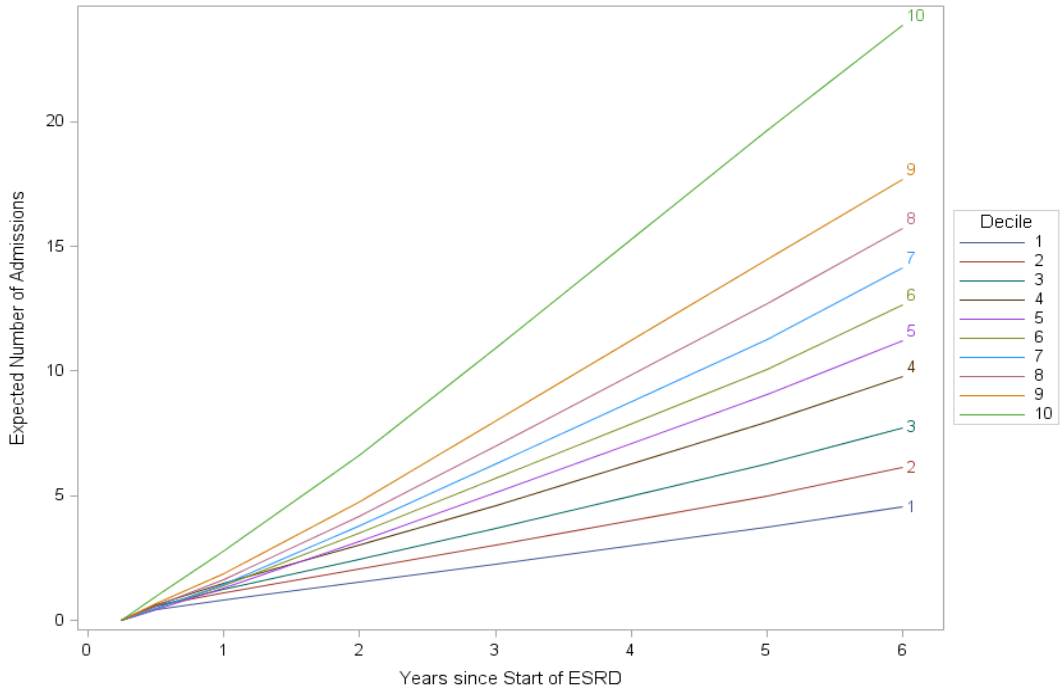
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

N/A

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Decile plots showing piecewise linear estimates of the cumulative rates by years since start of ESRD are plotted in Figure 4.

Figure 4. Decile Plot for SHR Admissions (2013 data).



Martingale residual plots were also examined (Figures 5-7).

Figure 5. Martingale Residuals by Age of Patient with LOESS Curve (2013 data).

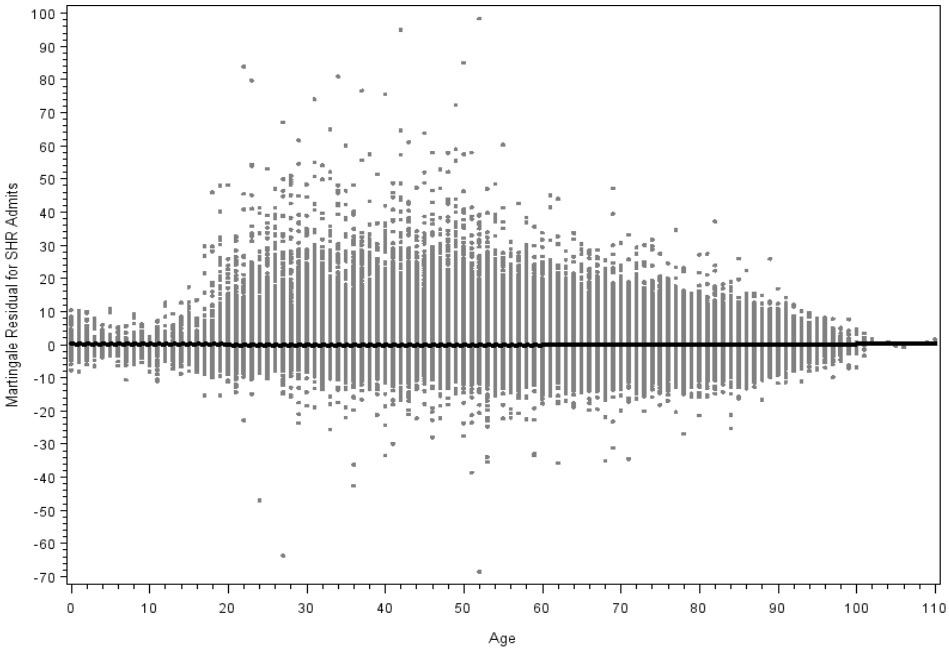


Figure 6. Martingale Residuals by BMI of Patient with LOESS Curve (2013 data).

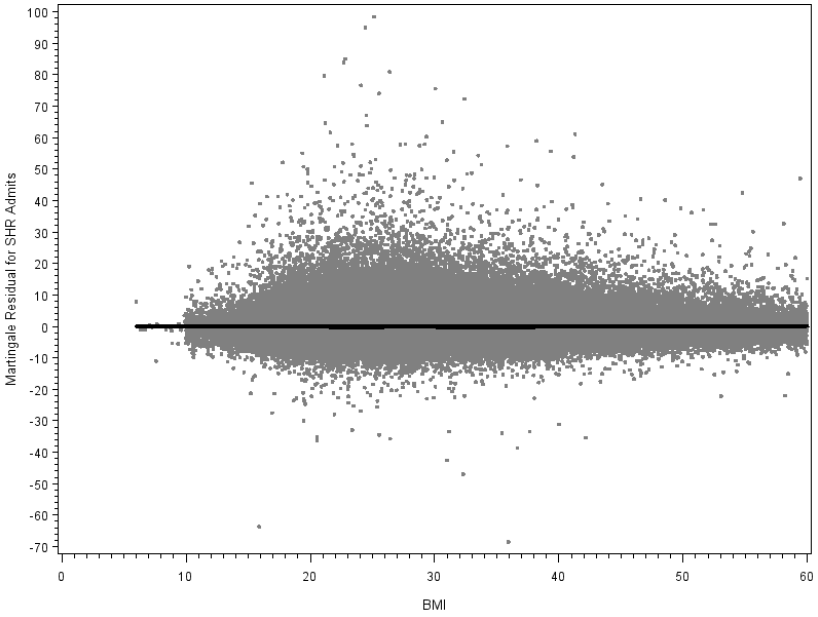
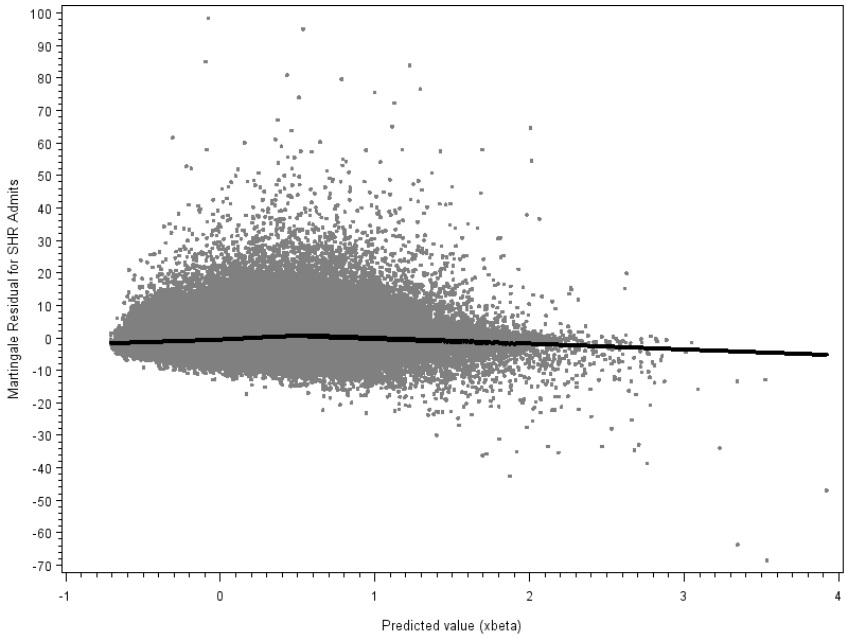


Figure 7. Martingale Residuals by Predicted Value of Patient with LOESS Curve (2013 data).



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. 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*)

The decile plot shows 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 hospitalization rates). The absolute differences between the groups is also large, with patients predicted to have the highest hospitalization rates (line 10) having 3 times higher hospitalization rates than those predicted to have the lowest rates (line 1).

The Martingale residual plots also did not indicate problems with the model fit. There was no pattern in the residuals that suggested lack of fit in any of the variables considered. In the LOESS plots attached, the LOESS curve for the mean of the residuals is flat indicating that there is no problem with the fit for each of the variables considered. The adjustment variables are highly predictive of the hospital admissions, and model extensions to examine interactions suggest a good overall fit.

2b4.11. 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

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

To adjust for over-dispersion of the data, we compute the p-value for our estimates using the empirical null distribution, a robust approach that takes account of the natural random variation among facilities that is not accounted for in the model (Efron, 2004; Kalbfleisch and Wolfe, 2013). Our algorithm consists of the following concrete steps. First, we fit an over-dispersed Poisson model (e.g., SAS PROC GENMOD with link=log, dist=poisson and scale=dscale) for the number of hospital admissions

$$\log(E[\mathbf{n}_{ik}]) = \log(\mathbf{E}_{ik}) + \boldsymbol{\theta}_k,$$

where $\mathbf{n}_{ik}$ is the observed number of events for patient *i* in facility *k*, $\mathbf{E}_{ik}$ is the expected number of events for patient *i* in facility *k* and $\boldsymbol{\theta}_k$ is the facility-specific intercept. Here, i ranges over the number of patients $\underline{N_k}$ who are treated in the *k*th facility. The natural log of the SHR for the *k*th facility is then given by the corresponding estimate of $\boldsymbol{\theta}_k$. The standard error of $\boldsymbol{\theta}_k$ is obtained from the robust estimate of variance arising from the overdispersed Poisson model.

Second, we obtain a z-score for each facility by dividing the natural log of its SHR by the standard error from the general linear model described above. These z-scores are then grouped into quartiles based on the number of patient years at risk for Medicare patients in each facility. Finally, using robust estimates of location and scale based on the normal curve fitted to the center of the z-scores for the SHR, we derive the mean and variance of a normal empirical null distribution for each quartile. This empirical null distribution is then used to calculate the p-value for a facility's SHR.

2b5.2. 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*)

Table 5. Number and percentage of facilities by classification of SHR, 2013. Categories stratified by facility size.

Number of patients	Better than expected	As expected	Worse than expected
< 51	0.26% (15)	31.86% (1,866)	1.47% (86)
51 - 87	0.39% (23)	31.71% (1,857)	1.79% (105)
> 87	0.43% (25)	30.46% (1,784)	1.64% (96)

2b5.3. 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?*)

Without empirical null methods, a large number of facilities will be flagged, including many larger facilities with a relatively small difference between the rates of hospitalization. In contrast, the methods based on the empirical null make appropriate adjustments for overdispersion. Using this method, facilities are flagged if they have outcomes that are extreme when compared to the variation in outcomes for other facilities of a similar size. Overall, most facilities are

flagged as expected (94.03%), while approximately 1% are better than expected, and approximately 5% are flagged as worse than expected.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

2b7.2. 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*)

2b7.3. 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)

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition
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)*
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.**

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.**
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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
N/A

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Dialysis Facility Compare (DFC) http://www.medicare.gov/dialysisfacilitycompare/

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting: Dialysis Facility Compare (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 5 patient years at risk. For the most recent DFC report, that was 5,992 facilities.

Patients included: All patients who meet the requirements to be included in the measure.

4a.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

4a.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

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Hospitalization rates have decreased over time as evidenced by the coefficients for calendar year from the SHR model. The hospitalization rate for 2011 decreased by 3% compared to 2010 (p-value <0.0001). Subsequent years had a larger decrease in the hospitalization rate compared to 2010 at 12.7% lower for 2012 and about 16.2% lower for 2013 (p-value<0.0001 for both).

SHR Calendar Year Model Coefficients, 2010-2013

2011: Coefficient = -0.03, P-value = <0.0001
2012: Coefficient = -0.127, P-value = <0.0001
2013: Coefficient = -0.162, P-value = <0.0001

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
In the past, a concern has been raised about patient selection relating to ensuring access to care for sicker patients. CMS considers ensuring patient access to dialysis care to be particularly important, and we continue to seek ways to ensure that access is unabated as part of the measures we develop, specifically outcome measures like SHR. The SHR measure incorporates a risk adjustment methodology that levels the playing field for facilities with different patient case-mixes in order to dis-incentivize patient cherry-picking. Given the adjustments for patients' prevalent comorbidities, which reflect a substantial modification to SHR, we think this additional risk adjustment strategy would discourage avoidance of treating sicker and more complex patients that are more likely to experience a hospital admission.

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.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0369 : Standardized Mortality Ratio for Dialysis Facilities
2496 : Standardized Readmission Ratio (SRR) for dialysis facilities

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
These measures are not completely harmonized. Each measure assesses different outcomes as reflected in certain differences across the measure specifications. SHR, SMR and SRR are harmonized to the population they measure (Medicare-covered ESRD patients),

methods (SMR and SHR) and certain risk adjustment factors specific to the ESRD population. SHR and SMR adjust for all the same comorbidity risk factors, a similar set of patient characteristics, and use fixed effects in their modeling approach. The differences between SHR, SMR and SRR reflect adjustment for factors specific to the outcome of each respective measure. Both SHR and SMR adjust for a set of prevalent comorbidities (observed in a prior year), however the complete set of comorbidities differs for SRR. SRR excludes planned readmissions; and adjusts for discharging hospital, acknowledging that for readmission, hospitals also bear accountability for properly coordinating care with the dialysis facility. These risk adjustments in SRR account for those characteristics specifically associated with readmission, and do not apply to SHR or SMR. SHR adjusts for sex to account for sex-age specific effects associated with higher hospitalization. Only SMR adjusts for state death rates, race, and ethnicity to account for these respective differences related to mortality outcomes and that are deemed outside of a facility's control.

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.)
N/A

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:** 1463_Appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, Sophia.Chan@cms.hhs.gov, 410-786-5050-
Co.3 **Measure Developer if different from Measure Steward:** University of Michigan Kidney Epidemiology and Cost Center
Co.4 **Point of Contact:** Casey, Parrotte, parrotte@med.umich.edu

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.
The following is a list of TEP members who participated in the End-Stage Renal Disease Evaluation of Potential Prevalent Comorbidity Adjustments in the Standardized Hospitalization Ratio (SHR) and the Standardized Mortality Ratio (SMR) TEP. In this advisory role, the primary duty of the TEP was to review any existing measures in terms of comorbidities included as adjusters, and determine if there was sufficient evidence to support the inclusion of specific proposed comorbidities as measure adjusters, and relatedly, suggest measure specifications.

Caroline Steward, APRN, CCRN, CNN
Advanced Practice Nurse (Hemodialysis)
Capital Health System
Trenton, NJ

Dana Miskulin, MD, MS
Staff Nephrologist
Turfts Medical Center
Boston, MA
Associate Professor of Medicine

Outcomes Monitoring Program, Dialysis Clinic Inc.
Nashville, TN

David Gilbertson, PhD
Co-Director of Epidemiology and Biostatistics
Chronic Disease Research Group
Minneapolis, MN

Eduardo Lacson Jr, MD, MPH
Nephrologist
American Society of Nephrology
Lexington, MA

Jennifer Flythe, MD, MPH
Research Fellow
University of North Carolina at Chapel Hill
Assistant Professor of Medicine
Chapel Hill, NC

Lorien Dalrymple, MD, MPH
Associate Professor
University of California, Davis
Division of Nephrology
Sacramento, CA

Mark Mitsnefes, MD, MS
Professor of Pediatrics
Cincinnati Children's Hospital Medical Center
Program Director
University of Cincinnati
Cincinnati, OH

Roberta Wager, MSN, RN
Renal Care Coordinator
Fresenius Medical Care
Member of Forum of ESRD Networks Beneficiary Council
Forum of ESRD Networks
Boerne, TX

Danielle Ward
Member of Forum of ESRD Networks Beneficiary Council
Forum of ESRD Networks
Board Member
Network 6
Wake Forest, NC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
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, 2017

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

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

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2977
Measure Title: Hemodialysis Vascular Access: Standardized Fistula Rate
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: Adjusted percentage of adult hemodialysis patient-months using an autogenous arteriovenous fistula (AVF) as the sole means of vascular access.
Developer Rationale: The NKF K/DOQI guidelines state the following: 1) AV fistulas have the lowest rate of thrombosis and require the fewest interventions, 2) cost of AV fistula use and maintenance is the lowest, 3) fistulas have the lowest rates of infection, and 4) fistulas are associated with the highest survival and lowest hospitalization rates. Indeed, a number of epidemiologic studies consistently demonstrate the reduced morbidity and mortality associated with greater use of AV fistulas for vascular access in maintenance hemodialysis.
As the accompanying literature review indicates, there are a growing number of studies reporting that creating AVF in some patients is less likely to be successful in the presence of certain comorbidities. In addition, certain patient groups may have less incremental benefit from an AV fistula relative to an AV graft. By adjusting the fistula rate for patient characteristics and comorbidities associated with low AV fistula success rates, this measure accounts for patients where a graft or even a catheter may be a more appropriate option.
This measure is intended to be jointly reported with Hemodialysis Vascular Access: Long-term Catheter Rate. These two vascular access quality measures, when used together, consider Arterial Venous Fistula (AVF) use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome. With the growing recognition that some patients have exhausted options for an AVF or have comorbidities that may limit the success of AVF creation, joint reporting of the measures accounts for all three vascular access options. The fistula measure adjusts for patient factors where fistula placement may be either more difficult or not appropriate and acknowledges that in certain circumstances an AV graft may be the best access option. This paired incentive structure that relies on both measures (SFR, long-term catheter rate) reflects consensus best practice, and supports maintenance of the gains in vascular access success achieved via the Fistula First/Catheter Last Project over the last decade.

Numerator Statement: The numerator is the adjusted count of adult patient-months using an AVF as the sole means of vascular access as of the last hemodialysis treatment session of the month.
Denominator Statement: All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) for the entire reporting month at the same facility.
Denominator Exclusions: Exclusions that are implicit in the denominator definition include:
•Pediatric patients (<18 years old)
•Patients on Peritoneal Dialysis
•Patient-months with in-center or home hemodialysis for less than a complete reporting month at the same facility

In addition, the following exclusions are applied to the denominator:
Patients with a catheter that have limited life expectancy:
•Patients under hospice care in the current reporting month
•Patients with metastatic cancer in the past 12 months
•Patients with end stage liver disease in the past 12 months
•Patients with coma or anoxic brain injury in the past 12 months

Measure Type: Intermediate Clinical Outcome
Data Source: Administrative claims, Electronic Clinical Data
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- The developer's rationale for this intermediate clinical outcome measure is that there is an association between type of vascular access used for hemodialysis and the risk of patient mortality. The developer provides evidence suggesting that measuring the AV Fistula rate will decrease the risk of patient morbidity and mortality. This evidence is derived from a systematic review which grades the quantity, quality, and consistency of the body of evidence examining the linkage between catheter type and patient morbidity and mortality.
- The developer provides results of a systematic review of the evidence, concluding that:
  - A number of epidemiologic studies consistently demonstrate the reduced morbidity and mortality associated with greater use of AV fistulas for vascular access in maintenance hemodialysis.
  - There are a growing number of studies reporting that creating AVF in some patients is less likely to be successful in the presence of certain comorbidities
- The developer notes that this measure is intended to be jointly reported with Hemodialysis Vascular Access: Long-term Catheter Rate. Used together, the two vascular access quality measures consider Arterial Venous Fistula (AVF) use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome.
- The developer provides this following linkage between the measure focus and a desired health outcome: Measure AV Fistula Rate→Assess value→Identify patients who do not have an AV Fistula→Evaluation for an AV fistula by a qualified dialysis vascular access provider→Increase Fistula Rate→Lower patient mortality.
- Developer references clinical practice guidelines (NKF: KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates). The order of preference for placement of fistulae in patients with kidney failure who choose HD as their initial mode of KRT are listed in descending order. The guidelines have all been assigned a Grade B (moderately strong evidence) rating.

Guidance from the Evidence Algorithm:
Intermediate outcome (Box 3) → Systematic review with QQC (Box 4) → SR concludes: "moderately strong"

Questions for the Committee:

- *Does the evidence indicate that measuring AV type leads to a reduction of the undesired health outcome?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- An analysis of CROWNWeb data from January 2014-December 2014 indicates the mean percentage of patient-months with a fistula was 64.6% (SD=10.4%).

Minimum	1st Quartile	Median	3rd Quartile	Maximum
9.2%	57.8%	64.8%	71.7%	97.5%

Disparities

- 2014 data indicate age, sex, race and ethnicity were evaluated in a logistic regression model for AV Fistula use.
  - Patients 75 years of age or older were 18% less likely to have an AV fistula when compared to the younger reference group while females are about half as likely to have fistulas as males.
  - Hispanic ethnicity was associated with higher odds of fistula use whereas blacks are about 33% less likely to have fistulas than whites.
  - The analysis results for age, race, and sex indicate potential disparity in fistula use.
- Below are the odds ratios for these patient characteristics:

Age	Odds Ratio (95% CI)	P-value
18-<25	1.08 (0.85, 1.36)	0.542
25-<60	1.07 (1.02, 1.12)	0.005
60-<75	used as the reference group	
75+	0.82 (0.78, 0.87)	.0001

Sex	Odds Ratio (95% CI)	P-value
Female	0.52 (0.50, 0.54)	<.0001
Male	used as the reference group	

Race	Odds Ratio (95% CI)	P-value
Black	0.67 (0.63, 0.71)	<.0001
Other race	1.07 (0.96, 1.19	0.206
White	used as the reference group	

Ethnicity	Odds Ratio (95% CI)	P-value
Hispanic	1.16 (1.08, 1.25)	<.0001
Non-Hispanic	used as the reference group	

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Using the provided disparities information, are you aware of any other evidence that would support the claim that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Process measure. Evidence relates strongly to this process metric - although it is largely retrospective observational data. While this measure excludes some classes of patients (ex.: hospice patients) it is possible that for some other subsets (ex: frail elderly with PVD) the evidence is not clear.

Preliminary rating: moderate

**The evidence linking AVF to outcomes is well recognized and documented in the literature. However, that is related to an AVF

rate at a population level not a statistically adjusted model or the ratio being proposed.

**A number of epidemiologic studies consistently demonstrate the reduced morbidity and mortality associated with greater use of AV fistulas for vascular access in maintenance hemodialysis. In addition ,there are a growing number of studies reporting that creating AVF in some patients is less likely to be successful in the presence of certain comorbidities. This measure is intended to be jointly reported with Hemodialysis Vascular Access: Long-term Catheter Rate. Used together, the two vascular access quality measures consider Arterial Venous Fistula (AVF) use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome. The developer provides this following linkage between the measure focus and a desired health outcome: Measure AV Fistula Rate-☐Assess value-☐Identify patients who do not have an AV Fistula--Evaluation for an AV fistula by a qualified dialysis vascular access provider---☐Increase Fistula Rate Lower patient mortality.

The structure therefore relates to the desired outcome.

**Good evidence review post K-DOQi (showing AVF is superior vascular access) with multiple references supporting notion that elderly fragile patients and others may not always be best served by a fistula .

**Measure focuses on intermediate clinical outcome: fistula rate (facility level measure);

Evidence directly relates to measure;

The intermediate outcome is related to important clinical outcomes.

**systematic review of evidence provided with studies both demonstrating benefits of AVF and data that dealt with comorbidities that minimize this benefit

**The evidence appears to be strong.

**Strong evidence base for association of fistula use with better outcomes-morbidity,mortality, hospitalization

**No concerns

**The evidence related directly to the outcome being measured. Including the denominator exclusions was strongly verbalized at the last Renal Standing Committee and CMS used 2015 Vascular access TEP to help identify this subset of patients. Adding this subset of patients makes this a more appropriate measure. Would like to have added to this subset the group of patients who have exhausted all other accesses and are hemodialysis catheter dependent.

1b. Performance Gap

Comments: **Interquartile differences in measured performance from CrownWeb are substantial. Disparities in age, sex, ethnicity and race are apparent.

Preliminary rating: high

**Based on the Dec 2015 Fistula First Dashboard (http://fistulafirst.org/ffcl/for-ffcl-professionals/) the prevalent fistula rate was 62.94% suggesting a gap does exist.

Given that access to surgeons in a local area is a key determinant of AVF placement, it is possible that disparities exist. To evaluate that one would need to examine the AVF placement rate by MSA or some similar breakdown which was not done by the developer.

**2014 data indicate age, sex, race and ethnicity were evaluated in a logistic regression model for AV Fistula use.

Patients 75 years of age or older were 18% less likely to have an AV fistula when compared to the younger reference group while females are about half as likely to have fistulas as males.

Hispanic ethnicity was associated with higher odds of fistula use whereas blacks are about 33% less likely to have fistulas than whites.

The analysis results for age, race, and sex indicate potential disparity in fistula use.

Therefore there is a performance gap.

**There is a wide range in the percentage of HD patients utilizing a fistula across facilities. While this is not the same as their measure, the focus is correct. Impressive disparities data which also emphasizes the performance gap

**Gap is demonstrated with variation across facilities, median 64.8% [IQR 57.8, 71.7]

**wide range of performance from CROWNWeb data

**There is a sizable gender and racial gap.

**Gap well demonstrated

**Agree with developers rationale

**Yes a significant gap in care which is less than optimal performance. Data on disparities is well demonstrated based on age, sex, ethnicity and race.

1c. High Priority (previously referred to as High Impact)

Comments: **Yes.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative Claims, Electronic Clinical Data

Specifications:

- The numerator of this measure is: *The adjusted count of adult patient-months using an AVF as the sole means of vascular access as of the last hemodialysis treatment session of the month.*
- The denominator of this measure is: *All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) for the entire reporting month at the same facility.*
- Both ICD-9 and ICD-10 codes are included in the Data Dictionary Code Table.
- The measure logic is included in the appendix and seems straightforward.
- This intermediate clinical outcome measure is risk adjusted, using a statistical risk model.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- To determine the measure's reliability, developers calculated the inter-unit reliability (IUR) for the annual performance scores. The IUR measures the proportion of the total variation of a measure attributable to between-facility variation. A small IUR reveals that most of the variation of the measure 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 indicates that most of the variation between facilities is due to real differences between facilities.
- The reliability of SFR calculation only included facilities with at least 11 patients during the entire year.

Results of reliability testing

- The IUR is 0.736. This indicates that about 74% of the variation in the annual SFR can be attributed to between-facility differences in performance (signal) and about 26% to the within-facility variation (noise). The developer states that the result of IUR testing suggests a high degree of reliability.

Guidance from the Reliability Algorithm

Precide specifications (Box 1) → empirical reliability testing of measure score (Boxes 2 and 4) → Appropriate method → moderate (perhaps high) certainty that scores are reliable

Questions for the Committee:

o *Is the test sample adequate to generalize for widespread implementation?*
o *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:
o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:
☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:
- The developer assessed validity using a Poisson regression model to measure the association between facility-level quintiles of performance scores and the 2014 Standardized Mortality Rate (SMR) for Dialysis Facilities(NQF 0369) as well as the Standardized Hospital Ratio for Dialysis Facilities (NQF1463). Performance scores at the facility level were divided into quintiles (Q1-Q5) and the relative risk of mortality was calculated for each quintile (Q4 and Q5 were used as reference group).

Validity testing results:
- Quintiles of the performance scores were defined as:
  - Q1: 9.2% - < 55.9%
  - Q2: 55.9% - < 62.0%
  - Q3: 62.0% - < 67.4%
  - Q4: 67.4% - < 73.4%
  - Q5: 73.4% - < 97.5%
- The developer provided the following as an analysis of the results:
  - The percent of patient-months with a fistula had a significant negative association with the risks of mortality and hospitalization.
  - For the 2014 SMR, the relative risk of mortality increased as the performance measure quintile decreased, with the highest risk in the Q1.

Quintile	Relative Reliability (RR)	95% CI	P-value
Q1	1.10	1.08, 1.12	<0.001
Q2	1.05	1.03, 1.07	<0.001
Q3	1.03	1.01, 1.05	0.003
Q4	Used as reference group		
Q5	Used as reference group		

- For Standardized Hospital Ratio, the relative risk of hospitalization increased as the performance measure quintile decreased from the reference group with the highest risk in Q1. These results suggest the predictive relationship of lower fistula use with higher mortality and hospitalization.

Quintile	Relative Reliability (RR)	95% CI	P-value
Q1	1.11	1.11, 1.11	<0.001
Q2	1.08	1.08, 1.09	<0.001
Q3	1.07	1.06, 1.07	<0.001
Q4	Used as reference group		
Q5	Used as reference group		

Questions for the Committee:

- *Are the correlations with SMR and SHR as expected?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- These following patients are excluded from the denominator:
  - Patients with a catheter that have limited life expectancy, which is defined as patients:
    - Under hospice care in the current reporting month,
    - With metastatic cancer in the past 12 months,
    - With end stage liver disease in the past 12 months, and
    - With coma or anoxic brain injury in the past 12 months
- The developer calculated and compared the facility-level standardized fistula rate with and without the patient-month exclusions.
- Developer states the exclusion criteria are necessary since the percentage of patients excluded at each facility are not evenly distributed across all facilities. The exclusions take in to account that some facilities treat a higher proportion of patients with limited life expectancy.
- The frequency of exclusions in 2014:
  - Patient years: 0.92% excluded
  - Patients: 2.35% excluded

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Statistical Risk Model with 19 patient-month level risk factors.

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- Because of the variation in the burden of comorbidities between different facilities, adjusting for these factors when calculating an AV fistula rate recognizes some patients are more likely to have AV grafts. Studies detail particular patient characteristics that are associated with a decreased likelihood of having a successful AV fistula created. Decisions of the risk factors were based on both the clinical and statistical association with the lower

likelihood of fistula use in patients with these risk factors, and that these factors were not likely to be associated with facility care.

- The risk adjustment is based on a multivariate logistic regression model. The adjustment is made for age, BMI at incident, nursing home status, nephrologist's care prior to ESRD, duration of ESRD, diabetes as primary cause of ESRD, and comorbidities. The common risk effects are assumed in order to improve computational stability in estimating facility-specific effects.
- In Table 4, results are listed from the adjusted model described. Table 5 shows the parameter estimates for patient and area level SDS/SES variables based on a logistic regression model for AV fistula use that included all these variables along with all the other clinical covariates used for adjustment in SFR.
- Area-level SDS factors were not included in the risk-adjustment model due to the absence of clinically meaningful or statistically observed differences on the fistula rate with these adjustments.
- Patient-level SDS/SES variables were not included in the risk-adjustment model due to the absence of a convincing biological or clinical rationale that warrant accounting for different outcomes on the basis of race, sex, or socioeconomic status.

Questions for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- For both the annual SFR and the overall national distribution, the majority (93%) of facilities have "as expected" performance, 2.4% facilities have performed "worse than expected," and 4.4% of facilities have performed better than expected.
- A higher rate of fistula use represents better quality of care. This analysis demonstrates both practical and statistically significant differences in performance across facilities based on their adjusted proportion of patient months with a fistula in use.

Question for the Committee:

- *Can stakeholders make meaningful conclusions about quality when 93% of facilities perform "as expected"?*
- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

None needed.

2b7. Missing Data

An analysis of missing data was not provided for this measure.

Guidance from the algorithm:
Specifications consistent with evidence (Box 1) → Potential threats to validity – meaningful differences might be an issue (Box 2) → empirical testing of the measure score (Boxes 3 and 6) → appropriate method (box 7) → high (possibly moderate) certainty

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: None

2a2. Reliability Testing

Comments:

**Sample is adequate for generalizability

IUR testing shows good signal:noise,

Preliminary rating: high

**To determine the measure's reliability, developers calculated the inter-unit reliability (IUR) for the annual performance scores. The IUR measures the proportion of the total variation of a measure attributable to between-facility variation. A small IUR reveals that most of the variation of the measure 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 indicates that most of the variation between facilities is due to real differences between facilities.

The reliability of SFR calculation only included facilities with at least 11 patients during the entire year.

The IUR is 0.736. This indicates that about 74% of the variation in the annual SFR can be attributed to between-facility differences in performance (signal) and about 26% to the within-facility variation (noise). Therefore IUR testing suggests a high degree of reliability.

**The IUR is 74% which seems good.

**IUR 0.736

**IUR of 74% -- testing uses data source and level of analysis indicated in measure

**Reliability testing is appropriate

**reliability well demonstrated

**No concerns

**IUR is 74% - moderate degree of reliability. Jan - Dec 2014 crownweb data used to calculate facility - level IUR. Appropriate number and method. Still concern about appropriate capturing of Medicare claims for the denominator exclusions.

2b2. Validity Testing

Comments:

**Correlation between AVF use and SMR, SHR are high and as expected.

Test sample is adequate.

Validity appears high and can be used to assess quality.

Preliminary rating: high

**Rather than rely on a quintiles comparison, it may be worthwhile to do a delta delta analysis as well which KECC has previously championed. That being said the correlation provides supports the measure.

**The developer assessed validity using a Poisson regression model to measure the association between facility-level quintiles of performance scores and the 2014 Standardized Mortality Rate (SMR) for Dialysis Facilities(NQF 0369) as well as the Standardized Hospital Ratio for Dialysis Facilities (NQF1463). Performance scores at the facility level were divided into quintiles (Q1-Q5) and the relative risk of mortality was calculated for each quintile (Q4 and Q5 were used as reference group).

The results suggest the predictive relationship of lower fistula use with higher mortality and hospitalization.

**Validty was assessed by correlating the measure with SMR and SHR – the differences were not large but were consistent and in the right direction and statistically significant.

**Validity assessed by examining the association between performance scores and the SMR or SHR.

**regression model relates quintiles of facility performance with this measure to 2014 SMR and SHR -- decreased performance with measure and lower SMR and SHR found statistically significant

**Validity testing is appropriate

**validity well demonstrated

**No concerns

**Poisson regression model measured association of AVF performance scores and 2014 Standardized Mortality Rate (SMR) & standardized hospital ratio for dialysis facilities & found a significant predictive relationship of lower AVF use with higher mortality and hospitalization in dialysis facilities. A 2015 vascular access TEP provided input of the development of access measures,

exclusions and risk adjustment factors. This measure was based on the final TEP recommendations. I do agree that higher use of AVF is an indicator of quality and improves health outcomes.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:

**Exclusion appear appropriate and data-supported - not certain there are not other subsets that are not evenly distributed across facilities and perhaps might be likewise excluded.
Risk adjustment appears to include the important factors.
Analyses showed that SDS factors impacted fistula rates on the individual patient level, but not at the area level. It thus seems appropriate not to adjust for SDS for this facility-level measure.
It is no surprise to me that only 2.4% facilities are worse than expected and 4.4% better than expected. It is clinically important to identify these "outliers" for remediation on one side, and for best practice learning on the other.
Preliminary rating: high
**The use of the 2728 data has the potential to create validity issues as mentioned above. In addition, the CW data source being referenced should be validated for missing data. Previous analytics have compared the correlation between CW and Claims based vascular access reporting and found a high correlation. This makes sense give that both data originate in the same source EMR systems. However, there have been known issues with data completeness for CW resulting in missing data. So a better data completeness metric is the number of patients with data reported divided by the unit census for the same time period.
With regards to the proposed exclusion criteria. It is not clear how these will be crisply extracted from claims data but one assumes that there has been some validation of the ICD9/10 codes against chart reviews for the sensitivity of the claims based appraoch

- ☐ Under hospice care in the current reporting month,
- ☐ With metastatic cancer in the past 12 months,
- ☐ With end stage liver disease in the past 12 months, and
- ☐ With coma or anoxic brain injury in the past 12 months

**Excluded from the denominator: Patients with a catheter that have limited life expectancy, which is defined as patients: Under hospice care in the current reporting month, With metastatic cancer in the past 12 months, With end stage liver disease in the past 12 months, and With coma or anoxic brain injury in the past 12 months. These are consistent and appropriate exclusions.
Decisions of the risk factors were based on both the clinical and statistical association with the lower likelihood of fistula use in patients with these risk factors, and that these factors were not likely to be associated with facility care. The risk adjustment is based on a multivariate logistic regression model. The adjustment is made for age, BMI at incident, nursing home status, nephrologist's care prior to ESRD, duration of ESRD, diabetes as primary cause of ESRD, and comorbidities. The common risk effects are assumed in order to improve computational stability in estimating facility-specific effects. Risk adjustments were appropriate.
A higher rate of fistula use represents better quality of care. The analysis demonstrates both practical and statistically significant differences in performance across facilities based on their adjusted proportion of patient months with a fistula in use despite 93% as expected.
**Threats to Validity: Showed the denominator exclusions were not a threat, or improved validity. After analysis of SDS they decided to include that info in the risk adjustment. Only a small percentage of units performed Worse or Better thean expected.
**Exclusions may need to be more specifically defined (specific codes);
What are the implications for only being able to apply exclusions among Medicare recipients?
4.4% facilities better than expected and 2.38% facilities worse than expected
**The only other consideration I would have is for patients who cannot practically get the AV fistula due to anatomic issues
**No concerns
**The stated exclusions are consistent with the evidence. Feel that the subgroup of patients who have exhausted all other accesses should be added to this exclusion group. Data collection burden is appropriate.
Risk adjustment is made for age, BMI, NH status, Nephrologist care prior to ESRD, diabetes as primary cause of ESRD and co-

morbidities using a multivariate logistic regression model.
THE CMS AVF target at the facility level is 68% rather than 100% which recognizes that 1/3 of patients will require a different type of access which includes use of AV grafts. The 2015 Vascular Access TEP wanted to ensure that AVGs were a strongly preferred outcome to catheters and should not be disincentivized.
SDS factors were included in the analysis including Area Deprivation Index (ADI). The risk adjustment seems to have been appropriated developed and tested with acceptable results - especially with the TEP input.
Meaningful difference: Standardized fistula rate of each facility was compared to the overall national distribution. Proportion of facilities with statistically significant differences (p values <0.5):
Better than expected - 261 facilities (4.4%), As expected - 5526 facilities (93.22%) and worse than expected - 141 facilities (2.3%).
Analysis does allow results to be appropriately compared and evaluated. The SDS/SES variables used a logistic regression model that included all these variables. Area level SDS factors were not included in the risk-adjustment model d/t the absence of clinically meaningful or statistically observed differences in the AVF rate with these adjustments.
CMS acknowledges that certain areas will have a higher percentage of patients with the denominator exclusions which may increase the difficulty achieving the performance goal. An analysis of missing data was not provided for this measure.

Criterion 3. Feasibility

3. Feasibility is the 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.

- All data elements are in defined fields in a combination of electronic sources and are collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: None

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details: CMS will determine if and when the measure will be implemented in a CMS program. Upon endorsement, CMS will consider retiring the currently endorsed measure of fistula use (#0257) in favor of this new

measure for implementation in a future performance year for the ESRD QIP and reporting period for Dialysis Facility Compare at the next available opportunity.

Improvement results: The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of the standardized fistula rate. Once implemented facility performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in promoting fistula use for the incident and prevalent populations, while also taking into account those patient risk factors that hinder successful fistula use in certain subpopulations.

Unexpected findings (positive or negative) during implementation

Potential harms: Potential unintended consequences would center on patients being pushed towards having an AVF created when they may not realize a benefit from this type of access due to either comorbidities or a limited life expectancy. It is incumbent on both dialysis facilities, as well as surgeons, to establish the most appropriate type of vascular access for patients based on their individual circumstances and preferences, rather than in response to quality measures.

Feedback:
No feedback from other sources available.

***Questions for the Committee**:*

- *The developer indicates that "This measure is intended to be jointly reported with Hemodialysis Vascular Access: Long-term Catheter Rate." Should the measure be formally paired, i.e. reporting of both measures is included in the NQF endorsement?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Metric can be easily used - - perhpas too easily...
As stated before, the basis of the data is retrospective analyses, and there may be subsets of patients other than those excluded for which fistula use is not as well correlated with poor outcomes. Patient choice is not at all considered - - and patients surely may be coerced into multiple procedures to establish fistulae when for their subset the evidence does not support its use.
Preliminary rating: moderate
**The use of a statistically and standardized rate creates an issue in usability. Namely that dialysis units are unable to calculate such metrics (as opposed to a raw unadjusted rate) for continuous improvement purposes.
**CMS will determine if and when the measure will be implemented in a CMS program. The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of the standardized fistula rate. Potential unintended consequences would center on patients being pushed towards having an AVF created when they may not realize a benefit from this type of access due to either comorbidities or a limited life.
**This is a credible case that it will be reported and used for payment in the future.
**Planned for public reporting and payment
**measure sponsor describes how measure may be used in upcoming QIP or as part of DFC
**There's not current public reporting. There may be potential unintended consequences of inappropriate AVF use, but overall the benefits outweigh the risks

**not currently reported but excellent potential for accountability.
if adopted along with companion catheter rate will allow retirement of previous fistula PM.
**Proposed for use with measure 2978
**Presently not publicly reported. If endorsed, CMS will consider retiring the currently endorsed measure of AVF use (#0257) in favor of this new measure for implementation in a future year for the ESRD QIP and for Dialysis Facility Compare.
Potential unintended consequence: promoting AVF placement in patients whom will not benefit from an AVF (exclusionary subgroup).

Criterion 5: Related and Competing Measures

Related or competing measures

- 0251: Vascular Access—Functional Arteriovenous Fistula (AVF) or AV Graft or Evaluation for Placement
- 2594: Optimal End Stage Renal Disease (ESRD) Starts

Harmonization

- Measure 0251 contains several components in addition to assessing fistula use. It is a referral process measure. The most basic requirement to get into the numerator is referral to a vascular surgeon (or other qualified physician). This has the potential for facilities to score well on the measure separate from whether patients are receiving treatment with a fistula, graft, or catheter, as long as the patient was referred to or evaluated by a vascular surgeon. We acknowledge this is an important step to fistula placement however it departs from the intent of this fistula measure to function as a more direct incentive to encourage fistula use. Moreover, consistent with the concerns and recommendations made by the vascular access TEP, the SFR is risk adjusted and includes risk factors to account for patients where fistula may not be the appropriate access type.
- Measure 2594 is not directed toward dialysis facilities. The setting focus addresses a different provider type which falls outside the purview of measures evaluating dialysis facility performance on fistula use. This suggests a fundamental difference in the measure target populations, setting and intent that cannot be harmonized. Additionally, the measure is limited to incident patients, while the SFR includes both incident and prevalent patients as the measured population.

Pre-meeting public and member comments

Comment by Joseph Vassalotti
Organization National Kidney Foundation
Comment #5687: The National Kidney Foundation (NKF) strongly supports this measure and its pairing with the long-term catheter rate measure (NQF #2978). NKF is particularly pleased with the additional exclusions that acknowledge catheter use for patients with limited life expectancy. These changes align with NKF's previous recommendations.

We do note that clarity around sole access use would strengthen this measure. Specifically, credit for this measure should only apply if the patient does not have a catheter. As written it could be interpreted that the facility would get credit for a patent with a catheter as long as the catheter was not being used for dialysis. The presence of a catheter increases patients risk for infection and therefore no credit should be given if the patient has a catheter. In contrast if a patient has an AV graft that is not being used credit for the measures should still apply as the risk of AV graft infection is low, but there is associated risk with removal.

Comment by Lisa McGonigal, MD, MPH
Organization National Kidney Care Partners
Comment #5716: As with the catheter measure, KCP used the existing arteriovenous fistula (AVF) measure, NQF 0257, for context in our review.

SPECIFICATIONS. The language in #0257 that specifically defines an autogenous AVF as using two needles has

been replaced with an autogenous AVF "as the sole means of vascular access." KCP believes the specifications are imprecise as to whether facilities would receive credit for patients using an AVF as the sole means of access, but who also have in place a graft or catheter that is no longer being used. We note patients with catheters remain at risk for infection and other adverse sequellae, so credit should not be not given when a catheter is present, even if an AVF is being used. A numerator that specifies the patient must be on maintenance hemodialysis "using an AVF with two needles and without a dialysis catheter present" would remove ambiguity. In contrast, removal of an AV graft is complex and not without risk of complications, so KCP believes credit should be received for a patient who is using an AVF as the sole means of access, but who also may have a non-functioning AV graft present.

KCP notes the 90-day ESRD requirement has been removed from the denominator statement as compared to #0257, which means the "clock" for the measure starts on the first day of dialysis in a non-hospital setting—but that the permitted timeframe for catheter use in the numerator is still 90 days; we support this change. Additionally, we commend the developer for adding an exclusion for patients with limited life expectancy and for now unambiguously identifying the four subcategories, both approaches that KCP had recommended.

VALIDITY. KCP believes this measure improves on #0257 and commends the developer for accepting KCP's recommendation in previous comments to remove the co-variate alcohol dependence from the model's risk variables. We continue to believe two additional vasculature risk variables would strengthen the model: a history of multiple prior accesses and the presence of a cardiac device.

KCP notes that the validity testing yielded an overall c-statistic of 0.71. We are concerned the model will not adequately discriminate performance—particularly that smaller units might look worse than reality. We believe a minimum c-statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Hemodialysis Vascular Access: Standardized Fistula Rate
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: 4/15/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
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

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☒ Intermediate clinical outcome (*e.g., lab value*): standardized fistula rate

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Several observational studies have demonstrated an association between type of vascular access used for hemodialysis and patient mortality. Arteriovenous fistulae (AVF) are associated with the lowest mortality risk while long term catheters have the highest mortality. Arteriovenous grafts (AVG) have been found to have a risk of death that is higher than AVF but lower than catheters.

The measure focus is the process of assessing AV Fistula use at chronic dialysis facilities.

This process leads to improvement in mortality as follows:
Measure AV Fistula Rate→ Assess value →Identify patients who do not have an AV Fistula→Evaluation for an AV fistula by a qualified dialysis vascular access provider →Increase Fistula Rate →Lower patient mortality.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

National Kidney Foundation KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates: Hemodialysis Adequacy, Peritoneal Dialysis Adequacy and Vascular Access. Am J Kidney Dis 48:S1-S322, 2006 (suppl 1). http://www.kidney.org/professionals/KDOQI/guidelines_commentaries

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

GUIDELINE 2. SELECTION AND PLACEMENT OF HEMODIALYSIS ACCESS

A structured approach to the type and location of long-term HD accesses should help optimize access survival and minimize complications. Options for fistula placement should be considered first, followed by prosthetic grafts if fistula placement is not possible. Catheters should be avoided for HD and used only when other options listed are not available.

2.1 The order of preference for placement of fistulae in patients with kidney failure who choose HD as their initial mode of KRT should be (in descending order of preference):

2.1.1 Preferred: Fistulae. (B)

2.1.2 Acceptable: AVG of synthetic or biological material. (B)

2.1.3 Avoid if possible: Long-term catheters. (B)

2.1.4 Patients should be considered for construction of a primary fistula after failure of every dialysis AV access. (B)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

KDOQI Guideline 2.1 was graded B, indicating moderate evidence supports the guideline.
The "B" rating indicates: It is recommended that clinicians routinely follow the guideline for eligible patients. There is moderately strong evidence that the practice improves health outcomes.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

The rating system defined in the KDOQI Guidelines was used to grade the strength of the Guideline recommendation.
KDOQI defined grades as follows:
Grade A: It is strongly recommended that clinicians routinely follow the guideline for eligible patients. There is strong evidence that the practice improves health outcomes.
Grade B: It is recommended that clinicians routinely follow the guideline for eligible patients. There is moderately strong evidence that the practice improves health outcomes.
Grade CPR: It is recommended that clinicians consider following the guideline for eligible patients. This recommendation is based on either weak evidence or on the opinions of the Work Group and reviewers that the practice might improve health outcomes.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

National Kidney Foundation. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates: Hemodialysis Adequacy, Peritoneal Dialysis Adequacy and Vascular Access. Am J Kidney Dis 48:S1-S322, 2006 (suppl 1).

http://www.kidney.org/professionals/KDOQI/guidelines_commentaries

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The evidence review focuses on the advantages of AV fistula compared to other types of vascular access and highlights the superior patency, reduced need for interventions, and lower infection rates associated with AV fistula.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

The quality of evidence was not explicitly graded in the KDOQI guidelines. However, it was implicitly assessed according to the criteria outlined in the table in 1a.7.3 below. The workgroup considered the overall methodological quality, the target population (e.g. patients on dialysis), and whether the health outcome was studied directly or not.

Overall, the evidence that supports the guideline was assessed as: Moderately Strong.

The workgroup defined "Moderately Strong" as: Evidence is sufficient to determine effects on health outcomes in the target population, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; OR evidence is from studies with some problems in design and/or analysis; OR evidence is from well-designed, well-conducted studies on surrogate endpoints for efficacy and/or safety in the target population.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

		Methadologic Quality		
Outcome	Population	Well designed and analyzed (little if any potential bias)	Some problems in design and/or analysis (some potential bias)	Poorly designed and/or analyzed (large potential bias)
Health Outcomes	Target Population	Strong	Moderately Strong	Weak
Health Outcomes	Other than target population	Moderately Strong	Moderately Strong	Weak
Surrogate Measure	Target Population	Moderately Strong	Weak	Weak
Surrogate Measure	Other than target population	Weak	Weak	Weak

Strong- Evidence includes results from well-designed, well-conducted study/studies in the target population that directly assess effects on health outcomes.
Moderately strong- Evidence is sufficient to determine effects on health outcomes in the target population, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; OR evidence is from a population other than the target population, but from well-designed, well conducted studies; OR evidence is from studies with some problems in design and/or analysis; OR evidence is from well-designed, well-conducted studies on surrogate endpoints for efficacy and/or safety in the target population.
Weak- Evidence is insufficient to assess the effects on net health outcomes because it is from studies with some problems in design and/or analysis on surrogate endpoints for efficacy and/or safety in the target population; OR the evidence is only for surrogate measures in a population other than the target population; OR the evidence is from studies that are poorly designed and/or analyzed.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: January 1997 – June 2005

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The 2006 Clinical Practice Guidelines for Vascular Access is an update to the original vascular access guidelines published in 1997 by the National Kidney Foundation. In the eight years that the literature review included for the update, there have been no randomized controlled trials for type of vascular access. Specifically, for the guideline used to support this measure, a total of 84 peer-reviewed publications are included in the body of evidence presented. While these are all observational studies, some are based on either national data such as the United States Renal Data System (USRDS) that includes all patients with end stage kidney disease in the US, or international data, such as the Dialysis Outcomes Practice Pattern Study (DOPPS) that provides a global perspective for US vascular access outcomes.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The overall quality of evidence is moderately strong. All studies are in the target population of hemodialysis patients. Some studies have evaluated health outcomes such as patient mortality, but have limitations due to the observational nature of the design. Other studies have more rigorous design, but use surrogate outcomes such as access thrombosis.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The 12 studies listed below highlight the core benefits such as reduced mortality and morbidity associated with using an AV fistula relative to either an AV graft or a tunneled catheter. Specifically, AV fistulae have:

- Lowest risk of thrombosis: in a systematic review of 34 studies evaluating access patency, AVF were found to have superior primary patency at 18 months compared to AV grafts (51% vs. 33%).[1]
- Lowest rate of angioplasty/intervention: Procedure rates have been reported as 0.53 procedures/patient/year for AV fistula compared to 0.92 procedures/patient/year for AV grafts.[2]
- Longest survival: Case-mix adjusted survival analysis indicated substantially better survival of AV fistula compared with AV grafts in the US [risk ratios (RR) of failure 0.56, $P < 0.0009$][3]
- Lowest Cost[4-6]: Based on 1990 costs to Medicare, graft recipients cost HCFA (CMS) $3,700 more than fistula patients when pro-rating graft reimbursements to the median fistula survival time.[5]
- Lowest rates of infection: AV fistula have the lowest rates of infection followed by AV grafts and then tunneled dialysis catheters[7]. Vascular access infections are common, and represent the second most common cause of death for patients receiving hemodialysis.[8]
- Lowest mortality and hospitalization: Patients using catheters (RR=2.3) and grafts (RR=1.47) have a greater mortality risk than patients dialyzed with fistulae[9]. Other studies have also found that use of fistulae reduces mortality and morbidity[10-12] compared to AV grafts or catheters.

References:

1. Huber TS, Carter JW, Carter RL, Seeger JM: Patency of autogenous and polytetrafluoroethylene upper extremity arteriovenous hemodialysis accesses: A systematic review. J Vasc Surg 38(5):1005-11, 2003
2. Perera GB, Mueller MP, Kubaska SM, Wilson SE, Lawrence PF, Fujitani RM: Superiority of autogenous arteriovenous hemodialysis access: Maintenance of function with fewer secondary interventions. Ann Vasc Surg 18:66-73, 2004
3. Pisoni RL, Young EW, Dykstra DM, et al: Vascular access use in Europe and the United States: Results from the DOPPS. Kidney Int 61:305-316, 2002
4. Mehta S: Statistical summary of clinical results of vascular access procedures for haemodialysis, in Sommer BG, Henry ML (eds): Vascular Access for Hemodialysis-II (ed 2). Chicago, IL, Gore, 1991, pp 145-157
5. The Cost Effectiveness of Alternative Types of Vascular access and the Economic Cost of ESRD. Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 1995, pp 139-157
6. Eggers P, Milam R: Trends in vascular access procedures and expenditures in Medicare's ESRD program, in Henry ML (ed): Vascular Access for Hemodialysis-VII. Chicago, IL, Gore, 2001, pp 133-143
7. Nassar GM, Ayus JC: Infectious complications of the hemodialysis access. Kidney Int 60:1-13, 2001

8. Gulati S, Sahu KM, Avula S, Sharma RK, Ayyagiri A, Pandey CM: Role of vascular access as a risk factor for infections in hemodialysis. Ren Fail 25:967-973, 2003
9. Dhingra RK, Young EW, Hulbert-Shearon TE, Leavey SF, Port FK: Type of vascular access and mortality in U.S. hemodialysis patients. Kidney Int 60:1443-1451, 2001
10. Woods JD, Port FK: The impact of vascular access for haemodialysis on patient morbidity and mortality. Nephrol Dial Transplant 12:657-659, 1997
11. Xue JL, Dahl D, Ebben JP, Collins AJ: The association of initial hemodialysis access type with mortality outcomes in elderly Medicare ESRD patients. Am J Kidney Dis 42:1013-1019, 2003
12. Polkinghorne KR, McDonald SP, Atkins RC, Kerr PG: Vascular access and all-cause mortality: A propensity score analysis. J Am Soc Nephrol 15:477-486, 2004

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The potential harms of placing an AV fistula include: (1) failure of the AV fistula to mature such that additional surgery is needed for vascular access, (2) steal syndrome where the distal arm becomes ischemic, and (3) prolonged maturation times that increase reliance on a tunneled dialysis catheter and its attendant risk of infection. Overall these risks associated with an AV fistula are considered to be small and overshadowed by the long-term benefits outlined above.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Casey JR, Hanson CS, Winkelmayer WC, et al. **Patients' perspectives on hemodialysis vascular access: a systematic review of qualitative studies.** *Am J Kidney Dis. 2014 Dec;64(6):937-53. doi: 10.1053/j.ajkd.2014.06.024. Epub 2014 Aug 10*.

This systematic review and thematic synthesis of qualitative studies describes patients' perspectives on vascular access initiation and maintenance in hemodialysis. 46 studies were reviewed and found that initiation of vascular access signifies kidney failure and imminent dialysis, which is emotionally confronting. Patients strive to preserve their vascular access for survival, but at the same time describe it as an agonizing reminder of their body's failings and "abnormality" of being amalgamated with a machine disrupting their identity and lifestyle. Timely education and counseling about vascular access and building patients' trust in health care providers may improve the quality of dialysis and lead to better outcomes for patients with chronic kidney disease requiring hemodialysis.

Impact: Adds the patient's perspective to the discussion on vascular access options.

Al-Jaishi AA, Oliver MJ, Thomas SM, et al. **Patency rates of the arteriovenous fistula for hemodialysis: a systematic review and meta-analysis.** *Am J Kidney Dis. 2014 Mar;63(3):464-78. doi: 10.1053/j.ajkd.2013.08.023. Epub 2013 Oct 30. Review.*

This systematic review and meta-analysis reported that in recent years AVFs had a high rate of primary failure and low to moderate primary and secondary patency rates. Consideration of these outcomes is required when choosing a patient's preferred access type.

Impact: Updates primary and secondary patency rates of AVF for more contemporary cohorts of dialysis patients. The lower success rates suggests that some patients may not realize the full benefits of AVF that have been previously reported in the KDOQI systematic review.

Oliver MJ, Quinn RR. **Recalibrating vascular access for elderly patients.** *Clin J Am Soc Nephrol. 2014 Apr;9(4):645-7. doi: 10.2215/CJN.01560214. Epub 2014 Mar 20.*

Governments in numerous jurisdictions have set targets for fistula utilization and some have tied reimbursement to attaining these targets. This creates an environment in which it is tempting to overemphasize the benefits of fistulas and the risks of catheters when discussing vascular access options with patients.

Impact: Highlights that not all older patients may benefit from an AVF.

Drew DA, Lok CE, Cohen JT, et al. **Vascular access choice in incident hemodialysis patients: a decision analysis.** *J Am Soc Nephrol. 2015 Jan;26(1):183-91. doi: 10.1681/ASN.2013111236. Epub 2014 Jul 25.*

Decision analysis evaluating AV fistula, AV graft, and central venous catheter (CVC) strategies for patients initiating hemodialysis with a CVC, a scenario occurring in over 70% of United States dialysis patients. An AV fistula attempt strategy was found to be superior to AV grafts and CVCs in regard to mortality and cost for the majority of patient characteristic combinations, especially younger men without diabetes. Women with diabetes and elderly men with diabetes had similar outcomes, regardless of access type. Overall, the advantages of an AV fistula attempt strategy lessened considerably among older patients, particularly women with diabetes, reflecting the effect of lower AV fistula success rates and lower life expectancy. These results suggest that vascular access-related outcomes may be optimized by considering individual patient characteristics.

Impact: Certain patient groups, such as women with diabtes, have lower reported success rates of AVF creation and may have equivalent outcomes with an AVG.

Wish JB. **Catheter last, fistula not-so-first.** *J Am Soc Nephrol. 2015 Jan;26(1):5-7. doi: 10.1681/ASN.2014060594. Epub 2014 Jul 25.*

The issue of vascular access choice is not as black and white as the Centers for Medicare & Medicaid Services (CMS) would like it to appear, with arteriovenous fistula (AVF) always being good or “first” and central venous catheters (CVCs) always being bad or “last.” Nonetheless,CMS has instituted a quality incentive program (QIP) for dialysis providers that rewards high AVF prevalence and penalizes high CVC prevalence without regard to patient mix. For payment year 2014, vascular access constitutes 30% of the total QIP score. This may have already led to access to care issues, as some dialysis providers are refusing to accept patients with CVCs. CMS has recently given ground on this issue by renaming the “Fistula First” initiative “Fistula First Catheter Last” (FFLC) to emphasize that CVC avoidance is as important or more important than AVF use.

Impact: Opinion piece on changes in the Fistula First initiative reflecting the implementation of the current NQF endorsed fistula and catheter vascular access measures in the CMS Quality Incentive Program (QIP). The empahsis of the opinion piece suggests a greater shift to catheter avoidance versus only prioritizing promotion of fistula use.

Grubbs V, Wasse H, Vittinghoff E, et al. **Health status as a potential mediator of the association between hemodialysis vascular access and mortality.** *Nephrol Dial Transplant. 2014 Apr;29(4):892-8. doi: 10.1093/ndt/gft438. Epub 2013 Nov 13.*

Selection of healthier patients for arteriovenous fistula (AVF) placement may explain higher observed catheter-associated mortality among elderly hemodialysis patients. A proportional hazard model was used to examine 117 277 incident hemodialysis patients aged 67-90 years from USRDS for the association of initial vascular access type and 5-year mortality after accounting for health status. Patients with catheter alone had more limited functional status (25.5 versus 10.8% of those with AVF) and 3-fold more prior hospital days than those with AVF (mean 18.0 versus 5.4). In a fully adjusted model including health status, mortality differences between access type were attenuated, but remained statistically significant <AVG [HR 1.18 (1.13-1.22)], catheter plus AVF [HR 1.20 (1.17-1.23)], catheter plus AVG {HR 1.38 [1.26 (1.21-1.31)]} and catheter only [HR 1.54 (1.50-1.58)], P < 0.001>.The observed attenuation in mortality differences previously attributed to access type alone suggests the existence of selection bias. Nevertheless, the persistence of an apparent survival advantage after adjustment for health status suggests that AVF should still be the access of choice for elderly individuals beginning hemodialysis until more definitive data eliminating selection bias become available.

Impact: Underscores the need to adjust for patient characteristics and comorbidities when evaluating the association between vascular access type and outcomes such as mortality.

Lok, Charmaine E & Foley, Robert. **Vascular access morbidity and mortality: trends of the last decade.** *Clin J Am Soc Nephrol. 2013 Jul;8(7):1213-9. doi: 10.2215/CJN.01690213.*

During the past decade, clear trends in the types of incident and prevalent hemodialysis vascular access can be observed. There has been a steady increase and recent stabilizaton of patients initiating hemodialysis with a central venous catheter, representing approximately 80% of all incident accesses. There has also been a steady increase in prevalent fistula use, currently greater than 50% within 4 months of hemodialysis initiation. Patient and vascular access related morbidity and mortality are reflected in the type of vascular access used at initiation and for long-term maintenance dialysis. There is a three- to fourfold increase in risk of infectious complications in patients initiating dialysis with a catheter compared with either a fistula or graft and a sevenfold higher risk when the catheter is used as a prevalent access. Procedure rates have increased two- to threefold for all types of access. There is a significant increased risk of mortality associated with catheter use, especially within the first year of dialysis initiation.

Impact: Despite longstanding KDOQI guidelines, many patients still start hemodialysis with a tunneled catheter and experience higher rates of infectious complications compared to those with an AVF.

Ravani, Pietro & Palmer, Suetonia C & Oliver, Matthew J et al. **Associations between hemodialysis access type and clinical outcomes: a systematic review.** *J Am Soc Nephrol. 2013 Feb;24(3):465-73. doi: 10.1681/ASN.2012070643. Epub 2013 Feb 21.*

Clinical practice guidelines recommend an arteriovenous fistula as the preferred vascular access for hemodialysis, but quantitative associations between vascular access type and various clinical outcomes remain controversial. This systematic review of cohort studies evaluates the associations between type of vascular access (arteriovenous fistula, arteriovenous graft, and central venous catheter) and risk for death, infection, and major cardiovascular events. 67 (62 cohort studies comprising 586,337 participants)studies were selected. In a random effects meta-analysis, compared with persons with fistulas, those individuals using catheters had higher risks for all-cause mortality (risk ratio=1.53, 95% CI=1.41-1.67), fatal infections (2.12, 1.79-2.52), and cardiovascular events (1.38, 1.24-1.54). Similarly, compared with persons with grafts, those individuals using catheters had higher risks for mortality (1.38, 1.25-1.52), fatal infections (1.49, 1.15-1.93), and cardiovascular events (1.26, 1.11-1.43). Compared with persons with fistulas, those individuals

with grafts had increased all-cause mortality (1.18, 1.09-1.27) and fatal infection (1.36, 1.17-1.58), but we did not detect a difference in the risk for cardiovascular events (1.07, 0.95-1.21). The risk for bias, especially selection bias, was high. In conclusion, persons using catheters for hemodialysis seem to have the highest risks for death, infections, and cardiovascular events compared with other vascular access types, and patients with usable fistulas have the lowest risk.

Impact: This study emphasizes that the body of evidence is consistent in the magnitude and direction of effect with regards to the benefits of AVF over central venous catheter.

Moist, Louise M & Lok, Charmaine E & Vachharajani, Tushar J et al. **Optimal hemodialysis vascular access in the elderly patient.** *Semin Dial. 2012 Nov-Dec;25(6):640-8. doi: 10.1111/sdi.12037.*

The optimal vascular access for elderly patients remains a challenge due to the difficulty balancing the benefits and risks in a population with increased comorbidity and decreased survival. Age is commonly associated with failure to mature in fistula and decreased rates of primary and secondary patency in both fistula and grafts. In the elderly, at 1 and 2 years, primary patency rates range from 43% to 74% and from 29% to 67%, respectively. Secondary patency rates at 1 and 2 years range from 56% to 82% and 44% to 67%, respectively. Cumulative fistula survival is no better than grafts survival when primary failures are included. Several observational studies consistently demonstrate a lower adjusted mortality among those using a fistula compared with a catheter; however, catheter use in the elderly is increasing in most countries with the exception of Japan. Both guidelines and quality initiatives do not acknowledge the trade-offs involved in managing the elderly patients with multiple chronic conditions and limited life expectancy or the value that patients place on achieving these outcomes. The framework for choice of vascular access presented in this article considers: (1) likelihood of disease progression before death, (2) patient life expectancy, (3) risks and benefits by vascular access type, and (4) patient preference. Future studies evaluating the timing and type of vascular access with careful assessments of complications, functionality, cost benefit, and patients' preference will provide relevant information to individualize and optimize care to improve morbidity, mortality, and quality of life in the elderly patient.

Impact: Outlines the importance of considering patient factors in vascular access options for elderly patients.

Schmidt, Rebecca J & Goldman, Richard S & Germain, Michael. **Pursuing permanent hemodialysis vascular access in patients with a poor prognosis: juxtaposing potential benefit and harm.** *Am J Kidney Dis. 2012 Dec;60(6):1023-31. doi: 10.1053/j.ajkd.2012.07.020. Epub 2012 Sep 19.*

For patients with end-stage renal disease requiring hemodialysis, the native arteriovenous fistula remains the gold standard of vascular access, with tunneled cuffed central venous catheters reserved for temporary use or as a last resort in patients for whom a permanent vascular access is not possible. It is expected that most patients receiving hemodialysis will be suitable for arteriovenous fistula placement, with suitable patients defined as those: (1) for whom long-term dialysis is expected to confer benefit, (2) with vascular anatomy amenable to arteriovenous fistula placement, and (3) with progressive irreversible kidney failure who are more likely to require dialysis than to die before reaching dialysis dependence. The present article reviews considerations for vascular access decision making, focusing on older patients and those with a poor prognosis, weighing the risks and benefits of arteriovenous fistulas, arteriovenous grafts, and central venous catheters and emphasizing that in the process of vascular access decision making for such patients, medical and ethical obligations to avoid central venous catheters must be balanced by the obligation to do no harm.

Impact: Risks and benefits of arteriovenous fistulas, relative to arteriovenous grafts, and central venous catheters need to be considered, particularly carefully in older patients and those with poor prognosis (limited life expectancy).

Vassalotti, Joseph A & Jennings, William C & Beathard, Gerald A et al. **Fistula first breakthrough initiative: targeting catheter last in fistula first.** *Semin Dial. 2012 May;25(3):303-10. doi: 10.1111/j.1525-139X.2012.01069.x. Epub 2012 Apr 4.*

An arteriovenous fistula (AVF) is the optimal vascular access for hemodialysis (HD), because it is associated with prolonged survival, fewer infections, lower hospitalization rates, and reduced costs. The AVF First breakthrough initiative (FFBI) has made dramatic progress, effectively promoting the increase in the national AVF prevalence since the program's inception from 32% in May 2003 to nearly 60% in 2011. Central venous catheter (CVC) use has stabilized and recently decreased slightly for prevalent patients (treated more than 90 days), while CVC usage in the first 90 days remains unacceptably high at nearly 80%. This high prevalence of CVC utilization suggests important specific improvement goals for FFBI. In addition to the current 66% AVF goal, the initiative should include specific CVC usage target(s), based on the KDOQI goal of less than 10% in patients undergoing HD for more than 90 days, and a substantially improved initial target from the current CVC proportion. These specific CVC targets would be disseminated through the ESRD networks to individual dialysis facilities, further emphasizing CVC avoidance in the transition from advanced CKD to chronic kidney failure, while continuing to decrease CVC by prompt conversion of CVC-based hemodialysis patients to permanent vascular access, utilizing an AVF whenever feasible.

Impact: Emphasizes that catheter avoidance should receive more attention than simply increasing the proportion of patients with an AVF.

Tamura, Manjula Kurella & Tan, Jane C & O'Hare, Ann M. **Optimizing renal replacement therapy in older adults: a framework for making individualized decisions.** *Kidney Int. 2012 Aug;82(3):261-9. doi: 10.1038/ki.2011.384. Epub 2011 Nov 16.*

It is often difficult to synthesize information about the risks and benefits of recommended management strategies in older patients with end-stage renal disease since they may have more comorbidity and lower life expectancy than patients described in clinical trials or practice guidelines. In this review, we outline a framework for individualizing end-stage renal disease management decisions in older patients. The framework considers three factors: life expectancy, the risks and benefits of competing treatment strategies, and patient preferences. We illustrate the use of this framework by applying it to three key end-stage renal disease decisions in older patients with varying life expectancy: choice of dialysis modality, choice of vascular access for hemodialysis, and referral for kidney transplantation. In several instances, this approach might provide support for treatment decisions that directly contradict available practice guidelines, illustrating circumstances when strict application of guidelines may be inappropriate for certain patients. By combining quantitative estimates of benefits and harms with qualitative assessments of patient preferences, clinicians may be better able to tailor treatment recommendations to individual older patients, thereby improving the overall quality of end-stage renal disease care.

Impact: An individualized approach to vascular access decisions that relies on both quantitative assessment of benefits and harms, as well as patient preference, can lead to treatement decisions that contradict practice guidelines.

Ng, Leslie J & Chen, Fangfei & Pisoni, Ronald L et al. **Hospitalization risks related to vascular access type among incident US hemodialysis patients.** *Nephrol Dial Transplant. 2011 Nov;26(11):3659-66. doi: 10.1093/ndt/gfr063. Epub 2011 Mar 3.*

The excess morbidity and mortality related to catheter utilization at and immediately following dialysis initiation may simply be a proxy for poor prognosis. This study examined hospitalization burden related to vascular access (VA) type among incident patients who received some predialysis care using the DOPPS patient cohort (1996-2004) who reported predialysis nephrologist care. VA utilization was assessed at baseline and throughout the first 6 months on dialysis. Poisson regression was used to estimate the risk of all-cause and cause-specific hospitalizations during the first 6 months. Among 2635 incident patients, 60% were dialyzing with a catheter, 22% with a graft and 18% with a fistula at baseline. Compared to fistulae, baseline catheter use was associated with an increased risk of all-cause hospitalization [adjusted relative risk (RR) = 1.30, 95% confidence interval (CI): 1.09-1.54] and graft use was not (RR = 1.07, 95% CI: 0.89-1.28). Allowing for VA changes over time, the risk of catheter versus fistula use was more pronounced (RR = 1.72, 95% CI: 1.42-2.08) and increased slightly for graft use (RR = 1.15, 95% CI: 0.94-1.41). Baseline catheter use was most strongly related to infection-related (RR = 1.47, 95% CI: 0.92-2.36) and VA-related hospitalizations (RR = 1.49, 95% CI: 1.06-2.11). These effects were further strengthened when VA use was allowed to vary over time (RR = 2.31, 95% CI: 1.48-3.61 and RR = 3.10, 95% CI: 1.95-4.91, respectively). A similar pattern was noted for VA-related hospitalizations with graft use. Among potentially healthier incident patients, hospitalization risk, particularly infection and VA-related, was highest for patients dialyzing with a catheter at initiation and throughout follow-up, providing further support to clinical practice recommendations to minimize catheter placement.

Impact: Additional support for the association between catheter use and risk of hospitalization, particularly infection related hospitalizations.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2977_Evidence_form.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
The NKF K/DOQI guidelines state the following: 1) AV fistulas have the lowest rate of thrombosis and require the fewest interventions, 2) cost of AV fistula use and maintenance is the lowest, 3) fistulas have the lowest rates of infection, and 4) fistulas are associated with the highest survival and lowest hospitalization rates. Indeed, a number of epidemiologic studies consistently demonstrate the reduced morbidity and mortality associated with greater use of AV fistulas for vascular access in maintenance

hemodialysis.
As the accompanying literature review indicates, there are a growing number of studies reporting that creating AVF in some patients is less likely to be successful in the presence of certain comorbidities. In addition, certain patient groups may have less incremental benefit from an AV fistula relative to an AV graft. By adjusting the fistula rate for patient characteristics and comorbidities associated with low AV fistula success rates, this measure accounts for patients where a graft or even a catheter may be a more appropriate option.
This measure is intended to be jointly reported with Hemodialysis Vascular Access: Long-term Catheter Rate. These two vascular access quality measures, when used together, consider Arterial Venous Fistula (AVF) use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome. With the growing recognition that some patients have exhausted options for an AVF or have comorbidities that may limit the success of AVF creation, joint reporting of the measures accounts for all three vascular access options. The fistula measure adjusts for patient factors where fistula placement may be either more difficult or not appropriate and acknowledges that in certain circumstances an AV graft may be the best access option. This paired incentive structure that relies on both measures (SFR, long-term catheter rate) reflects consensus best practice, and supports maintenance of the gains in vascular access success achieved via the Fistula First/Catheter Last Project over the last decade.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Analysis of CROWNWeb data from January 2014- December 2014 indicated the facility level mean percentage of patient-months with a fistula was 64.6% (SD=10.4%). Distribution: Min=9.2%, 1st quartile=57.8%, median=64.8%, 3rd quartile=71.7%, Max=97.5%. Information about the data used in these analyses can be found under "Scientific Acceptability".

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Using data from calendar year 2014, age, sex, race and ethnicity were evaluated in a logistic regression model for AV Fistula use. Below are the odds ratios for these patient characteristics. The other covariates included in the model are not shown here as the odds ratios were very similar to those reported in Table 5 of the testing form (risk adjusted model results). Age, sex, race, and ethnicity are all statistically significant predictors of AVF use. Specifically, patients 75 years of age or older were 18% less likely to have an AV fistula when compared to the younger reference group while females are about half as likely to have fistulas as males. Hispanic ethnicity was associated with higher odds of fistula use whereas blacks are about 33% less likely to have fistulas than whites. The analysis results for age, race, and sex indicate potential disparity in fistula use. In the absence of biological effects explaining these differences, risk adjustment for these demographic factors could potentially mask disparities in care.

Age:
For the 18-<25 age group, the Odds Ratio (95% CI) is 1.08 (0.85, 1.36), P-value is 0.542.
For the 25-<60 age group, the Odds Ratio (95% CI) is 1.07 (1.02, 1.12), P-value is 0.005
The 60-<75 age group was used as the reference group.
For the 75+ age group, the Odds Ratio (95% CI) is 0.82 (0.78, 0.87), P-value is <.0001

Sex:
For Female: the Odds Ratio (95% CI) is 0.52 (0.50, 0.54), P-value is <.0001
Male was used as the reference group.

Race:
White was used as the reference group.
For Black: the Odds Ratio (95% CI) is 0.67 (0.63, 0.71), P-value is <.0001
For Other race: the Odds Ratio (95% CI) is 1.07 (0.96, 1.19), P-value is 0.206

Ethnicity:

For Hispanic: the Odds Ratio (95% CI) is 1.16 (1.08, 1.25), P-value is <.0001
Non-Hispanic was used as the reference group.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Numerous studies demonstrate that the use of AV fistulas have the best 5-year patency rates and require the fewest interventions compared with other access types. The advantages of AV fistula over other accesses are clearly delineated in the NKF K/DOQI guidelines, summarized as follows: 1) AV fistulas have the lowest rate of thrombosis and require the fewest interventions, 2) cost of AV fistula use and maintenance is the lowest, 3) fistulas have the lowest rates of infection, and 4) fistulas are associated with the highest survival and lowest hospitalization rates. Indeed, a number of epidemiologic studies consistently demonstrate the reduced morbidity and mortality associated with greater use of AV fistulas for vascular access in maintenance hemodialysis.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. National Kidney Foundation: DOQI Clinical Practice Guidelines for Vascular Access. http://www.kidney.org/Professionals/kdoqi/guideline_upHD_PD_VA/index.htm

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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:** 2977_Data_Dictionary_Code_Table.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The numerator is the adjusted count of adult patient-months using an AVF as the sole means of vascular access as of the last hemodialysis treatment session of the month.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months

S.6. **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, 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.
The number of patient-months using an AVF as the sole means of vascular access at a given facility, adjusted for patient-mix.
An AVF is considered in use if the CROWNWeb "Access Type IDs" of 14 or 22 has been recorded for a given month, where "14" represents AV fistula only (with 2 needles) and "22" represents AV fistula only with an approved single needle device.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) for the entire reporting month at the same facility.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
For each patient, we identify the dialysis provider at each month using a combination of Medicare-paid dialysis claims, the Medical Evidence Form (Form CMS-2728), and data from CROWNWeb. These sources are used to identify patients that are on in-center or home hemodialysis for the entire reporting month. Patients are required to have been treated by the same facility for the complete month in order to be assigned to that facility for the reporting month.
To be included in the denominator for a particular reporting month, the patient must be receiving home or in-center hemodialysis for the complete reporting month at the facility, and be at least 18 years old as of the first day of the month.
The monthly patient count at a facility includes all eligible prevalent and incident patients. The number of patient-months over a time period is the sum of patients reported for the months covered by the time period. An individual patient may contribute up to 12 patient-months per year.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclusions that are implicit in the denominator definition include:
•Pediatric patients (<18 years old)

•Patients on Peritoneal Dialysis
•Patient-months with in-center or home hemodialysis for less than a complete reporting month at the same facility

In addition, the following exclusions are applied to the denominator:
Patients with a catheter that have limited life expectancy:
•Patients under hospice care in the current reporting month
•Patients with metastatic cancer in the past 12 months
•Patients with end stage liver disease in the past 12 months
•Patients with coma or anoxic brain injury in the past 12 months

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Determination of peritoneal dialysis treatment modality is derived from a combination of Medicare-paid dialysis claims, the Medical Evidence Form (Form CMS-2728), and data from CROWNWeb. These sources also determine patient assignment to the facility. Patients not treated by the facility for the entire month are excluded for that reporting month.
The patient's age is determined by subtracting the patient's date of birth from the first day of the reporting month. Patients that are <18 years old as of the first day of the reporting month are excluded.
For the exclusion of catheter patients with limited life expectancy, catheter use in the reporting month is defined as the CROWNWeb "Access Type ID" having any of the following values: (16,18,19,20,21,"·"), where Access_Type_ID "16" represents AV Fistula combined with a Catheter, "18" represents AV Graft combined with a Catheter, "19" represents Catheter only, "20" represents Port access only, "21" represents other/unknown, and "·" represents missing.
Hospice status is determined from a separate CMS file that contains final action claims submitted by Hospice providers. Once a beneficiary elects Hospice, all Hospice related claims will be found in this file, regardless if the beneficiary is in Medicare fee-for-service or in a Medicare managed care plan. Patients are identified as receiving hospice care if they have any final action claims submitted to Medicare by hospice providers in the current month.
Diagnoses of metastatic cancer, end stage liver disease, or coma in the past 12 months were determined from Medicare claims. Medicare claim types include inpatient admissions, outpatient claims (including dialysis claims) and physician services. Claims from providers, such as laboratories that report diagnosis codes when testing for the presence of a condition are excluded. A detailed list of ICD-9/ICD-10 diagnostic codes used to identify these comorbidities is included in the attached data dictionary code table (excel file).

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The proposed SFR measure is a directly standardized percentage, in that each facility's percentage of AVF use is adjusted to the national distribution of covariates (risk factors) (with 'national' here referring to all-facilities-combined). The SFR for facility i is an estimate of what the facility's percentage of AVF would equal if the facility's patient mix was equal to that of the nation as a whole. The measure is adjusted for patient demographic and clinical characteristics based on a logistic regression model. This model includes the facility indicators and assumes that the regression coefficients of risk factors are the same across all facilities. The common risk effects are assumed in order to improve computational stability in estimating facility-specific effects.

The patient characteristics included in the logistic regression model as covariates are:
•Age
•BMI at incidence
•Nursing home status in previous year
•Nephrologist's care prior to ESRD
•Duration of ESRD

•Inability to ambulate/transfer at ESRD incidence (CMS-2728 form)
•Comorbidities either at ESRD incidence (CMS-2728 form) or prevalent comorbidities based on Medicare claims filed in prior 12 months
  oDiabetes
  oHeart diseases
  oPeripheral vascular disease
  oCerebrovascular disease
  oChronic obstructive pulmonary disease
  oAnemia (unrelated to ESRD/CKD)
  oNon-Vascular Access-Related Infections
  oDrug dependence
•Indicator for Medicare coverage for at least 6 months during the past 12 months

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
See calculation flowchart in Appendix.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patients with a missing vascular access type are counted in the denominator, but not the numerator. For comorbidities, if the patient had missing comorbidity values both in the preceding 12 months of Medicare claims and in the Medical Evidence Form for the corresponding comorbidity, we assume this patient did not have the comorbidity in that reporting month. The same methodology is applied to the comorbidity exclusions and the hospice exclusion.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.24.
Administrative claims, Electronic Clinical Data

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
CROWNWeb, Medicare Claims and the CMS Medical Evidence form 2728 are used as the data sources for establishing the denominator. CROWNWeb is the data source for establishing the numerator. Medicare claims and the CMS Medical Evidence form 2728 are data sources for the risk adjustment factors. Medicare claims and CROWNWeb are used for the exclusion criteria.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Dialysis Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2977_NQF_Testing_form.docx

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)
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)*
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.**

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.**
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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
N/A

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Hemodialysis Vascular Access: Standardized Fistula Rate
Date of Submission: 4/15/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *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.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or

whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/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*).

National CROWNWeb data from January 2014-December 2014 and Medicare claims data from January 2013 – December 2014

1.3. What are the dates of the data used in testing? January 2013-December 2014

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.26***)	**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 <u>measured entities</u> 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*)

Patients on both home and in-center hemodialysis during the last HD treatment of the month from January 2014- December 2014 were included in the analyses. The number of facilities per month ranged from 5,736-5,871 and the total number of patients per month ranged from 344,945- 363,257.

Public reporting of this measure on DFC or in the ESRD QIP would be restricted to facilities with at least 11 eligible patients throughout the year for the measure. We have applied this restriction to all the reliability and validity testing reported here.

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*)

There were a total of 4,274,619 eligible patient-months. Among those patient-months over the whole year, the average age was 62.7 years, 43.8% of patient-months were female, 56.3% were white, 37.1% were black, 66.7% had race listed as other, 18.4% were Hispanic and 46.4% had type II diabetes as the primary cause of ESRD.

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.

N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patient level:

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

**Assessed at a specific time point (e.g., at the reporting month). Medicare coverage in model was defined as:*
1. Medicare as primary and Medicaid
2. Medicare as primary and NO Medicaid
3. Medicare as secondary or Medicare HMO (e.g. Medicare Advantage)
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
- Income disparity
- Families below the poverty level (%)
- Single-parent households with children <18 years old (%)
- Home ownership rate (%)
- Median home value
- Median monthly mortgage
- Median gross rent
- Population (aged 25+) with <9 years of education (%)
- Population (aged 25+) without high school diploma (%)

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*)

We used January 2014 – December 2014 CROWNWeb data to calculate facility-level annual performance scores. The NQF-recommended approach for determining measure reliability is a one-way analysis of variance (ANOVA), in which the between-facility variation (σ_b^2) and the within-facility variation ($\sigma_{t,w}^2$) in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of a measure (i.e., $\sigma_b^2 + \sigma_{t,w}^2$) that is attributable to the between-facility variation, the true signal reflecting the differences across facilities. We assessed reliability by calculating inter-unit reliability (IUR) for the annual performance scores. A small IUR (near 0) reveals that most of the variation of the measure 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 Standardized Fistula Rate (SFR) for N facilities. Since the variation in $T_1,\ldots,T_N$ is mainly driven by the estimates of facility-specific intercepts ($\alpha_1,\ldots, \alpha_N$), we use their asymptotic distributions to estimate the within-facility variation in SFR. Applying the delta method, we estimate the variance of T_i and denote the estimate as S_i^2. Calling on formulas from the one-way ANOVA, the within-facility variance in SFR can be estimated by

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

and the total variation in SFR can be estimated by

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

where n_i is the number of subjects in the *i*th facility, $\bar{T} = \Sigma\, n_i\, T_i / \Sigma\, n_i$, 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). 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 reliability of SFR calculation only included facilities with at least 11 patients during the entire year.

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 IUR is 0.736 which indicates that about 74% of the variation in the annual SFR can be attributed to between-facility differences in performance (signal) and about 26% to the within-facility variation (noise).

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?*)

The result of IUR testing suggests a high degree of reliability.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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 quintiles of performance scores and the 2014 Standardized Mortality Ratio (SMR, NQF 0369) and 2014 Standardized Hospitalization Ratio (SHR, NQF 1463). Facility-level performance scores were divided into

quintiles (Q1 to Q5) and the relative risk (RR) of mortality (and hospitalization, separately) was calculated for each quintile, using the combined Q4 and Q5 as the reference group. Thus, a RR>1.0 would indicate a higher relative risk of mortality or hospitalization, compared to the highest performance score quintiles.

In 2015 a vascular access TEP was convened to provide input on the development of access measures, and specifically input on exclusions for both catheter and fistula measures, and for fistula, risk adjustment factors to be considered. Ultimately, evaluation and selection of the clinical and patient risk factors for this measure was informed by the final TEP recommendations. The TEP recognized that while fistulas are preferred, an unintended consequence of a fistula measure that doesn't account for the patient's overall health status could harm patients by subjecting them to fistula surgery that is less likely to succeed or limit access to care for patients with more comorbidities. To accomplish this goal the TEP discussed adjusting the measure for conditions or scenarios where a graft may be an acceptable or preferred alternative to a fistula. The candidate measure was reviewed and validated by the Technical Expert Panel (TEP) in 2015.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Quintiles of the performance scores were defined as follows:

Q1: 9.2%-<55.9%

Q2: 55.9%-<62.0%

Q3: 62.0-<67.4%

Q4*: 67.4%-<73.4%

Q5*: 73.4%-<97.5%

*Q4 and Q5 as Reference

Results from the Poisson model indicated that the percent of patient-months with a fistula was significantly associated with the risks of mortality and hospitalization.

For the 2014 SMR, the relative risk of mortality increased as the performance measure quintile decreased from the reference group (combined Q4 and Q5) with the highest risk in quintile 1. For quintile 3, RR=1.03 (95% CI: 1.01, 1.05; p=0.003), quintile 2, RR=1.05 (95% CI: 1.03, 1.07; p<0.001), and quintile 1, RR=1.10 (95% CI: 1.08, 1.12; p<0.001).

Similarly for 2014 SHR, the relative risk of hospitalization increased as the performance measure quintile decreased from the reference group (combined Q4 and Q5) with the highest risk in quintile 1. For quintile 3, RR=1.07 (95% CI: 1.06, 1.07; p<0.001), quintile 2, RR=1.08 (95% CI: 1.08, 1.09; p<0.001), and quintile 1, RR=1.11 (95% CI: 1.11, 1.11; p<0.001).

2b2.4. 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?*)

These results of the Poisson regression suggest the predictive relationship of lower fistula use with higher

mortality and hospitalization, as measured by the respective standardized mortality and hospitalization rates, and compared to facilities with higher fistula use.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

The following exclusions are applied to the denominator:

Patients with a catheter that have limited life expectancy. Limited life expectancy is defined as:

- Patients under hospice care in the current reporting month
- Patients with metastatic cancer in the past 12 months
- Patients with end stage liver disease in the past 12 months
- Patients with coma or anoxic brain injury in the past 12 months

The facility-level standardized fistula rate with and without the patient-month exclusions are calculated and compared.

2b3.2. 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*)

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

Table 1: Percent of patient-months at risk excluded

Year	Before Exclusion	After Exclusion	Percent
2014	4,314,450	4,274,619	0.92%

Table 2: Number and percent of unique patients excluded

Year	Before Exclusion	After Exclusion	Percent
2014	468,910	457,902	2.35%

Table 3: Distribution of performance scores before and after the exclusion

Standardized Fistula Rate	N	Mean	Standard Deviation	Minimum	Maximum
Before exclusion	5928	0.640	0.103	0.092	0.975
After exclusion	5928	0.646	0.104	0.092	0.975

Figure 1: Scatterplot –SFR with and without measure exclusions

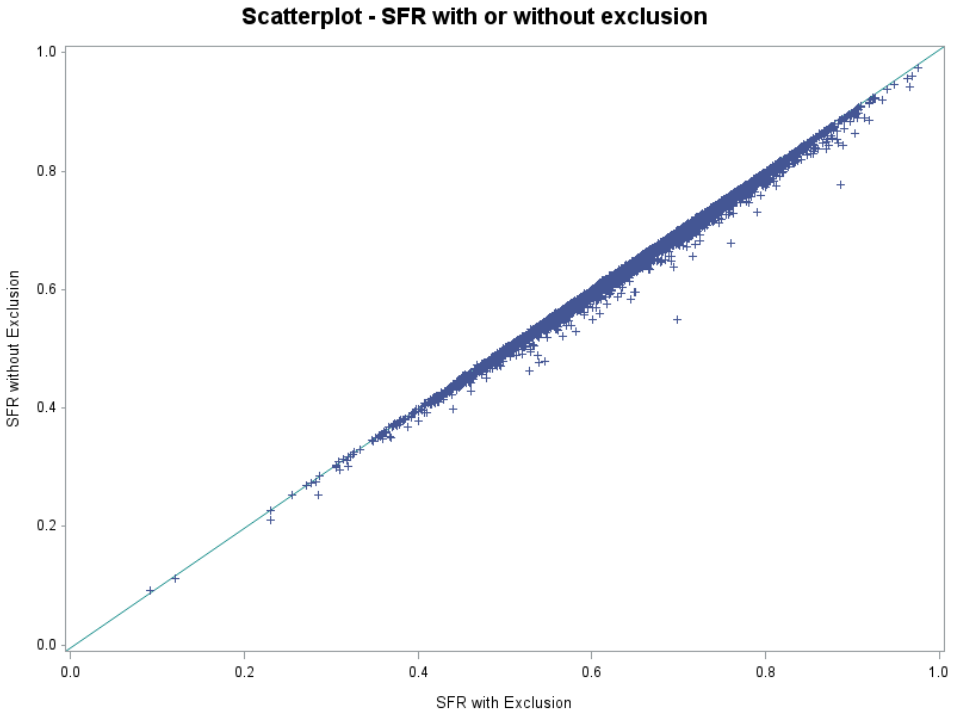
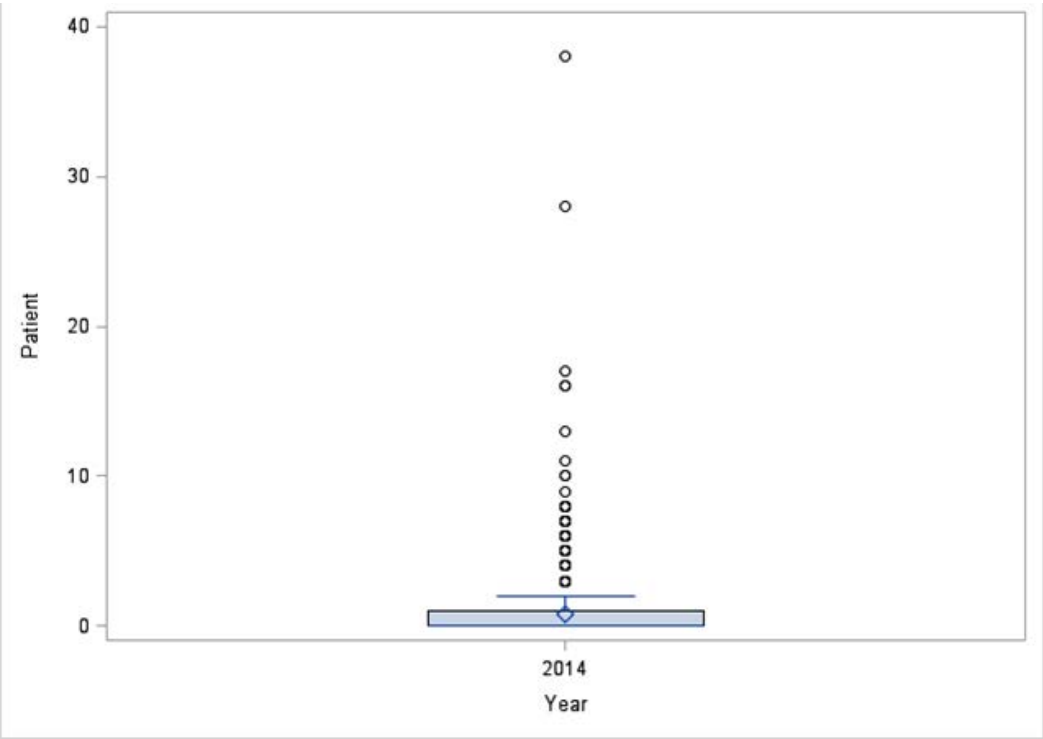


Figure 2. Distribution of Excluded Patients at facility level for 2014



2b3.3. 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.* <u>*Note*</u>*: **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 exclusion criteria are necessary since the percentage of patients excluded at each facility is not evenly distributed across facilities (Distribution shown in the boxplot above). Due to the unequal distribution across facilities, the exclusion criteria take into account that some facilities treat a higher portion of patients with limited life expectancy. Additionally, our results shown in both the scatter-plot (Figure 1) as well as the Pearson Correlation Coefficient of 0.998 (p-value <0.0001) between SFRs with and without the exclusion suggests that the overall impact of the exclusion on the measure's validity is not substantial since the two are highly correlated.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 19 patient-month level risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Although there have been significant gains in the proportion of dialysis patients that have an AV fistula, it is generally recognized that some patients on hemodialysis will need to have an AV graft or even a catheter. As evidence, the CMS AV fistula target at the facility level is 68%, rather than 100%, which recognizes that one third of patients will require a different type of access. Given that there is variation in the burden of comorbidities between different facilities, adjusting for these factors when calculating an AV fistula rate implicitly recognizes that some patients are more likely to have AV grafts. Several of the studies (listed in 1a.7.9) detail particular patient characteristics that are associated with a decreased likelihood of having a successful AV fistula created. Ultimately, evaluation and selection of the clinical and patient risk factors was informed by the final TEP recommendations. The TEP recognized that while fistulas are preferred, an unintended consequence of a fistula measure that doesn't account for the patient's overall health status could harm patients by subjecting them to fistula surgery that is less likely to succeed or limit access to care for patients with more comorbidities. The TEP recognized that they could not make the statement that fistulas and grafts are truly equivalent in all patients, but wanted to ensure that grafts were a strongly preferred outcome to catheters and should not be disincentivized. To accomplish this goal the TEP discussed adjusting the measure for conditions or scenarios where a graft may be an acceptable or preferred alternative to a fistula. The covariates in the final model represent a combination of those recommended by the TEP for inclusion as well as factors that empiric analyses indicated were predictive of AV fistula use. Final decisions of the risk factors were based on both the clinical and statistical association with the lower likelihood of fistula use in patients with these risk factors, and that these factors were not likely to be associated with facility care.

Risk adjustment is based on a multivariate logistic regression model. The adjustment is made for age, BMI at incident, nursing home status, nephrologist's care prior to ESRD, duration of ESRD, diabetes as primary cause of ESRD, and comorbidities. Although covariates are assumed to have the same effects across facilities, the adjustment model is fitted with different facility effects (through facility-specific intercept terms), which provides valid estimates even if the distribution of adjustment variables differs across facilities. The common risk effects are assumed in order to improve computational stability in estimating facility-specific effects. All analyses are done using SAS.

In general, adjustment factors for the SFR were selected based on several considerations. We began with a large set of patient characteristics, including demographics, comorbidities at ESRD incidence or past 12 months, and other characteristics. We used an indicator to identify patients with < 6 months of Medicare coverage in the past 12 months as part of the analyses that relied on Medicare claims for comorbidities. Factors considered appropriate were then investigated with statistical models to determine if they were related to AVF use. Factors related to the SFR were also evaluated for face validity before being included.

We used two data sources to collect comorbidity information: CMS-2728 and Medicare claims filed in prior 12 months. The covariates for comorbidities included in the final model take a value of 1 if there was any evidence of the condition in either CMS-2728 or Medicare claims, otherwise 0. Some patient characteristics or comorbidities are only available in CMS-2728, some are only available in Medicare claims, and some are available from both sources. We considered the condition to be present if it was noted in either the CMS-2728 form, or Medicare claims, or both. Table 4 shows that most all of the comorbidities had a statistically significant association with AVF use. As a comparison, using data from January 2014 we compared analysis results of two additional risk adjustment models that included: 1) no comorbidity adjustment at all (denoted as Model 0), and 2) comorbidities defined by CMS-2728 only (denoted as Model 1). Table A1 in the Appendix shows that the c-statistic of our final model was the highest, compared with Model 0 and Model 1 (c-statistic=0.688 for Model 0; 0.691 for Model 1; and 0.700 for our final model). In Table A2 of the Appendix, some of regression coefficients (especially for age, nursing home status and peripheral vascular disease) increased or decreased from those in Models 0 and 1.

In response to the requirements for NQF's Trial Period for the assessment of sociodemographic risk adjustment factors for quality measures, we investigated several patient and zip code level data elements (see list in 1.8). Sociodemographic factors included in the analysis were based on conceptual criteria and empirically demonstrated findings in the literature which have shown differences in fistula use exist among racial minorities, women and the poor. In addition, the particular patient and area level variables chosen were based on availability of data for the analyses. We were able to acquire individual area-level variables included in the Area Deprivation Index (ADI) developed by Singh and colleagues at the University of Wisconsin[1].

1. Singh, GK. Area deprivation and widening inequalities in US mortality, 1969–1998. Am J Public Health. 2003;93(7):1137–1143.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

In the table below, we list results from the adjusted model described above. For a given covariate, the regression coefficient represents the logit of the rate. We also report the odds ratio for each covariate. With a few exceptions (youngest age group, other heart diseases and anemia), all main effects are statistically significant at the 0.05 level.

Table 4. Model Coefficients and Odds Ratios, Data Year 2014

Covariate	Coefficient	Odds Ratio	P-value
Age			
18-<25	0.073	1.076	0.530
25-<59	0.087	1.091	0.000
60-<75	reference		
75+	-0.202	0.817	<.0001
BMI			
underweight(< 18.5)	-0.215	0.806	0.001
normal(18.5 - 24.9)	reference		

Covariate	Coefficient	Odds Ratio	P-value
overweight(>24.9)	0.054	1.055	0.026
Nursing home status*	-0.321	0.726	<.0001
Nephrologist's Care prior to ESRD*	0.257	1.293	<.0001
Duration of ESRD			
<1 year	-1.171	0.310	0.000
1-<5 years	reference		
5-<9 years	-0.234	0.792	0.000
9+	-0.602	0.548	0.000
Primary Cause of ESRD			
Diabetes	-0.053	0.948	0.034
Other	reference		
Comorbidities*			
Diabetes (NOT as primary cause of ESRD)	-0.121	0.886	<.0001
Heart Failure	-0.046	0.955	0.038
Other Heart Diseases	-0.037	0.963	0.114
Peripheral Vascular Disease	-0.340	0.712	<.0001
Cerebrovascular Disease	-0.113	0.893	<.0001
Chronic Obstructive Pulmonary Disease	-0.083	0.921	0.001
Drug Dependence	-0.207	0.813	<.0001
Inability to ambulate/transfer	-0.497	0.609	<.0001
Anemia (unrelated to ESRD/CKD)	-0.049	0.952	0.228
Non-Vascular Access-Related Infections: Pneumonia/Hepatitis/HIV/Tuberculosis	-0.286	0.751	<.0001
Less than 6-months of Medicare coverage in past 12 months	-0.447	0.640	<.0001
* 'No' was used as reference.			

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

The table below shows the parameter estimates for patient and area level SDS/SES variables based on a logistic regression model for AV fistula use that included all these variables along with all the other clinical covariates used for adjustment in SFR. Here we only report results for the SDS/SES factors.

Table 5. Coefficients and odds ratios for SDS/SES variables

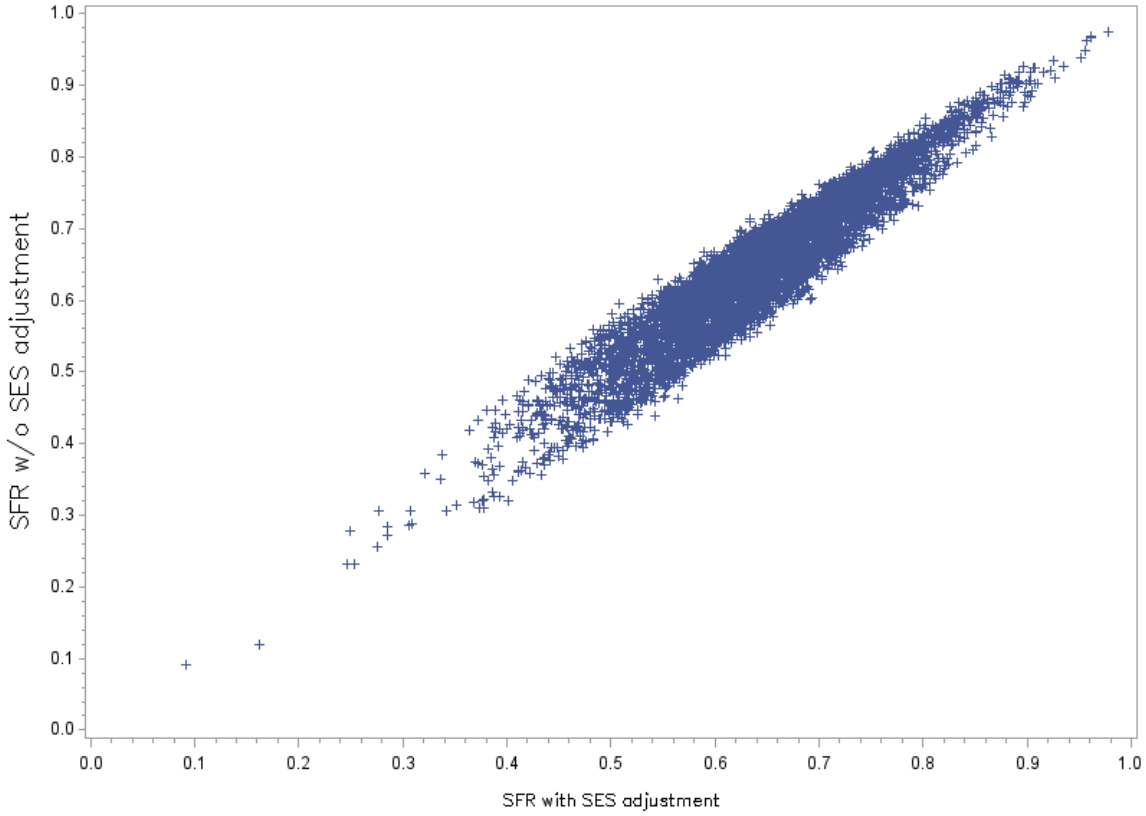
Variable	**Estimate**	**Odds Ratio**	**P-value**
Sex			
Female	-0.647	0.524	<.0001
Male	Reference		
Ethnicity			
Hispanic	0.160	1.173	<.0001
Non-Hispanic	Reference		
Race			
White	Reference		
Black	-0.406	0.666	<.0001
Other	0.081	1.085	0.126
Employment Status (2728)			
Employed	Reference		
Unemployed	-0.139	0.870	<.0001
Other	-0.172	0.842	<.0001
Medicare Coverage			
Medicare as primary without Medicaid	Reference		
Medicare as primary with Medicaid	-0.008	0.992	0.770
Medicare as secondary or Medicare HMO	0.044	1.045	0.135
Non-Medicare/missing	-0.212	0.809	<.0001
ADI (zipcode_level)			
Unemployment rate (%)	0.003	1.003	0.688
Median family income	-0.005	0.995	0.695
Families below the poverty level (%)	0.001	1.001	0.861
Single-parent households w/ children <18 (%)	-0.001	0.999	0.590
Home ownership rate (%)	0.000	1.000	0.929
Median home value	0.004	0.948	0.871
Median monthly mortgage	-0.002	0.998	0.981
Median gross rent	-0.037	0.964	0.680
Population (aged 25+) without High School diploma (%)	0.002	1.002	0.589
Income disparity	-0.026	0.975	0.434

Patient-level SDS/SES: Compared to males, females were less likely to have an autogenous arteriovenous fistula (AVF) (OR=0.52; p<0.01). Hispanics were more likely to have an AVF (OR=1.17; p<0.01), compared to non-Hispanics. Compared to white patients, black patients were less likely to have an AVF (OR=0.67, p<0.01). As for employment status, unemployed and other patients were less likely to have an AVF (OR=0.87; p<0.01; OR=0.84; 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. Compared to Medicare only patients, patients with both Medicare and Medicaid or patients with Medicare as secondary coverage/Medicare HMO have no significant difference in fistula use (OR=0.99, p=0.77; OR=1.05, p=0.14), while patients classified as no Medicare/missing were less likely to have an AVF (OR=0.81, p<0.01).

Area-level SDS/SES: Area-level effects were generally very small and were not statistically significant.

Figure 3. Correlation between SFRs with and without SDS/SES adjustment



The standard and SDS/SES-adjusted SFRs were highly correlated at 0.95 (p<.001).

	SFR with SDS/SES			
SFR w/o SDS/SES	**Better than expected**	**As expected**	**Worse than expected**	**Total**
Better than expected	196 (3.3%)	65 (1.1%)	0	261 (4.4%)
As expected	64 (1.1%)	5409 (91.2%)	53 (0.9%)	5526 (93.2%)
Worse than expected	0	52 (0.9%)	89 (1.5%)	141 (2.4%)
Total	260 (4.4%)	5526 (93.2%)	142 (2.4%)	5928

After adjustment for SDS/SES, 234 facilities (4.0%) changed performance categories. 118 (2.0%) facilities were down-graded and 116 (2.0%) facilities were upgraded.

These analyses indicate that patient-level, but not area-level, variables for SDS/SES impact expected AVF rates. Furthermore, we observed that adjustment for SDS/SES shifted facility performance, but more facilities declined in performance ranking with SDS/SES adjustment than improved.

Area-level factors are not included as adjustments due to the absence of clinically meaningful or statistically observed differences on the fistula rate with these adjustments. Patient-level SDS/SES variables are not included as adjustments in the measure due to the absence of a convincing biological or clinical rationale that warrant accounting for different outcomes on the basis of race, sex, or socioeconomic status. For example, some providers in the dialysis community believe that women are less likely to have AVF due to smaller vessels and hypothesize that this may be a biologic explanation for subsequent higher primary failure rates seen in women. While several studies have reported that women have smaller vasculature than men [1,2] , this has

not been a consistent finding, and may be isolated to forearm vessels. Studies that have focused on upper arm AVF have demonstrated similar AVF rates between men and women, suggesting a lack of sufficient biological or clinical support for different outcomes in fistula rates between female and males [3].

1. Jemcov, TK Morphologic and functional vessel characteristics assessed by ultrasonography for prediction of radiocephalic fistula maturation. J Vasc Access 2013; 14(4):356-363
2. Allon, M et al. Effect of preoperative sonographic mapping on vascular access outcomes in hemodialysis patients. Kidney International, Vol. 60 (2001), pp. 2013–2020
3. Caplin, N et al. Venous Access:Women Are Equal. Am J Kidney Dis 2003. 41:429-432.

2b4.5. 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 2b4.9

Risk factors were selected for the final model based on the magnitude of the coefficients, evaluation of their statistical significance, and the model c-statistic.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

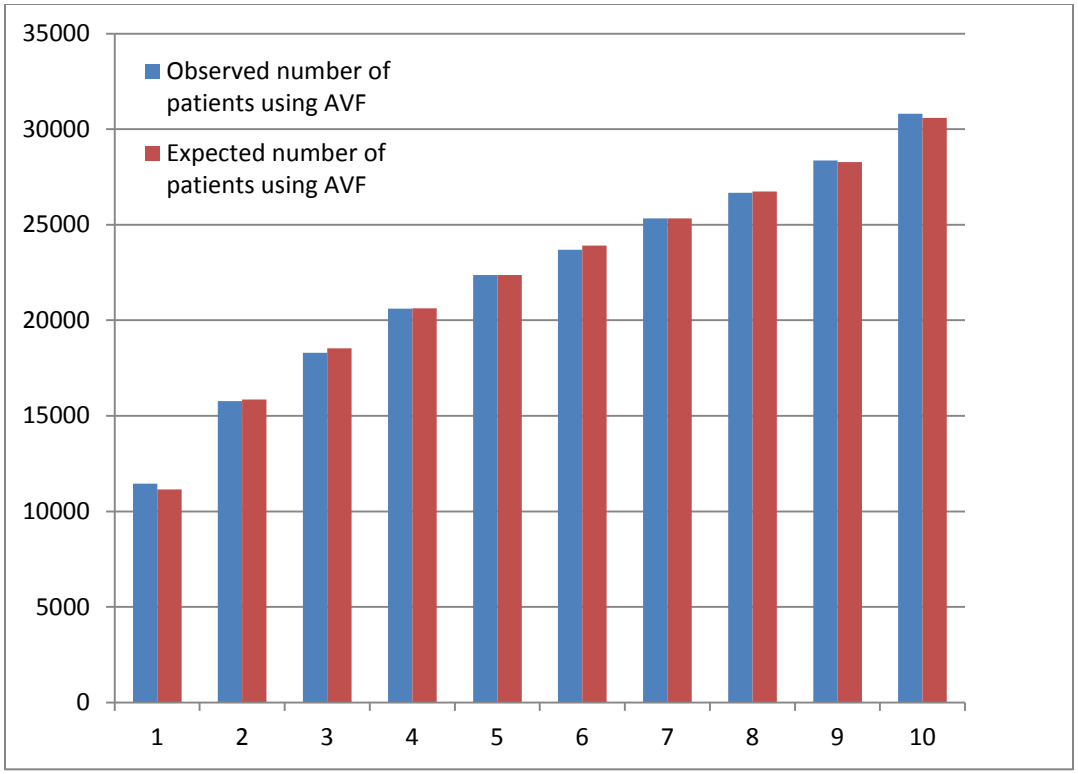
The C-statistic (also known as the Index of Concordance) was 0.70. This indicates that the model correctly ordered 70% of the pairs of patient-months that were discordant with respect to the response variate.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

The Hosmer-Lemeshow test statistic based on deciles of risk is 39.1 with p-value <0.0001. In very large samples such as this even relatively small departures from the model will lead to significant results. The c-statistic and risk decile plot show that the model provides an overall good fit to the data.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Figure 4: Decile plots for the number of patients using AVF



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. 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*)

The decile plot (Figure 4) shows that the risk factors in the model are discriminating well between patients. There is good separation among all 10 groups by risk scores, and the ordering is as predicted by the model (i.e., patients predicted to have a lower probability of AVF use actually do have a lower percentage of AVF use). The absolute differences between the risk groups are also large, with patients predicted to have the highest likelihood of AVF use (Group 10) having 3 times higher AVF rate than those predicted to have the lowest likelihood (Group 1). This means that the model fit is good and therefore adequately adjusts for patient characteristics (case mix).

2b4.11. 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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

Differences in measure performance were evaluated separately for each facility, where the annual standardized fistula rate (SFR) of each facility was compared to the overall national distribution.

Here we describe our approach for testing of statistical significance. Let $T_1,\ldots,T_N$ be the Standardized Fistula Rate (SFR) for N facilities. Since the variation in $T_1,\ldots,T_N$ is mainly driven by the estimates of facility-specific intercepts ($\alpha_1,\ldots, \alpha_N$), we use their asymptotic distributions and apply the delta method to estimate the standard errors of SFRs. Let S_i denote the standard error estimate of T_i. The test-statistic is then calculated by (T_i - national average of SFR)/ S_i, which asymptotically follows the standard normal distribution under the null hypothesis. A two-sided test with significant level 0.05 was used. As the reference null distribution, we used Efron's empirical null distribution in lieu of the theoretical null distribution since the empirical null method is more robust approach that takes account of the national random variation among facilities not accounted for in the model (Efron, 2004; Kalbfleisch and Wolfe, 2013). It essentially rescales the critical value for the test statistic. The rescaling multiple is estimated by the slope (estimated via robust regression) correlating the empirical and theoretical Z-score quantiles (e.g., with a multiple of 1 indicating that in fact no rescaling is required). In this approach, facilities are flagged if they have outcomes that are more extreme when compared to the variation in national AVF rate.

Efron, Bradley. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. Journal of the American Statistical Association. Vol. 99, No. 465 (Mar., 2004), pp. 96-104

Kalbfleisch JD, Wolfe RA. On monitoring outcomes of medical providers. Statistics in the Biosciences. November 2013, Volume 5, Issue 2, pp 286-302

2b5.2. 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*)

Proportion of facilities with statistically significant differences (p-values < 0.05) is shown as follows:

Category	Number of facilities	Percent of facilities
Better than expected	261	4.40%
As expected	5526	93.22%
Worse than expected	141	2.38%

2b5.3. 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?*)

For the annual SFR, 5,526 (93%) facilities have achieved expected performance, 141 (2.4%) facilities have performed worse than expected (lower fistula rate), 261 (4.4%) facilities have performed better than expected (higher fistula rate).

In general, a higher rate of fistula use represents better quality of care. This analysis demonstrates both practical and statistically significant differences in performance across facilities based on their adjusted proportion of patient months with a fistula in use.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

2b7.2. 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*)

2b7.3. 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*)

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.

Planned	Current Use (for current use provide URL)
Public Reporting Payment Program	

4a.1. For each CURRENT use, checked above, provide:

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

N/A

4a.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?*)

Measure is currently under development.

4a.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.*)

CMS will determine if and when the measure will be implemented in a CMS program. Upon endorsement, CMS will consider retiring the currently endorsed measure of fistula use (#0257) in favor of this new measure for implementation in a future performance year for the ESRD QIP and reporting period for Dialysis Facility Compare at the next available opportunity.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **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.**
The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of the standardized fistula rate. Once implemented facility performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in promoting fistula use for the incident and prevalent populations, while also taking into account those patient risk factors that hinder successful fistula use in certain subpopulations.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Potential unintended consequences would center on patients being pushed towards having an AVF created when they may not realize a benefit from this type of access due to either comorbidities or a limited life expectancy. It is incumbent on both dialysis facilities, as well as surgeons, to establish the most appropriate type of vascular access for patients based on their individual circumstances and preferences, rather than in response to quality measures.

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.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0251 : Vascular Access—Functional Arteriovenous Fistula (AVF) or AV Graft or Evaluation for Placement
2594 : Optimal End Stage Renal Disease (ESRD) Starts

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measure 0251 contains several components in addition to assessing fistula use. It is a referral process measure. The most basic requirement to get into the numerator is referral to a vascular surgeon (or other qualified physician). This has the potential for facilities to score well on the measure separate from whether patients are receiving treatment with a fistula, graft, or catheter, as long as the patient was referred to or evaluated by a vascular surgeon. We acknowledge this is an important step to fistula placement however it departs from the intent of this fistula measure to function as a more direct incentive to encourage fistula use. Moreover, consistent with the concerns and recommendations made by the vascular access TEP, the SFR is risk adjusted and includes risk factors to account for patients where fistula may not be the appropriate access type. Measure 2594 is not directed toward dialysis facilities. The setting focus addresses a different provider type which falls outside the purview of measures evaluating dialysis facility performance on fistula use. This suggests a fundamental difference in the measure target populations, setting and intent that cannot be harmonized. Additionally, the measure is limited to incident patients, while the SFR includes both incident and prevalent patients as the measured population.

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.)
There are no competing measures.

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:** Appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, Sophia.chan@cms.hhs.gov
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.
According to the CMS Measure Management System Blueprint, TEPs are advisory to the measure contractor. 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.

Joseph Vassalotti, MD, FASN, FNKF
Chief Medical Officer, National Kidney Foundation
Associate Professor of Medicine, Division of Nephrology

Mount Sinai Medical Center
New York, NY

Monet Carnahan, RN, BSN, CDN
Renal Care Coordinator Program Manager
Fresenius Medical Center (FMC)
Franklin, TN
American Nephrology Nurses Association

Derek Forfang
Patient Leadership Committee Chair
ESRD Network 17
Board Member
Intermountain End State Renal Disease Network Inc.
Beneficiary Advisory Council (Vice Chair)
The National Forum of ESRD Networks
Board Member
The National Forum of ERSD Networks
San Pablo, CA

Lee Kirskey, MD
Attending staff, Department of Vascular Surgery
Cleveland Clinic Foundation
Cleveland, OH

Nance Lehman
Board Member
Dialysis Patient Citizens (DPC)
Billings, MT

Charmaine Lok, MD, MSc, FRCPC (C)
Medical Director of Hemodialysis and Renal Management Clinics
University Health Network
Professor of Medicine
University of Toronto
Toronto, ON

Lynn Poole, FNP-BC, CNN NCC
Fistula First Catheter Last Project Clinical Lead
ESRD National Coordinating Center
Lake Success, NY

Rudy Valentini, MD
Chief Medical Officer
Children's Hospital of Michigan (CHM)
Professor of Pediatrics, Division of Nephrology
Wayne State University School of Medicine

Daniel Weiner, MD, MS
Nephrologist, Tufts Medical Center
Associate Medical Director
DCI Boston
Associate Professor of Medicine
Tufts University School of Medicine
Boston, MA

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, 2017
Ad.6 **Copyright statement:** Ad.7 **Disclaimers:**
Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2978
Measure Title: Hemodialysis Vascular Access: Long-term Catheter Rate
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: Percentage of adult hemodialysis patient-months using a catheter continuously for three months or longer for vascular access.
Developer Rationale: Based upon data from the CMS Fistula First/Catheter Last initiative, a gradual trend towards lower catheter use has been observed among prevalent maintenance HD patients in the US, declining from approximately 28% in 2006 to approximately 18% by August 2015. Furthermore, the percentage of maintenance HD patients using a catheter for at least three months has declined as well over this time period from nearly 12% to 10.8%. Continued monitoring of chronic catheter use is needed to sustain this trend.
This measure is intended to be jointly reported with the Hemodialysis Vascular Access- Standardized Fistula Rate. These two vascular access quality measures, when used together, consider Arterial Venous (AV) fistula use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome. With the growing recognition that some patients have exhausted options for an arteriovenous fistula, or have comorbidities that may limit the success of AVF creation, joint reporting of the measures accounts for all three vascular access options. The fistula measure adjusts for patient factors where fistula placement may be either more difficult or not appropriate and acknowledges that in certain circumstances an AV graft may be the best access option. This paired incentive structure that relies on both measures reflects consensus best practice, and supports maintenance of the gains in vascular access success achieved via the Fistula First/Catheter Last Project over the last decade.

Numerator Statement: The numerator is the number of adult patient-months in the denominator who were on maintenance hemodialysis using a catheter continuously for three months or longer as of the last hemodialysis session of the reporting month.
Denominator Statement: All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) for the complete reporting month at the same facility.
Denominator Exclusions: Exclusions that are implicit in the denominator definition include:
-Pediatric patients (<18 years old)
-Patients on Peritoneal Dialysis
-Patient-months under in-center or home hemodialysis for less than a complete reporting month at the same facility

In addition, the following exclusions are applied to the denominator:
Patients with a catheter that have limited life expectancy:
-Patients under hospice care in the current reporting month
-Patients with metastatic cancer in the past 12 months
-Patients with end stage liver disease in the past 12 months
-Patients with coma or anoxic brain injury in the past 12 months

Measure Type: Intermediate Clinical Outcome
Data Source: Administrative claims, Electronic Clinical Data
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- The developer's rationale for this intermediate clinical outcome measure is that there is an association between type of vascular access used for hemodialysis and patient mortality; this measure focuses on the process of assessing long term catheter use at chronic dialysis facilities. The developer provides evidence suggesting that long term catheter use is correlated with the highest mortality risk and arteriovenous fistula use has the lowest risk of mortality. Arteriovenous grafts (AVG) have been found to have a risk of death that is higher than AVF but lower than catheters.
- The developer notes that this measure is intended to be jointly reported with the Hemodialysis Vascular Access-Standardized Fistula Rate. Used together, the two vascular access quality measures consider Arterial Venous (AV) fistula use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome.
- The developer provides the following linkage between the measure focus and a desired health outcome: Measure long term catheter rate→ Assess value →Identify patients who do not have an AV Fistula or AV graft→Evaluation for an AV fistula or graft by a qualified dialysis vascular access provider →Increase Fistula/Graft Rate → Lower catheter rate →Lower patient mortality.
- The developer references clinical practice guidelines (NKF: KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates). The order of preference for placement of fistulae in patients with kidney failure who choose HD as their initial mode of KRT are listed in descending order. The guidelines have all been assigned a Grade B (moderately strong evidence) rating.

Guidance from the Evidence Algorithm:
Intermediate outcome (Box 3) → SR with QQC (Box 4) → concludes "moderately strong"

Questions for the Committee:

- *Does the Committee agree there is no change in the evidence since the last evaluation?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- An analysis of CROWNWeb data from January 2014-December 2014 indicates the facility-level mean percentage of patient-months with a long-term catheter was 11.6% (SD=6.6%).

Minimum	1st Quartile	Median	3rd Quartile	Maximum
0%	7.0%	10.5%	14.9%	58.2%

Disparities

- 2014 data indicate age, sex, race, ethnicity and dialysis vintage were evaluated in a logistic regression model for long-term catheter use. The analysis results indicate potential disparity in prolonged use of a tunneled catheter:
    - Specifically, females are about 55% more likely to have a long-term catheter as males.
    - Individuals 75 years of age and older were 14% more likely to have a long-term catheter and younger individuals 18-25 years of age were 46% more likely to have a long-term catheter when compared to patients 60-75 years of age.
    - Those whose race is reported as "Other" were less likely to have a long-term catheter when compared to whites, as were Hispanics, when compared to non-Hispanics.
    - Individuals whose duration of ESRD < 1 year and whose duration of ESRD are 9+ year were almost four times and 26% more likely to have a long-term catheter, respectively.
    - Patients whose duration of ESRD are 5-<9 years were 8% less likely to have a long-term catheter when compared to patients whose duration of ESRD are 1-<5 years.

- Below are the odds ratios for these patient characteristics (having a catheter for at least three months):

Age	Odds Ratio (95% CI)	P-value
18-<25	1.46 (1.12, 1.90)	0.005
25-<59	1.06 (1.00, 1.121	0.057
60-<75	used as the reference group	
75+	1.14 (1.07, 1.23)	<.0001

Sex	Odds Ratio (95% CI)	P-value
Female	1.55 (1.47, 1.63)	<.0001
Male	used as the reference group	

Race	Odds Ratio (95% CI)	P-value
Black	0.98 (0.91, 1.05)	0.586
Other race	0.77 (0.67, 0.88)	<.0001
White	used as the reference group	

Ethnicity	Odds Ratio (95% CI)	P-value
Hispanic	0.81 (0.74, 0.89)	<.0001
Non-Hispanic	used as the reference group	

Duration of ESRD	Odds Ratio	P-value
<1 year	3.97 (3.78, 4.18)	<.0001
1-<5 years	used as the reference group	
5-<9 years	0.92 (0.84, 1.00)	0.041
9+ years	1.26 (1.15, 1.38)	<.0001

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This is a Process measure, not an intermediate outcome measure.

The major evidence issue is that the evidence is virtually all retrospective observational data. For the whole population the evidence is clear that catheyter use is associated with worse outcomes than AVF use. However, for subpopulations there is no clear evidence. For a facility-level measure, this point may be inportant.

There has been no substantial change in evidence (to my knowledge) since the last review

Preliminary rating: Moderate

**There is an association between type of vascular access used for hemodialysis and patient mortality; this measure focuses on the process of assessing long term catheter use at chronic dialysis facilities. Evidence suggests that long term catheter use is correlated with the highest mortality risk and arteriovenous fistula use has the lowest risk of mortality. Arteriovenous grafts (AVG) have been found to have a risk of death that is higher than AVF but lower than catheters. This measure is intended to be jointly reported with the Hemodialysis Vascular Access- Standardized Fistula Rate. Used together, the two vascular access quality measures consider Arterial Venous (AV) fistula use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome.

**Yes, catheters have been linked to mortality and hospitalization

**Medical evidence is strong that catheters are associated with higher mortality

**Review of evidence provided with data showing increased mortality rates with long-term catheter use as well as KDOQI guidelines

**Evidence is based on systematic review

**Intermediate outcome, evidence relates directly to measure as specified;

The intermediate outcome is related to desired outcomes (lowering hospitalization and mortality)

**strong evidence showing association of catheters with worse patient outcomes

**Good evidence

**intermediate outcome measure; based on association between type of vascular access used for hemodialysis and patient mortality; It focuses on the process of assessing long term catheter use at chronic units.

Evidence supports long term catheter use and higher mortality risk while AVF use has lowest risk. AVG use has higher risk of death than AVF bu lower than catheters.

Evidence is supported.

1b. Performance Gap

Comments: **Interquartile differences remain substantial.

Differences in catheter use by vintage or age may be appropriate rather than disparities.

Preliminary rating for opportunity for improvement: High

**2014 data indicate age, sex, race, ethnicity and dialysis vintage were evaluated in a logistic regression model for long-term catheter use. The analysis results indicate potential disparity in prolonged use of a tunneled catheter.

**Yes, Based on the data submitted there is a performance gap Thought the maximum rate of 58.2% seems very high and may be a function of small numbers.

The analysis did not identify a disparities gap, but was done at a population level. Given that the measure is a facility level measure, a facility level analysis should be done to determine if a disparities gap exists for those same factors

**They show a wide range of measure results and also striking SDS results

**CROWNWeb data from 2014 showing 11% population with catheter with fair IQ range

Disparities data provided as to subgroups with potential for even greater performance improvement

**There is still a modest performance gap, but it does not huge, especially considering this number is unlikely to reach 0. There's a significant gender disparity.

**Variability is demonstrated. In 2014, facility percentage of patient-months with long-term catheter: median 10.5% [IQR 7, 14.9%] with a min of 0% and max of 58.2%

Data on measure by population subgroups provided - differences by age, sex, race, ethnicity and vintage shown

**Gap well demonstrated

**No concerns

**Potential disparity in prolonged use f a tunneled catheter.

Numerous variable associated with disparities provided.

1c. High Priority (previously referred to as High Impact)

Comments: None

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

Specifications:

- The numerator of the measure is: *The number of adult patient-months in the denominator who were on maintenance hemodialysis using a catheter continuously for three months or longer as of the last hemodialysis session on the reporting month.*
- The denominator of the measure is: *All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) for the complete reporting month at the same facility.*
- Both ICD-9 and ICD-10 codes are included in the Data Dictionary Code Table.
- The measure logic is included in the appendix and seems straightforward.
- This intermediate clinical outcome measure is not risk adjusted.
- A lower rate is considered to be better quality.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- To determine the measures reliability, developers calculated the inter-unit reliability (IUR) for the annual performance scores. The IUR measures the proportion of the total variation of a measure attributable to the between-facility variation. A small IUR reveals that most of the variation of the measure 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 indicates that most of the variation between facilities is due to the real difference between facilities.
- The reliability calculation only included facilities with at least 11 patients during the entire year.

Results of reliability testing

- The IUR is 0.765, which indicates that 76.5% of the variation in the annual long-term catheter rate can be attributed to between-facility differences in performance (signal) and about 23.5% to the within-facility variation (noise).
- The developer states that the result of IUR testing suggests a high degree of reliability.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm
Precise specifications (Box 1) → empirical testing of the measure score (Boxes 2 and 4) →appropriate method (Box 5) → high certainty

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:
- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:
- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:
- The Developer assessed validity using a Poisson regression model to measure the association between facility level quintiles of performance scores and the 2014 Standardized Mortality Ratio (SMR) for Dialysis Facilities (NQF 0369) as well as the 2014 Standardized Hospitalization Ratio (SHR)for Dialysis Facilities (NQF 1463). Performance scores at the facility level were divided into quintiles (Q1-Q5) and the relative risk of mortality was calculated for each quintile (Q1 and Q2 as the reference group).

Validity testing results:
- Quintiles of the performance scores were defined as follows (*Q1 and Q2 as Reference):
  - Q1*: 0.0%-<6.24%
  - Q2*: 6.24%-<9.12%
  - Q3: 9.12-<12.00%
  - Q4: 12.00%-<16.21%
  - Q5: 16.21%-<58.16%
- The developer provided the following as an analysis of the results:
  - The percent of patient-months with a long-term catheter was significantly associated with the risks of mortality and hospitalization.
  - For the 2014 SMR, the relative risk of mortality increased as the performance measure quintile increased from the reference group.

Quintile	Relative Reliability (RR)	95% CI	P-value
Q1	Used as reference group		
Q2	Used as reference group		
Q3	1.03	1.01, 1.05	0.006
Q4	1.03	1.01, 1.05	0.008
Q5	1.09	1.07, 1.12	<0.001

- For the 2014 Standardized Hospital Ratio, the relative risk of hospitalization increased as the performance measure quintile increased from the reference group. These results suggest the predictive relationship of higher catheter use with higher mortality and hospitalization.

Quintile	Relative Reliability (RR)	95% CI	P-value
Q1	Used as reference group		
Q2	Used as reference group		
Q3	1.08	1.08, 1.08	<0.001
Q4	1.10	1.10, 1.10	<0.001
Q5	1.16	1.15, 1.16	<0.001

Questions for the Committee:

- *What is the expected correlation between this measure and SMR and SHR?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The following exclusions are applicable to the denominator:
  - Patients with a catheter that have limited life expectancy, which is defined as:
    - Patients under hospice care in the current reporting month
    - Patients with metastatic cancer in the past 12 months
    - Patients with end stage liver disease in the past 12 months
    - Patients with coma or anoxic brain injury in the past 12 months
- The developer calculated and compared the facility-level mean percentage of patient-months for at least three months, with and without the patient-month exclusions.
- The developer states that the exclusion criteria are necessary because the percentage of patients excluded at each facility are not evenly distributed across all facilities. The exclusions take into account that some facilities treat a higher proportion of patients with limited life expectancy.
- The frequency of exclusions in 2014:
  - Patient years: 0.92% excluded
  - Patients: 2.35% excluded
- The mean measure results went from 0.121 before exclusions to 0.118 with exclusions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

o *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Risk adjustment summary : Measure is not risk-adjusted. Developer provided the following:

- Risk adjustment is not appropriate for this measure because of the primary goal of disincentivizing catheter use for incident and particularly prevalent dialysis patients. This measure was reviewed by the 2015 vascular access TEP which also did not recommend risk adjustment.

Questions for the Committee:

o *Do you agree with the developer that no risk adjustment is indicated?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- For both the proportion of patient-months with catheter greater than or equal to three months and the overall national distribution, a majority of the facilities had 87.9% facilities have "as expected" performance, and 12% facilities have performed "worse than expected."
- Lower rates of catheter use for three months or more represent better quality of care. This analysis demonstrates both practical and statistically significant differences in performance across facilities based on their proportion of patient months with a catheter for three months or greater.

Question for the Committee:

o *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

None needed.

2b7. Missing Data

An analysis of missing data was not provided for this measure.

Guidance from the algorithm:
Specifications consistent with the evidence (Box 1) → threats to validity addressed (Box 2) → empirical validity testing of the measure score (Box 6) → appropriate method (Box 7) → moderate or high certainty the measure is a valid indicator of quality

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: None

2a2. Reliability Testing

Comments:

**IUR testing suggests a high degree of reliability - - I agree

Test sample is adequate to generalize

Test results show sufficient reliability

Preliminary rating: High

**To determine the measures reliability, developers calculated the inter-unit reliability (IUR) for the annual performance scores. The IUR measures the proportion of the total variation of a measure attributable to the between-facility variation. A small IUR reveals that most of the variation of the measure 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 indicates that most of the variation between facilities is due

to the real difference between facilities. The reliability calculation only included facilities with at least 11 patients during the entire year.
The IUR is 0.765, which indicates that 76.5% of the variation in the annual long-term catheter rate can be attributed to between-facility differences in performance (signal) and about 23.5% to the within-facility variation (noise).
This suggests a high degree of reliability.
**Yes. No issues with Reliability testing
**IUR is around 75%
**IUR of 0.765 -- testing done with data source and with level of analysis indicated
**The reliability testing appears to be high
**IUR 0.765
**reliable
**No concerns
**Only facilities with at least 11 patients during the year were used.

2b2. Validity Testing
Comments:
**The correlation between SMR and catheter use, and SHR and catheter use both are significant.
SMR and SHR likely correlate with each other - - so these observations could be confounded by that likely association.
Test sample is adequate
Results show substantial validity
**The Developer assessed validity using a Poisson regression model to measure the association between facility level quintiles of performance scores and the 2014 Standardized Mortality Ratio (SMR) for Dialysis Facilities (NQF 0369) as well as the 2014 Standardized Hospitalization Ratio (SHR)for Dialysis Facilities (NQF 1463). Performance scores at the facility level were divided into quintiles (Q1-Q5) and the relative risk of mortality was calculated for each quintile (Q1 and Q2 as the reference group).
The percent of patient-months with a long-term catheter was significantly associated with the risks of mortality and hospitalization. The relative risk of hospitalization increased as the performance measure quintile increased from the reference group. These results suggest the predictive relationship of higher catheter use with higher mortality and hospitalization.
**No issues with Validity testing. Though there will be an intractable catheter rate defined as a floor. This represent patient refusal and new patients. We estimate that to be ~8% for the population. the developer should take that into account
**Validity testing: correlated to SMR amd SHR - looks OK
**Poisoon regression model measuring facility quintile performance on this measure vs SMR and SHR; higher catheter rates linked with worse SMR and SHR in statistically significant fashion
**Validity testing appears to have been performed well.
**To assess validity, association between performance score and SHR or SMR examined. Higher catheter rates were associated with higher SHR and SMR.
**demonstrated validity
**No concerns
**Appropriate. Percent of patient months with a long-term catheter was associated with risks of morality and hospitalization.
High degree of reliability.
2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:

**Exclusions to the denominator have good face value - - - question remains whether there are other subgroups (like frail elderly) that should be considered for exclusion.
Exclusions are consistent with the evidence - - and are needed given the uneven distribution across providers
Measure doies idientify meaningful differences in quality of care

Preliminary rating: Moderate
**The following exclusions are applicable to the denominator: Patients with a catheter that have limited life expectancy, which is defined as: Patients under hospice care in the current reporting month; Patients with metastatic cancer in the past 12 months; Patients with end stage liver disease in the past 12 months; Patients with coma or anoxic brain injury in the past 12 months . The exclusions take into account that some facilities treat a higher proportion of patients with limited life expectancy.
Risk adjustment is not appropriate for this measure because of the primary goal of disincentivizing catheter use for incident and particularly prevalent dialysis patients.
Meaningful difference was shown.
**There should be a mechanism to allow for patient refusal and for short term catheter use for patients with an imminent transplant
**No risk adjustments - this seems appropriate as nearly all long-term catheters should be be replaced with a different vascular access.
**My concern is that the exclusion criteria do not account for inability to provide an alternate access for example due to anatomic restrictions.
**Codes selected to represent exclusions may need further refinement;
What is the impact of only applying exclusions to the Medicare beneficiaries with available claims?
No risk adjustment (as compared with AVF measure)
Counting missing access type in the numerator may be inappropriate - unclear impact
**exclusions improve validity of this measure because patients not appropriate for interventions are excluded
**Limited life expectancy exclusion needs clear definition.
**2b3. Appropriate exclusions.
2b4. Risk adjustment is not appropriate for this measure; TEP did not recommend risk adjustment. Agree that none is required.
2b5. 87.9% facilities had expected performance with 12% having worse than expected.
Lower rates of catheter use for 3 months or more represent better quality of care. Overall the analysis demonstrates both practical and statistically significant differences in performance across facilities.
2b6. None needed.
2b7. no data presented.

Criterion 3. Feasibility

3. Feasibility is the 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.

- All data elements are in defined fields in a combination of electronic sources and are collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: None

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details: CMS will determine if and when the measure will be implemented in a CMS program. Upon endorsement, CMS will consider retiring the currently endorsed measure of catheter use (#0256) in favor of this new measure for implementation in a future performance year for the ESRD QIP and reporting period for Dialysis Facility Compare at the next available opportunity.

Improvement results: The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of this catheter measure. Once implemented facility performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing prolonged catheter use, while accounting for patients where a long-term catheter may be an appropriate vascular access choice.

Unexpected findings (positive or negative) during implementation

Potential harms: Potential unintended consequences would center on patients being pushed towards having an AVF or AVG created when they may not realize a benefit from this type of access due to either comorbidities or a limited life expectancy. It is incumbent on both dialysis facilities, as well as surgeons, to establish the most appropriate type of vascular access for patients based on their individual circumstances and preferences, rather than in response to quality measures.

Feedback :
No feedback from other sources available.

***Questions for the Committee**:*
- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments:
**Usability appears excellent.
I do remain a bit concerned that there are subsets of the population for whom a catheter is a good choice - - not identified in this measure - - that will be pushed toward another access.
Preliminary rating: Moderate
**CMS will determine if and when the measure will be implemented in a CMS program. The measure is not yet implemented in a

public reporting program, so improvement could not be evaluated.
**No concerns
**Credible plan for accountability
**measure sponsor states intention to consider use as part of QIP or DFC -- not currently used in public reporting
**This is not currently in use or publically reported
**Planned for public reporting and payment, may replace current measures
**not publicly reported
when used with new fistula measure will allow retirement of old access PM's
**Proposed to use with measure 2977
**Not reported publicly; Facility performance can be evaluated to determine is the measure affects quality improvement in reducing long term catheter use.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0251: Vascular Access—Functional Arteriovenous Fistula (AVF) or AV Graft or Evaluation for Placement
- 2594: Optimal End Stage Renal Disease (ESRD) Starts

Harmonization

- Measure 0251 contains several components including AV fistula use, AV graft use or referral to a vascular surgeon (or other qualified physician) if using a long-term catheter. It is a referral process measure for those patients with a catheter. This has the potential for facilities to score well on the measure even if they have patients with a catheter, as long as the patient was referred to or evaluated by a vascular surgeon. We acknowledge this is an important step to fistula placement however it departs from the intent of the catheter measure to function as a more direct disincentive to prolonged catheter use, consistent with the concerns and recommendations made by the vascular access TEP.
- Measure 2594 is not directed toward dialysis facilities. The setting focus addresses a different provider type which falls outside the purview of measures evaluating dialysis facility performance on prolonged catheter use. These suggest fundamental differences in measure target populations, setting and intent that cannot be harmonized. Additionally, the measure is limited to incident patients, while the long-term catheter rate measure includes both incident and prevalent patients as the measured population.

Pre-meeting public and member comments

Comment by Joseph Vassalotti
Organization National Kidney Foundation
Comment #5700: National Kidney Foundation (NKF) strongly supports this measure and its pairing with the standardized fistula rate measure (NQF #2977). NKF is particularly pleased with the four additional exclusions that acknowledge catheter use is appropriate for patients with limited life expectancy. These changes align with NKF's previous recommendations.

We do note that clarity around catheter use continuously would strengthen this measure. Specifically, the numerator should include all patients with a catheter in place for the reporting period, whether the hemodialysis catheter is in continuous use or not. The presence of a catheter increases the risk for infection even if it is not in use.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners

Comment #5688: As with the AVF measure, KCP used the existing catheter measure, NQF 0256, for context in our review.

SPECIFICATIONS. As with the AVF measure, KCP notes the 90-day ESRD requirement has been removed from the denominator statement as compared to #0256, which means the "clock" for the measure starts on the first day of dialysis in a non-hospital setting—but that the permitted timeframe for catheter use in the numerator is still 90 days; we support this change. Additionally, we commend the developer for adding an exclusion for patients with limited life expectancy and for now unambiguously identifying the four subcategories, both approaches that KCP had recommended.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:**
Measure Title: Hemodialysis Vascular Access: Long-term Catheter Rate
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: 4/15/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
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
☐ Health outcome: Click here to name the health outcome
☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☒ Intermediate clinical outcome (*e.g., lab value*): catheter rate

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Several observational studies have demonstrated an association between type of vascular access used for hemodialysis and patient mortality. Long term catheter use is associated with the highest mortality risk while arteriovenous fistula use has the lowest mortality risk. Arteriovenous grafts (AVG) have been found to have a risk of death that is higher than AVF but lower than catheters.

The measure focus is the process of assessing long term catheter use at chronic dialysis facilities.

This process leads to improvement in mortality as follows:
Measure long term catheter rate→ Assess value →Identify patients who do not have an AV Fistula or AV graft→Evaluation for an AV fistula or graft by a qualified dialysis vascular access provider →Increase Fistula/Graft Rate → Lower catheter rate →Lower patient mortality.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

National Kidney Foundation KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates: Hemodialysis Adequacy, Peritoneal Dialysis Adequacy and Vascular Access. Am J Kidney Dis 48:S1-S322, 2006 (suppl 1).

http://www.kidney.org/professionals/KDOQI/guidelines_commentaries

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

GUIDELINE 2. SELECTION AND PLACEMENT OF HEMODIALYSIS ACCESS

A structured approach to the type and location of long-term HD accesses should help optimize access survival and minimize complications. Options for fistula placement should be considered first, followed by prosthetic grafts if fistula placement is not possible. Catheters should be avoided for HD and used only when other options listed are not available.

2.1 The order of preference for placement of fistulae in patients with kidney failure who choose HD as their initial mode of KRT should be (in descending order of preference):

2.1.1 Preferred: Fistulae. (B)

2.1.2 Acceptable: AVG of synthetic or biological material. (B)

2.1.3 Avoid if possible: Long-term catheters. (B)

2.1.4 Patients should be considered for construction of a primary fistula after failure of every dialysis AV access. (B)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

KDOQI Guideline 2.1 was graded B, indicating moderate evidence supports the guideline.
The "B" rating indicates: It is recommended that clinicians routinely follow the guideline for eligible patients. There is moderately strong evidence that the practice improves health outcomes.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

The rating system defined in the KDOQI Guidelines was used to grade the strength of the Guideline recommendation. KDOQI defined grades as follows:

Grade A: It is strongly recommended that clinicians routinely follow the guideline for eligible patients. There is strong evidence that the practice improves health outcomes.

Grade B: It is recommended that clinicians routinely follow the guideline for eligible patients. There is moderately strong evidence that the practice improves health outcomes.

Grade CPR: It is recommended that clinicians consider following the guideline for eligible patients. This recommendation is based on either weak evidence or on the opinions of the Work Group and reviewers that the practice might improve health outcomes.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

National Kidney Foundation. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates: Hemodialysis Adequacy, Peritoneal Dialysis Adequacy and Vascular Access. Am J Kidney Dis 48:S1-S322, 2006 (suppl 1).

http://www.kidney.org/professionals/KDOQI/guidelines_commentaries

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The evidence review focuses on the advantages of AV fistula compared to other types of vascular access, specifically over catheters as the means of vascular access. The review highlights the superior patency, reduced need for interventions, and lower infection rates associated with AV fistula.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

The quality of evidence was not explicitly graded in the KDOQI guidelines. However, it was implicitly assessed according to the criteria outlined in the table in 1a.7.3 below. The workgroup considered the overall methodological quality, the target population (e.g. patients on dialysis), and whether the health outcome was studied directly or not.

Overall, the evidence that supports the guideline was assessed as: Moderately Strong.

The workgroup defined "Moderately Strong" as: Evidence is sufficient to determine effects on health outcomes in the target population, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; OR evidence is from studies with some problems in design and/or analysis; OR evidence is from well-designed, well-conducted studies on surrogate endpoints for efficacy and/or safety in the target population.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

		Methadologic Quality		
Outcome	Population	Well designed and analyzed (little if any potential bias)	Some problems in design and/or analysis (some potential bias)	Poorly designed and/or analyzed (large potential bias)
Health Outcomes	Target Population	Strong	Moderately Strong	Weak
Health Outcomes	Other than target population	Moderately Strong	Moderately Strong	Weak
Surrogate Measure	Target Population	Moderately Strong	Weak	Weak
Surrogate Measure	Other than target population	Weak	Weak	Weak

Strong- Evidence includes results from well-designed, well-conducted study/studies in the target population that directly assess effects on health outcomes.

Moderately strong- Evidence is sufficient to determine effects on health outcomes in the target population, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; OR evidence is from a population other than the target population, but from well-designed, well conducted studies; OR evidence is from studies with some problems in design and/or analysis; OR evidence is from well-designed, well-conducted studies on surrogate endpoints for efficacy and/or safety in the target population.

Weak- Evidence is insufficient to assess the effects on net health outcomes because it is from studies with some problems in design and/or analysis on surrogate endpoints for efficacy and/or safety in the target population; OR the

evidence is only for surrogate measures in a population other than the target population; OR the evidence is from studies that are poorly designed and/or analyzed.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: January 1997 – June 2005

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The 2006 Clinical Practice Guidelines for Vascular Access is an update to the original vascular access guidelines published in 1997 by the National Kidney Foundation. In the eight years that the literature review included for the update, there have been no randomized controlled trials for type of vascular access. Specifically, for the guideline used to support this measure, a total of 84 peer-reviewed publications are included in the body of evidence presented. While these are all observational studies, some are based on either national data such as the United States Renal Data System (USRDS) that includes all patients with end stage kidney disease in the US, or international data, such as the Dialysis Outcomes Practice Pattern Study (DOPPS) that provides a global perspective for US vascular access outcomes.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The overall quality of evidence is moderately strong. All studies are in the target population of hemodialysis patients. Some studies have evaluated health outcomes such as patient mortality, but have limitations due to the observational nature of the design. Other studies have more rigorous design, but use surrogate outcomes such as access thrombosis.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The 12 studies listed below highlight the core benefits associated with using an AV fistula or graft such as reduced mortality and morbidity relative to using a tunneled catheter. Specifically, AV fistula have:

- Lowest Cost[1-3]: Compared to catheters, Medicare expenditures for AVF are approximately $17,000 less per person per year.
- Lowest rates of infection: AV fistula have the lowest rates of infection followed by AV grafts and then tunneled dialysis catheters[4]. Vascular access infections are common, and represent the second most common cause of death for patients receiving hemodialysis.[5]
- Lowest mortality and hospitalization: Patients using catheters (RR=2.3) and grafts (RR=1.47) have a greater mortality risk than patients dialyzed with fistulae[6-9]. Other studies have also found that use of fistulae reduces mortality and morbidity[10-12] compared to AV grafts or catheters.

References:

1. Mehta S: Statistical summary of clinical results of vascular access procedures for haemodialysis, in Sommer BG, Henry ML (eds): Vascular Access for Hemodialysis-II (ed 2). Chicago, IL, Gore, 1991, pp 145-157
2. The Cost Effectiveness of Alternative Types of Vascular access and the Economic Cost of ESRD. Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 1995, pp 139-157
3. Eggers P, Milam R: Trends in vascular access procedures and expenditures in Medicare's ESRD program, in Henry ML (ed): Vascular Access for Hemodialysis-VII. Chicago, IL, Gore, 2001, pp 133-143
4. Nassar GM, Ayus JC: Infectious complications of the hemodialysis access. Kidney Int 60:1-13, 2001
5. Gulati S, Sahu KM, Avula S, Sharma RK, Ayyagiri A, Pandey CM: Role of vascular access as a risk factor for infections in hemodialysis. Ren Fail 25:967-973, 2003
6. Dhingra RK, Young EW, Hulbert-Shearon TE, Leavey SF, Port FK: Type of vascular access and mortality in U.S. hemodialysis patients. Kidney Int 60:1443-1451, 2001
7. Woods JD, Port FK: The impact of vascular access for haemodialysis on patient morbidity and mortality. Nephrol Dial Transplant 12:657-659, 1997
8. Xue JL, Dahl D, Ebben JP, Collins AJ: The association of initial hemodialysis access type with mortality outcomes in elderly Medicare ESRD patients. Am J Kidney Dis 42:1013-1019, 2003
9. Polkinghorne KR, McDonald SP, Atkins RC, Kerr PG: Vascular access and all-cause mortality: A propensity score analysis. J Am Soc Nephrol 15:477-486, 2004
10. Huber TS, Carter JW, Carter RL, Seeger JM: Patency of autogenous and polytetrafluoroethylene upper extremity arteriovenous hemodialysis accesses: A systematic review. J Vasc Surg 38(5):1005-11, 2003
11. Perera GB, Mueller MP, Kubaska SM, Wilson SE, Lawrence PF, Fujitani RM: Superiority of autogenous arteriovenous hemodialysis access: Maintenance of function with fewer secondary interventions. Ann Vasc Surg 18:66-73, 2004
12. Pisoni RL, Young EW, Dykstra DM, et al: Vascular access use in Europe and the United States: Results from the DOPPS. Kidney Int 61:305-316, 2002

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Unintended consequences of catheter avoidance strategies were not well studied at the time when the clinical practice guidelines were developed. More recently, members of the dialysis community have voiced concern that an aggressive agenda to create AVF in most all patients would lead to unnecessary surgery for some patients that have a high risk of mortality either before starting dialysis or within the first year of treatment. Despite these concerns, the overall risk associated with AV fistula creation to avoid long term catheter use are considered to be small and overshadowed by the long-term benefits outlined above for fistula use.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Casey JR, Hanson CS, Winkelmayer WC, et al. **Patients' perspectives on hemodialysis vascular access: a systematic review of qualitative studies.** *Am J Kidney Dis. 2014 Dec;64(6):937-53. doi: 10.1053/j.ajkd.2014.06.024. Epub 2014 Aug 10*.

This systematic review and thematic synthesis of qualitative studies describes patients' perspectives on vascular access initiation and maintenance in hemodialysis. 46 studies were reviewed and found that initiation of vascular access signifies kidney failure and imminent dialysis, which is emotionally confronting. Patients strive to preserve their vascular access for survival, but at the same time describe it as an agonizing reminder of their body's failings and "abnormality" of being amalgamated with a machine disrupting their identity and lifestyle. Timely education and counseling about

vascular access and building patients' trust in health care providers may improve the quality of dialysis and lead to better outcomes for patients with chronic kidney disease requiring hemodialysis.

Impact: Adds the patient's perspective to the discussion on vascular access options.

Al-Jaishi AA, Oliver MJ, Thomas SM, et al. **Patency rates of the arteriovenous fistula for hemodialysis: a systematic review and meta-analysis.** *Am J Kidney Dis. 2014 Mar;63(3):464-78. doi: 10.1053/j.ajkd.2013.08.023. Epub 2013 Oct 30. Review.*

This systematic review and meta-analysis reported that in recent years AVFs had a high rate of primary failure and low to moderate primary and secondary patency rates. Consideration of these outcomes is required when choosing a patient's preferred access type.

Impact: Updates primary and secondary patency rates of AVF for more contemporary cohorts of dialysis patients. The lower success rates suggests that some patients may not realize the full benefits of AVF that have been previously reported in the KDOQI systematic review.

Oliver MJ, Quinn RR. **Recalibrating vascular access for elderly patients.** *Clin J Am Soc Nephrol. 2014 Apr;9(4):645-7. doi: 10.2215/CJN.01560214. Epub 2014 Mar 20.*

Governments in numerous jurisdictions have set targets for fistula utilization and some have tied reimbursement to attaining these targets. This creates an environment in which it is tempting to overemphasize the benefits of fistulas and the risks of catheters when discussing vascular access options with patients.

Impact: Highlights that not all older patients may benefit from an AVF.

Drew DA, Lok CE, Cohen JT, et al. **Vascular access choice in incident hemodialysis patients: a decision analysis.** *J Am Soc Nephrol. 2015 Jan;26(1):183-91. doi: 10.1681/ASN.2013111236. Epub 2014 Jul 25.*

Decision analysis evaluating AV fistula, AV graft, and central venous catheter (CVC) strategies for patients initiating hemodialysis with a CVC, a scenario occurring in over 70% of United States dialysis patients. An AV fistula attempt strategy was found to be superior to AV grafts and CVCs in regard to mortality and cost for the majority of patient characteristic combinations, especially younger men without diabetes. Women with diabetes and elderly men with diabetes had similar outcomes, regardless of access type. Overall, the advantages of an AV fistula attempt strategy lessened considerably among older patients, particularly women with diabetes, reflecting the effect of lower AV fistula success rates and lower life expectancy. These results suggest that vascular access-related outcomes may be optimized by considering individual patient characteristics.

Impact: Certain patient groups, such as women with diabetes, have lower reported success rates of AVF creation and may have equivalent outcomes with an AVG.

Wish JB. **Catheter last, fistula not-so-first.** *J Am Soc Nephrol. 2015 Jan;26(1):5-7. doi: 10.1681/ASN.2014060594. Epub 2014 Jul 25.*

The issue of vascular access choice is not as black and white as the Centers for Medicare & Medicaid Services (CMS) would like it to appear, with arteriovenous fistula (AVF) always being good or "first" and central venous catheters (CVCs) always being bad or "last." Nonetheless,CMS has instituted a quality incentive program (QIP) for dialysis providers that rewards high AVF prevalence and penalizes high CVC prevalence without regard to patient mix. For payment year 2014, vascular access constitutes 30% of the total QIP score. This may have already led to access to care issues, as some dialysis providers are refusing to accept patients with CVCs. CMS has recently given ground on this issue by renaming the "Fistula First" initiative "Fistula First Catheter Last" (FFLC) to emphasize that CVC avoidance is as important or more important than AVF use.

Impact: Opinion piece on changes in the Fistula First initiative reflecting the implementation of the current NQF endorsed fistula and catheter vascular access measures in the CMS Quality Incentive Program (QIP). The empahsis of the opinion piece suggests a greater shift to catheter avoidance versus only prioritizing promotion of fistula use.

Grubbs V, Wasse H, Vittinghoff E, et al. **Health status as a potential mediator of the association between hemodialysis vascular access and mortality.** *Nephrol Dial Transplant. 2014 Apr;29(4):892-8. doi: 10.1093/ndt/gft438. Epub 2013 Nov 13.*

Selection of healthier patients for arteriovenous fistula (AVF) placement may explain higher observed catheter-associated mortality among elderly hemodialysis patients. A proportional hazard model was used to examine 117,277 incident hemodialysis patients aged 67-90 years from USRDS for the association of initial vascular access type and 5-year mortality after accounting for health status. Patients with catheter alone had more limited functional status (25.5 versus 10.8% of those with AVF) and 3-fold more prior hospital days than those with AVF (mean 18.0 versus 5.4). In a fully adjusted model including health status, mortality differences between access type were attenuated, but remained statistically significant <AVG [HR 1.18 (1.13-1.22)], catheter plus AVF [HR 1.20 (1.17-1.23)], catheter plus AVG {HR 1.38 [1.26 (1.21-1.31)]} and catheter only [HR 1.54 (1.50-1.58)], P < 0.001>.The observed attenuation in mortality differences previously attributed to access type alone suggests the existence of selection bias. Nevertheless, the persistence of an apparent survival advantage after adjustment for health status suggests that AVF should still be the access of choice for elderly individuals beginning hemodialysis until more definitive data eliminating selection bias become available.

Impact: Underscores the need to adjust for patient characteristics and comorbidities when evaluating the association between vascular access type and outcomes such as mortality.

Lok, Charmaine E & Foley, Robert. **Vascular access morbidity and mortality: trends of the last decade.** *Clin J Am Soc Nephrol. 2013 Jul;8(7):1213-9. doi: 10.2215/CJN.01690213.*

During the past decade, clear trends in the types of incident and prevalent hemodialysis vascular access can be observed. There has been a steady increase and recent stabilizaton of patients initiating hemodialysis with a central venous catheter, representing approximately 80% of all incident accesses. There has also been a steady increase in prevalent fistula use, currently greater than 50% within 4 months of hemodialysis initiation. Patient and vascular access related morbidity and mortality are reflected in the type of vascular access used at initiation and for long-term maintenance dialysis. There is a three- to fourfold increase in risk of infectious complications in patients initiating dialysis with a catheter compared with either a fistula or graft and a sevenfold higher risk when the catheter is used as a prevalent access. Procedure rates have increased two- to threefold for all types of access. There is a significant increased risk of mortality associated with catheter use, especially within the first year of dialysis initiation.

Impact: Despite longstanding KDOQI guidelines, many patients still start hemodialysis with a tunneled catheter and experience higher rates of infectious complications compared to those with an AVF.

Ravani, Pietro & Palmer, Suetonia C & Oliver, Matthew J et al. **Associations between hemodialysis access type and clinical outcomes: a systematic review.** *J Am Soc Nephrol. 2013 Feb;24(3):465-73. doi: 10.1681/ASN.2012070643. Epub 2013 Feb 21.*

Clinical practice guidelines recommend an arteriovenous fistula as the preferred vascular access for hemodialysis, but quantitative associations between vascular access type and various clinical outcomes remain controversial. This systematic review of cohort studies evaluates the associations between type of vascular access (arteriovenous fistula, arteriovenous graft, and central venous catheter) and risk for death, infection, and major cardiovascular events. 67 (62 cohort studies comprising 586,337 participants)studies were selected. In a random effects meta-analysis, compared with persons with fistulas, those individuals using catheters had higher risks for all-cause mortality (risk ratio=1.53, 95% CI=1.41-1.67), fatal infections (2.12, 1.79-2.52), and cardiovascular events (1.38, 1.24-1.54). Similarly, compared with persons with grafts, those individuals using catheters had higher risks for mortality (1.38, 1.25-1.52), fatal infections (1.49, 1.15-1.93), and cardiovascular events (1.26, 1.11-1.43). Compared with persons with fistulas, those individuals with grafts had increased all-cause mortality (1.18, 1.09-1.27) and fatal infection (1.36, 1.17-1.58), but we did not detect a difference in the risk for cardiovascular events (1.07, 0.95-1.21). The risk for bias, especially selection bias, was high. In conclusion, persons using catheters for hemodialysis seem to have the highest risks for death, infections, and cardiovascular events compared with other vascular access types, and patients with usable fistulas have the lowest risk.

Impact: This study emphasizes that the body of evidence is consistent in the magnitude and direction of effect with regards to the benefits of AVF over central venous catheter.

Moist, Louise M & Lok, Charmaine E & Vachharajani, Tushar J et al. **Optimal hemodialysis vascular access in the elderly patient.** *Semin Dial. 2012 Nov-Dec;25(6):640-8. doi: 10.1111/sdi.12037.*

The optimal vascular access for elderly patients remains a challenge due to the difficulty balancing the benefits and risks in a population with increased comorbidity and decreased survival. Age is commonly associated with failure to mature in fistula and decreased rates of primary and secondary patency in both fistula and grafts. In the elderly, at 1 and 2 years, primary patency rates range from 43% to 74% and from 29% to 67%, respectively. Secondary patency rates at 1 and 2 years range from 56% to 82% and 44% to 67%, respectively. Cumulative fistula survival is no better than grafts survival when primary failures are included. Several observational studies consistently demonstrate a lower adjusted mortality among those using a fistula compared with a catheter; however, catheter use in the elderly is increasing in most countries with the exception of Japan. Both guidelines and quality initiatives do not acknowledge the trade-offs involved in managing the elderly patients with multiple chronic conditions and limited life expectancy or the value that patients place on achieving these outcomes. The framework for choice of vascular access presented in this article considers: (1) likelihood of disease progression before death, (2) patient life expectancy, (3) risks and benefits by vascular access type, and (4) patient preference. Future studies evaluating the timing and type of vascular access with careful assessments of complications, functionality, cost benefit, and patients' preference will provide relevant information to individualize and optimize care to improve morbidity, mortality, and quality of life in the elderly patient.

Impact: Outlines the importance of considering patient factors in vascular access options for elderly patients.

Schmidt, Rebecca J & Goldman, Richard S & Germain, Michael. **Pursuing permanent hemodialysis vascular access in patients with a poor prognosis: juxtaposing potential benefit and harm.** *Am J Kidney Dis. 2012 Dec;60(6):1023-31. doi: 10.1053/j.ajkd.2012.07.020. Epub 2012 Sep 19.*

For patients with end-stage renal disease requiring hemodialysis, the native arteriovenous fistula remains the gold standard of vascular access, with tunneled cuffed central venous catheters reserved for temporary use or as a last resort in patients for whom a permanent vascular access is not possible. It is expected that most patients receiving hemodialysis will be suitable for arteriovenous fistula placement, with suitable patients defined as those: (1) for whom long-term dialysis is expected to confer benefit, (2) with vascular anatomy amenable to arteriovenous fistula placement, and (3) with progressive irreversible kidney failure who are more likely to require dialysis than to die before reaching dialysis dependence. The present article reviews considerations for vascular access decision making, focusing on older patients and those with a poor prognosis, weighing the risks and benefits of arteriovenous fistulas, arteriovenous grafts, and central venous catheters and emphasizing that in the process of vascular access decision making for such patients, medical and ethical obligations to avoid central venous catheters must be balanced by the obligation to do no harm.

Impact: Risks and benefits of arteriovenous fistulas, relative to arteriovenous grafts, and central venous catheters need to be considered, particularly carefully in older patients and those with poor prognosis (limited life expectancy).

Vassalotti, Joseph A & Jennings, William C & Beathard, Gerald A et al. **Fistula first breakthrough initiative: targeting catheter last in fistula first.** *Semin Dial. 2012 May;25(3):303-10. doi: 10.1111/j.1525-139X.2012.01069.x. Epub 2012 Apr 4.*

An arteriovenous fistula (AVF) is the optimal vascular access for hemodialysis (HD), because it is associated with prolonged survival, fewer infections, lower hospitalization rates, and reduced costs. The AVF First breakthrough initiative (FFBI) has made dramatic progress, effectively promoting the increase in the national AVF prevalence since the program's inception from 32% in May 2003 to nearly 60% in 2011. Central venous catheter (CVC) use has stabilized and recently decreased slightly for prevalent patients (treated more than three months), while CVC usage in the first three months remains unacceptably high at nearly 80%. This high prevalence of CVC utilization suggests important specific improvement goals for FFBI. In addition to the current 66% AVF goal, the initiative should include specific CVC usage target(s), based on the KDOQI goal of less than 10% in patients undergoing HD for more than three months, and a substantially improved initial target from the current CVC proportion. These specific CVC targets would be disseminated through the ESRD networks to individual dialysis facilities, further emphasizing CVC avoidance in the transition from advanced CKD to chronic kidney failure, while continuing to decrease CVC by prompt conversion of CVC-based hemodialysis patients to permanent vascular access, utilizing an AVF whenever feasible.

Impact: Emphasizes that catheter avoidance should receive more attention than simply increasing the proportion of patients with an AVF.

Tamura, Manjula Kurella & Tan, Jane C & O'Hare, Ann M. **Optimizing renal replacement therapy in older adults: a framework for making individualized decisions.** *Kidney Int. 2012 Aug;82(3):261-9. doi: 10.1038/ki.2011.384. Epub 2011 Nov 16.*

It is often difficult to synthesize information about the risks and benefits of recommended management strategies in older patients with end-stage renal disease since they may have more comorbidity and lower life expectancy than patients described in clinical trials or practice guidelines. In this review, we outline a framework for individualizing end-

stage renal disease management decisions in older patients. The framework considers three factors: life expectancy, the risks and benefits of competing treatment strategies, and patient preferences. We illustrate the use of this framework by applying it to three key end-stage renal disease decisions in older patients with varying life expectancy: choice of dialysis modality, choice of vascular access for hemodialysis, and referral for kidney transplantation. In several instances, this approach might provide support for treatment decisions that directly contradict available practice guidelines, illustrating circumstances when strict application of guidelines may be inappropriate for certain patients. By combining quantitative estimates of benefits and harms with qualitative assessments of patient preferences, clinicians may be better able to tailor treatment recommendations to individual older patients, thereby improving the overall quality of end-stage renal disease care.

Impact: An individualized approach to vascular access decisions that relies on both quantitative assessment of benefits and harms, as well as patient preference, can lead to treatement decisions that contradict practice guidelines.

Ng, Leslie J & Chen, Fangfei & Pisoni, Ronald L et al. **Hospitalization risks related to vascular access type among incident US hemodialysis patients.** *Nephrol Dial Transplant. 2011 Nov;26(11):3659-66. doi: 10.1093/ndt/gfr063. Epub 2011 Mar 3.*

The excess morbidity and mortality related to catheter utilization at and immediately following dialysis initiation may simply be a proxy for poor prognosis. This study examined hospitalization burden related to vascular access (VA) type among incident patients who received some predialysis care using the DOPPS patient cohort (1996-2004) who reported predialysis nephrologist care. VA utilization was assessed at baseline and throughout the first 6 months on dialysis. Poisson regression was used to estimate the risk of all-cause and cause-specific hospitalizations during the first 6 months. Among 2635 incident patients, 60% were dialyzing with a catheter, 22% with a graft and 18% with a fistula at baseline. Compared to fistulae, baseline catheter use was associated with an increased risk of all-cause hospitalization [adjusted relative risk (RR) = 1.30, 95% confidence interval (CI): 1.09-1.54] and graft use was not (RR = 1.07, 95% CI: 0.89-1.28). Allowing for VA changes over time, the risk of catheter versus fistula use was more pronounced (RR = 1.72, 95% CI: 1.42-2.08) and increased slightly for graft use (RR = 1.15, 95% CI: 0.94-1.41). Baseline catheter use was most strongly related to infection-related (RR = 1.47, 95% CI: 0.92-2.36) and VA-related hospitalizations (RR = 1.49, 95% CI: 1.06-2.11). These effects were further strengthened when VA use was allowed to vary over time (RR = 2.31, 95% CI: 1.48-3.61 and RR = 3.10, 95% CI: 1.95-4.91, respectively). A similar pattern was noted for VA-related hospitalizations with graft use. Among potentially healthier incident patients, hospitalization risk, particularly infection and VA-related, was highest for patients dialyzing with a catheter at initiation and throughout follow-up, providing further support to clinical practice recommendations to minimize catheter placement.

Impact: Additional support for the association between catheter use and risk of hospitalization, particularly infection related hospitalizations.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2978_Evidence_form-635963068330082994.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Based upon data from the CMS Fistula First/Catheter Last initiative, a gradual trend towards lower catheter use has been observed among prevalent maintenance HD patients in the US, declining from approximately 28% in 2006 to approximately 18% by August 2015. Furthermore, the percentage of maintenance HD patients using a catheter for at least three months has declined as well over this time period from nearly 12% to 10.8%. Continued monitoring of chronic catheter use is needed to sustain this trend.
This measure is intended to be jointly reported with the Hemodialysis Vascular Access- Standardized Fistula Rate. These two vascular access quality measures, when used together, consider Arterial Venous (AV) fistula use as a positive outcome and prolonged use of a tunneled catheter as a negative outcome. With the growing recognition that some patients have exhausted options for an arteriovenous fistula, or have comorbidities that may limit the success of AVF creation, joint reporting of the measures accounts for all three vascular access options. The fistula measure adjusts for patient factors where fistula placement may be either more difficult or not appropriate and acknowledges that in certain circumstances an AV graft may be the best access option. This paired incentive structure that relies on both measures reflects consensus best practice, and supports maintenance of the gains in vascular access success achieved via the Fistula First/Catheter Last Project over the last decade.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Analysis of CROWNWeb data from January 2014- December 2014 indicated the facility-level mean percentage of patient-months with a long-term catheter was 11.6% (SD=6.6%). Distribution: Min=0%, 1st quartile=7.0 %, median=10.5%, 3rd quartile=14.9%, Max=58.2%.

Information about the data used in these analyses can be found under "Scientific Acceptability".

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Using the data from Jan-Dec 2014, age, sex, race, ethnicity and dialysis vintage were evaluated in a logistic regression model for long-term catheter use. Below we report the odds ratios for these patient characteristics. Age, sex, race, ethnicity and dialysis vintage are all statistically significant predictors of long-term catheter use. The analysis results indicate potential disparity in prolonged use of a tunneled catheter among these groups. Specifically, females are about 55% more likely to have a long-term catheter as males. Individuals 75 years of age and older were 14% more likely to have a long-term catheter and younger individuals 18-25 years of age were 46% more likely to have a long-term catheter when compared to patients 60-75 years of age. Those whose race is reported as "Other" were less likely to have a long-term catheter when compared to whites, as were Hispanics, when compared to non-Hispanics. Individuals whose duration of ESRD < 1 year and whose duration of ESRD are 9+ year were almost four times and 26% more likely to have a long-term catheter, respectively. Patients whose duration of ESRD are 5-<9 years were 8% less likely to have a

long-term catheter when compared to patients whose duration of ESRD are 1-<5 years. In the absence of biological effects explaining these differences, risk adjustment for these demographic factors could potentially mask disparities in care.

Odds ratio of having a catheter for at least three months:

Age:
For the 18-<25 age group, the Odds Ratio (95% CI) is 1.46 (1.12, 1.90), P-value is 0.005.
For the 25-<59 age group, the Odds Ratio (95% CI) is 1.06 (1.00, 1.121), P-value is 0.057.
The 60-<75 age group was used as the reference group.
For the 75+ age group, the Odds Ratio (95% CI) is 1.14 (1.07, 1.23), P-value is <.0001.

Sex:
For Female: the Odds Ratio (95% CI) is 1.55 (1.47, 1.63), P-value is <.0001.
Male was used as the reference group.

Race:
White was used as the reference group.
For Black: the Odds Ratio (95% CI) is 0.98 (0.91, 1.05), P-value is 0.586.
For Other race: the Odds Ratio (95% CI) is 0.77 (0.67, 0.88), P-value is <.0001.

Ethnicity:
For Hispanic: the Odds Ratio (95% CI) is 0.81 (0.74, 0.89), P-value is <.0001.
Non-Hispanic was used as the reference group.

Duration of ESRD:
For <1 year: the Odds Ratio was 3.97 (3.78, 4.18), P-value is <.0001.
1-<5 years was used as the reference group.
For 5-<9 years: the Odds Ratio was 0.92 (0.84, 1.00), P-value is 0.041.
For 9+ years: the Odds Ratio was 1.26 (1.15, 1.38), P-value is <.0001.

IN RESPONSE TO THE REQUIREMENTS FOR THE SDS TRIAL PERIOD*
*placing these results here per instructions from NQF staff, as the measure is not risk adjusted.

We performed the following analyses specifically to address the requirement for the NQF Trial Period for assessment of sociodemographic factors as potential risk adjustors. Sociodemographic factors included in the analysis were based on conceptual criteria and empirically demonstrated findings in the literature, which have shown that differences in long-term (= three months) catheter use exist among racial minorities, women and the poor. In addition, the particular patient and area level SDS/SES variables tested were based on availability of data for the analyses. We were able to acquire individual and area-level variables included in the Area Deprivation Index (ADI) developed by Singh and colleagues at the University of Wisconsin (Singh, GK. Area deprivation and widening inequalities in US mortality, 1969–1998. Am J Public Health. 2003;93(7):1137–1143).

The results below show the parameter estimates for patient- and area-level variables based on a logistic regression model for long-term catheter use (at least three months) that included these variables.

Sex:
For Female: The Odds Ratio is 1.47, and the P-value is <.0001.
Male was used as the reference group.

Race:
White was used as the reference group.
For Black: The Odds Ratio is 0.87, and the P-value is <.0001.
For Other: The Odds Ratio is 0.73, and the P-value is <.0001.

Ethnicity:
For Hispanic: The Odds Ratio is 0.77, and the P-value is <.0001.

Non-Hispanic was used as the reference group.

Employment Status:
Employed was used as the reference group.
For Unemployed: The Odds Ratio is 1.36, and the P-value is <.0001.
For Other: The Odds Ratio is 1.37, and the P-value is <.0001.

Medicare Coverage:
Medicare as primary w/o Medicaid was used as the reference group.
Medicare as primary with Medicaid: The Odds Ratio is 0.66, and the P-value is <.0001.
Medicare as secondary/Medicare HMO: The Odds Ratio is 0.94, and the P-value is 0.060.
For Non-Medicare/missing: The Odds Ratio is 1.41, and the P-value is <.0001.

ADI (zipcode-level):
Unemployment rate (%): The Odds Ratio is 0.997, and the P-value is 0.690.

Median family income: The Odds Ratio is 1.002, and the P-value is 0.881.

Families below the poverty level (%): The Odds Ratio is 0.999, and the P-value is 0.876.

Single-parent households with children <18 (%): The Odds Ratio is 0.998, and the P-value is 0.488.

Home ownership rate (%): The Odds Ratio is 0.997, and the P-value is 0.019.

Median home value: The Odds Ratio is 0.948, and the P-value is 0.052.

Median monthly mortgage: The Odds Ratio is 1.154, and the P-value is 0.058.

Median gross rent: The Odds Ratio is 0.887, and the P-value is 0.205.

Population (aged 25+) without High School diploma (%): The Odds Ratio is 0.997, and the P-value is 0.390.

Income disparity: The Odds Ratio is 1.003, and the P-value is 0.944.

Patient-level SDS/SES: Compared to males, females were more likely to have long-term catheter use (OR=1.47, p<0.01). Hispanics were less likely to have long-term catheter use (OR=0.77, p<0.01), compared to non-Hispanics. Compared to white patients, black patients and patients reporting other race were less likely to have long-term catheter use (OR=0.87, p<0.01; OR=0.73, p<0.01). As for employment status, unemployed patients and those with unknown or "other" status patients were more likely to have long-term catheter use (OR=1.36; p<0.01; OR=1.37; p<0.01), compared to employed patients. Note that for employment categories, the "Other" category represents diverse patient groups with regards to SDS/SES, such as students, homemakers, and those who are retired. Compared to Medicare only patients, patients with both Medicare and Medicaid and patients with Medicare as secondary/Medicare HMO were less likely to have long-term catheter use (OR=0.66, p<0.01; OR=0.94, p=0.06), and patients with no Medicare/missing were more likely to have long-term catheter use (OR=1.41, p<0.01). This latter result suggests that patients without Medicare coverage/missing are likely to have no insurance coverage and be in poorer health due to reduced access to health care.

Area-level SDS/SES: Area-level effects were all very small, and only three were statistically significant (p<0.05).

These analyses indicate that patient-level, but not area-level, variables for SDS/SES affect long-term catheter use. However, patient-level SDS/SES variables are not included as risk adjustment factors in the measure due to the absence of a convincing biological or clinical rationale that warrants accounting for different outcomes on the basis of race, sex, and the other patient-level SDS/SES factors tested. Area-level factors of SDS/SES are not included as adjustments due to the absence of clinically meaningful or statistically observed differences in long-term catheter use with these adjustment factors.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Numerous studies demonstrate that the long-term use of central venous catheters for HD access is associated with greater morbidity and higher mortality. Whereas catheters have the advantage of immediate use without need for maturation time, as enumerated in the Kidney Disease Outcomes Quality Initiative (KDOQI) guidelines, the long-term use of catheters is associated with substantially higher rates of infection-related complications and increased risk for central venous thrombosis, stenosis and occlusion. Numerous studies have shown that patients receiving dialysis using catheters have been found to have greater mortality risk than patients dialyzed with fistulas or grafts, whether or not diabetes mellitus was present. Higher case-mix adjusted mortality rates have been seen for HD patients dialyzing in facilities having greater catheter use.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. National Kidney Foundation: DOQI Clinical Practice Guidelines for Vascular Access. http://www.kidney.org/professionals/KDOQI/guidelines_commentaries

2. Grubbs V, Wasse H, Vittinghoff E, et al. Health status as a potential mediator of the association between hemodialysis vascular access and mortality. Nephrol Dial Transplant. 2014 Apr;29(4):892-8. doi: 10.1093/ndt/gft438. Epub 2013 Nov 13.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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:** 2978_Data_Dictionary_Code_Table.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The numerator is the number of adult patient-months in the denominator who were on maintenance hemodialysis using a catheter continuously for three months or longer as of the last hemodialysis session of the reporting month.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months

S.6. **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, 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.
The number of patient-months with a long-term catheter in use. Long-term catheter use is defined as using a catheter, at the same facility, for at least three consecutive complete months as of the last day of the reporting month.
For a given month, if any of the following CROWNWeb "Access Type IDs" (16,18,19,20,21,"·") has been recorded, a catheter is considered in use. If a catheter has been observed for three consecutive months (i.e., in the reporting month and the immediate two preceding months) at the same facility, the reporting month is counted in the numerator. Access Type ID "16" represents AV Fistula combined with a Catheter, "18" represents AV Graft combined with a Catheter, "19" represents Catheter only, "20" represents Port access only, "21" represents other/unknown, and "·" represents missing. If a patient changes dialysis facilities, the counting of the three consecutive complete months restarts at the new facility.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) for the complete reporting month at the same facility.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
For each patient, we identify the dialysis provider at each month using a combination of Medicare-paid dialysis claims, the Medical Evidence Form (Form CMS-2728), and data from CROWNWeb. These sources are used to identify patients that are receiving in-center or home hemodialysis for the entire reporting month. Patients are required to have been treated by the same facility for the complete month in order to be assigned to that facility for the reporting month.
To be included in the denominator for a particular reporting month, the patient must be receiving home or in-center hemodialysis for the complete reporting month at the facility, and be at least 18 years old as of the first day of the month.
The monthly patient count at a facility includes all eligible prevalent and incident patients. The number of patient-months over a time period is the sum of patients reported for the months covered by the time period. An individual patient may contribute up to 12 patient-months per year.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclusions that are implicit in the denominator definition include:

-Pediatric patients (<18 years old)
-Patients on Peritoneal Dialysis
-Patient-months under in-center or home hemodialysis for less than a complete reporting month at the same facility

In addition, the following exclusions are applied to the denominator:
Patients with a catheter that have limited life expectancy:
-Patients under hospice care in the current reporting month
-Patients with metastatic cancer in the past 12 months
-Patients with end stage liver disease in the past 12 months
-Patients with coma or anoxic brain injury in the past 12 months

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Determination of peritoneal dialysis treatment modality is derived from a combination of Medicare-paid dialysis claims, the Medical Evidence Form (Form CMS-2728), and data from CROWNWeb. These sources also determine patient assignment to the facility. Patients not treated by the facility for the entire month are excluded for that reporting month.

The patient's age is determined by subtracting the patient's date of birth from the first day of the reporting month. Patients that are < 18 years old as of the first day of the reporting month are excluded.

For the exclusion of catheter patients with limited life expectancy, catheter use in the reporting month is defined as the CROWNWeb "Access Type ID" having any of the following values: (16,18,19,20,21,"·"), where Access_Type_ID "16" represents AV Fistula combined with a Catheter, "18" represents AV Graft combined with a Catheter, "19" represents Catheter only, "20" represents Port access only, "21" represents other/unknown, and "·" represents missing.

Hospice status is determined from a separate CMS file that contains final action claims submitted by Hospice providers. Once a beneficiary elects Hospice, all Hospice related claims will be found in this file, regardless if the beneficiary is in Medicare fee-for-service or in a Medicare managed care plan. Patients are identified as receiving hospice care if they have any final action claims submitted to Medicare by hospice providers in the current month.

Diagnoses of metastatic cancer, end stage liver disease, or coma in the past 12 months were determined from Medicare claim types. Medicare claims include inpatient hospitalizations, outpatient claims (including dialysis claims), and physician services. Claims from providers, such as laboratories, that report diagnosis codes when testing for the presence of a condition are excluded. A detailed list of ICD-9/ICD-10 diagnostic codes used to identify these comorbidities is included in the attached data dictionary code table (excel file)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
See calculation flowchart in Appendix.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
We count patients with missing vascular access type in both the denominator and the numerator. Therefore missing vascular access type is counted as a catheter. For comorbidities used to determine the exclusions, if the patient had missing comorbidity values in the preceding 12 months of Medicare claims, we assume this patient did not have the comorbidity in that reporting month. The same methodology is applied to the hospice exclusion.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
CROWNWeb, Medicare Claims and the CMS Medical Evidence form 2728 are used as the data sources for establishing the denominator. CROWNWeb is the data source for establishing the numerator. Medicare claims are used for the comorbidity conditions exclusion criteria.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Dialysis Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2978_Testing_form.docx

Measure Number (*if previously endorsed*)**:**
Measure Title: Hemodialysis Vascular Access: Long-term Catheter Rate
Date of Submission: 4/15/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *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.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or

whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/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*).

National CROWNWeb data from January 2014-December 2014 and Medicare claims data from January 2013 – December 2014.

1.3. What are the dates of the data used in testing? January 2013-December 2014

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.26***)	**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 <u>measured entities</u> 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*)

Patients on both home and in-center hemodialysis during the last HD treatment of the month from January 2014-December 2014 were included in the analyses. The number of facilities per month ranged from 5,736-5,871 and the total number of patient-months ranged from 344,945- 363,257.

Public reporting of this measure on DFC or in the ESRD QIP would be restricted to facilities with at least 11 eligible patients throughout the year for the measure. We have applied this restriction to all the reliability and validity testing reported here.

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*)

There were a total of 4,274,619 eligible patient-months. Among those patient-months over the whole year, the average age was 62.7 years, 43.79% of patient-months were female, 56.27% were white, 37.05% were black, 6.68% reported race as "other", 18.41% were Hispanic and 46.37% had type II diabetes as the primary cause of ESRD.

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.

N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patient level:

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

**Assessed at a specific time point (e.g., at the reporting month). Medicare coverage in model was defined as:*
1. Medicare as primary and Medicaid
2. Medicare as primary and NO Medicaid

3. Medicare as secondary or Medicare HMO (e.g. Medicare Advantage)
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
- Income disparity
- Families below the poverty level (%)
- Single-parent households with children <18 years old (%)
- Home ownership rate (%)
- Median home value
- Median monthly mortgage
- Median gross rent
- Population (aged 25+) with <9 years of education (%)
- Population (aged 25+) without high school diploma (%)

NOTE: As this measure is not risk adjusted, the analysis results and interpretation for the above SDS factors are included in the response to question **1b.4** (Disparities) in the submission form.

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*)

We used January 2014 – December 2014 CROWNWeb data to calculate facility-level annual performance scores. The NQF-recommended approach for determining measure reliability is a one-way analysis of variance (ANOVA), in which the between-facility variation (σ_b^2) and the within-facility variation ($\sigma_{t,w}^2$) in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of a measure (i.e., $\sigma_b^2 + \sigma_{t,w}^2$) that is attributable to the between-facility variation, the true signal reflecting the differences across facilities. We assessed reliability by calculating inter-unit reliability (IUR) for the annual performance scores. If the measure were a simple average across individuals in the facility, the usual ANOVA approach would be used. The yearly based measure, 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 annual catheter rate for *N* facilities. To generate re-sampled data, we randomly draw patients from the national population B times (we set B=100). Using each re-sampled dataset, for the *i*th facility, we calculate an annual catheter rate ($T^*_{i,1},\ldots, T^*_{i,B}$) and their sample variance (S_i^*). From this it can be seen that

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

is a bootstrap estimate of the within-facility variance in the catheter rate, where n_i is the number of subjects in the *i*th facility. Calling on formulas from the one-way ANOVA, the total variation in the annual catheter rate (i.e., $\sigma_b^2 + \sigma_{t,w}^2$) can be estimated by

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

where the overall weighted average of catheter rate is $\bar{T} = \Sigma\, n_i\, T_i / \Sigma\, n_i$,, 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). 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 reliability calculation only included facilities with at least 11 patients during the entire year.

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 IUR is 0.765, which indicates that 76.5% of the variation in the annual long-term catheter rate can be attributed to between-facility differences in performance (signal) and about 23.5% to the within-facility variation (noise).

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?*)

The result of IUR testing suggests a high degree of reliability.

2b2. VALIDITY TESTING

2b2.1. 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 <u>performance measure score</u> 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*)

2b2.2. 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 quintiles of performance scores and the 2014 Standardized Mortality Ratio (SMR, NQF 0369) and 2014 Standardized Hospitalization Ratio (SHR, NQF 1463). Facility-level performance scores were divided into quintiles (Q1 to Q5), and the relative risk (RR) of mortality (and hospitalization, separately) was calculated for each quintile, using the combined Q1 and Q2 as the reference group. Thus, a RR>1.0 would indicate a higher relative risk of mortality or hospitalization, compared to the lowest performance score quintiles.

In 2015 a vascular access TEP was convened to provide input on the development of access measures, and specifically input on exclusions for both catheter and fistula measures, and for fistula, risk adjustment factors to be considered. The TEP felt that minimizing catheter use is paramount and that while catheters may potentially be acceptable for some patients, they addressed this through identifying patient level exclusion criteria rather than risk adjustment. The candidate catheter measure was reviewed and validated by the Technical Expert Panel (TEP) in 2015.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Quintiles of the performance scores were defined as follows:

Q1*: 0.0%-<6.24%

Q2*: 6.24%-<9.12%

Q3: 9.12-<12.00%

Q4: 12.00%-<16.21%

Q5: 16.21%-<58.16%

*Q1 and Q2 as Reference

Results from the Poisson model indicated that the percent of patient-months with a long-term catheter was significantly associated with the risks of mortality and hospitalization.

For the 2014 SMR, the relative risk of mortality increased as the performance measure quintile increased from the reference group (combined Q1 and Q2). For quintile 3, RR=1.03 (95% CI: 1.01, 1.05; p=0.006), quintile 4, RR=1.03 (95% CI: 1.01, 1.05; p=0.008), and quintile 5, RR=1.09 (95% CI: 1.07, 1.12; p<0.001).

Similarly for the 2014 SHR, the relative risk of hospitalization increased as the performance measure quintile increased from the reference group (combined Q1 and Q2). For quintile 3, RR=1.08 (95% CI: 1.08, 1.08; p<0.001), quintile 4, RR=1.10 (95% CI: 1.10, 1.10; p<0.001), and quintile 5, RR=1.16 (95% CI: 1.15, 1.16; p<0.001).

2b2.4. 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?*)

Results of the Poisson regression suggest the predictive relationship of higher catheter use with higher mortality and hospitalization, as measured by the respective standardized mortality and hospitalization rates, compared to facilities with a lower proportion of patients with a long-term catheter.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

The following exclusions are applied to the denominator:

Patients with a catheter that have limited life expectancy. Limited life expectancy is defined as:

- Patients under hospice care in the current reporting month
- Patients with metastatic cancer in the past 12 months
- Patients with end stage liver disease in the past 12 months
- Patients with coma or anoxic brain injury in the past 12 months

The facility-level mean percentage of patient-months with a catheter for at least three months with and without the patient-month exclusions are calculated and compared.

2b3.2. 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*)

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

Table 1: Percent of patient-months at risk excluded

Year	Before Exclusion	After Exclusion	Percent
2014	4,314,450	4,274,619	0.92%

Table 2: Number and percent of unique patients excluded

Year	Before Exclusion	After Exclusion	Percent
2014	468,910	457,902	2.35%

Table 3: Distribution of performance scores before and after the exclusion

Catheter Rate	N	Mean	Standard Deviation	Minimum	Maximum
Before exclusion	5928	0.121	0.068	0.000	0.597
After exclusion	5928	0.118	0.066	0.000	0.582

Figure 1: Scatterplot – Facility Catheter Rate with and without Exclusions

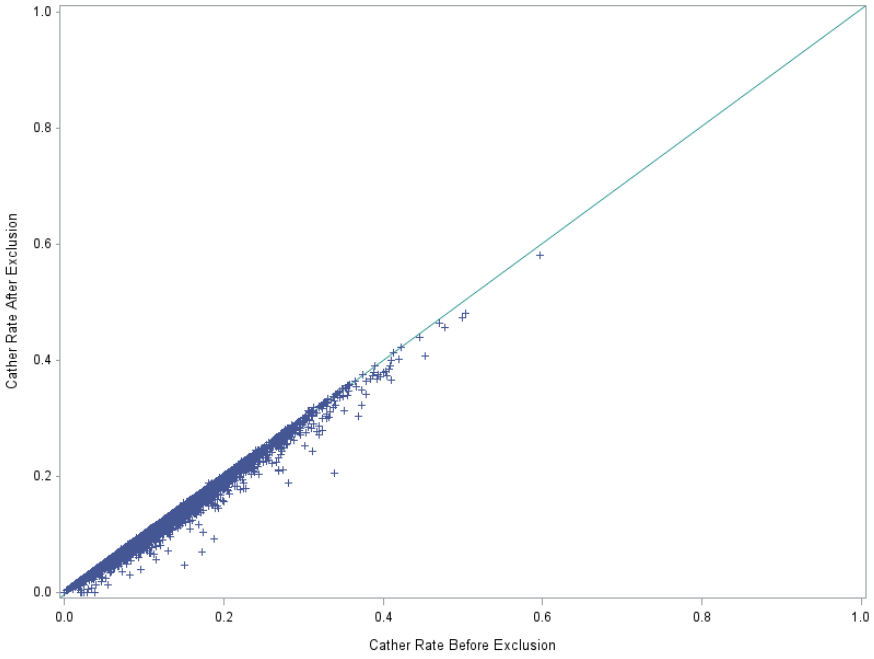
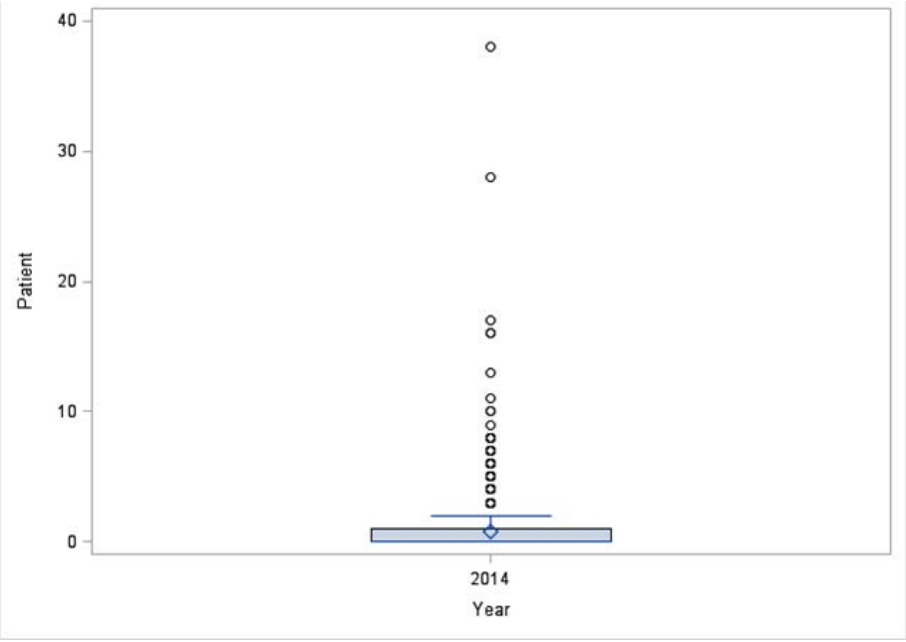


Figure 2. Distribution of Excluded Patients at the facility level for 2014



2b3.3. 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 exclusion criteria are necessary since the percentage of patients excluded at each facility is not evenly distributed across facilities (Distribution shown in the boxplot). Due to the unequal distribution across facilities, the exclusion criteria take into account that some facilities treat a higher portion of patients with limited life expectancy. Additionally, our results shown in both the scatter-plot (Figure 1) as well as the Pearson Correlation Coefficient of 0.993 (p-value <0.0001) between the mean percentage of patient months with a long-term catheter with and without the exclusion suggests that the overall impact of the exclusion on the measure's validity is not substantial since the two are highly correlated.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Risk adjustment is not appropriate for this measure because of the primary goal of disincentivizing catheter use for incident and particularly prevalent dialysis patients. This measure was reviewed by the 2015 vascular access TEP which also did not recommend risk adjustment.

The TEP felt that minimizing catheter use is paramount and that while catheters may potentially be acceptable for some patients, they addressed this through identifying patient level exclusion criteria rather than risk adjustment, so as not to penalize providers that treat patients that have limited life expectancy or limit those patients' access to care.

Consistent with the TEP's concerns, potential risk adjustors in a catheter measure would apply to a large portion of both incident and prevalent ESRD patients, and therefore may not function as a disincentive to reduce catheter use, which is the intent of the measure. Applying the exclusions more appropriately accounts for conditions in a very specific subset of patients where a catheter may be the only or an acceptable access type. Additionally, the fistula measure (intended to be reported with the catheter measure) includes risk adjustment based on the TEP's recommendation that facility success in fistula use (versus graft or catheter) will be limited in patients with certain comorbidities and other patient characteristics.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

N/A

2b4.4a. What were the statistical results of the analyses used to select risk factors?

N/A

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

N/A

2b4.5. 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 2b4.9

N/A

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

N/A

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

N/A

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

N/A

2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. 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*)

N/A

2b4.11. 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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

Differences in measure performance were evaluated separately for each facility using patient level analyses. For each facility, the proportion of patient-months with catheter ≥ three months, calculated at the year-level, was compared to the overall national distribution.

Note that the monthly based measure is a simple average of binary outcomes across individuals in the facility, for which the binary outcome equals 0 if no catheter is present, and equals 1 if a catheter ≥ three months is present. The differences in proportions can be compared using Fisher's Exact tests or its normal approximation. The yearly based measure, however, is not a simple average of binary outcomes and we instead used a re-sampling based exact test, with re-sampling generated from the population distribution of the patient level outcomes. Due to the non-symmetric structure of the measure distributions, a one-sided test with significance level 0.025 is used (corresponding to a cutoff=0.05 in a two-sided test). To calculate the p-value, we assess the probability that patients in each facility would experience a number of events (i.e., months dialyzing with catheter ≥ three months) more extreme than what was actually observed if the null hypothesis were true, where the null hypothesis is that a patient in each facility will follow the overall national distribution.

2b5.2. 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*)

Proportion of facilities with statistically significant differences (p-value < 0.025) is shown as follows:

Category	Number of facilities	Percent of facilities
As expected	5,211	87.9%
Worse than expected	717	12.1%

2b5.3. 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?*)

For the annual percentage of patients with a long-term catheter as the performance measure, 5,211 (87.9%) facilities have achieved expected performance, and 717 (12%) facilities have performed worse than expected (higher catheter rate).

In general, lower rates of catheter use for three months or more represent better quality of care. This analysis demonstrates both practical and statistically significant differences in performance across facilities based on their proportion of patient months with a catheter for three months or greater.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

2b7.2. 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)

2b7.3. 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*)

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)

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)*

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.

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.

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. Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

N/A

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.

Planned	Current Use (for current use provide URL)
Public Reporting Payment Program	

4a.1. For each CURRENT use, checked above, provide:

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

N/A

4a.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?*)

Measure is currently under development.

4a.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.*)

CMS will determine if and when the measure will be implemented in a CMS program. Upon endorsement, CMS will consider retiring the currently endorsed measure of catheter use (#0256) in favor of this new measure for implementation in a future performance year for the ESRD QIP and reporting period for Dialysis Facility Compare at the next available opportunity.

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. 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.

The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of this catheter measure. Once implemented facility performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing prolonged catheter use, while accounting for patients where a long-term catheter may be an appropriate vascular access choice.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

Potential unintended consequences would center on patients being pushed towards having an AVF or AVG created when they may not realize a benefit from this type of access due to either comorbidities or a limited life expectancy. It is incumbent on both dialysis facilities, as well as surgeons, to establish the most appropriate type of vascular access for patients based on their individual circumstances and preferences, rather than in response to quality measures.

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.

Yes

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

0251 : Vascular Access—Functional Arteriovenous Fistula (AVF) or AV Graft or Evaluation for Placement

2594 : Optimal End Stage Renal Disease (ESRD) Starts

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

5a. Harmonization

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 completely harmonized?

No

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

Measure 0251 contains several components including AV fistula use, AV graft use or referral to a vascular surgeon (or other qualified physician) if using a long-term catheter. It is a referral process measure for those patients with a catheter. This has the potential for facilities to score well on the measure even if they have patients with a catheter, as long as the patient was referred to or evaluated by a vascular surgeon. We acknowledge this is an important step to fistula placement however it departs from the intent of the catheter measure to function as a more direct disincentive to prolonged catheter use, consistent with the concerns and recommendations made by the vascular access TEP. Measure 2594 is not directed toward dialysis facilities. The setting focus addresses a different provider type which falls outside the purview of measures evaluating dialysis facility performance on prolonged catheter use. These suggest fundamental differences in measure target populations, setting and intent that cannot be harmonized. Additionally, the measure is limited to incident patients, while the long-term catheter rate measure includes both incident and

prevalent patients as the measured population.

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.)
There are no competing measures.

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:** 2978_Appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, sophia.chan@cms.hhs.gov
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, 734-548-3057-

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.
According to the CMS Measure Management System Blueprint, TEPs are advisory to the measure contractor. 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.

Joseph Vassalotti, MD, FASN, FNKF
Chief Medical Officer, National Kidney Foundation
Associate Professor of Medicine, Division of Nephrology
Mount Sinai Medical Center
New York, NY

Monet Carnahan, RN, BSN, CDN
Renal Care Coordinator Program Manager
Fresenius Medical Center (FMC)
Franklin, TN
American Nephrology Nurses Association

Derek Forfang
Patient Leadership Committee Chair
ESRD Network 17
Board Member
Intermountain End State Renal Disease Network Inc.
Beneficiary Advisory Council (Vice Chair)
The National Forum of ESRD Networks

Board Member
The National Forum of ERSD Networks
San Pablo, CA

Lee Kirskey, MD
Attending staff, Department of Vascular Surgery
Cleveland Clinic Foundation
Cleveland, OH

Nance Lehman
Board Member
Dialysis Patient Citizens (DPC)
Billings, MT

Charmaine Lok, MD, MSc, FRCPC (C)
Medical Director of Hemodialysis and Renal Management Clinics
University Health Network
Professor of Medicine
University of Toronto
Toronto, ON

Lynn Poole, FNP-BC, CNN NCC
Fistula First Catheter Last Project Clinical Lead
ESRD National Coordinating Center
Lake Success, NY

Rudy Valentini, MD
Chief Medical Officer
Children's Hospital of Michigan (CHM)
Professor of Pediatrics, Division of Nephrology
Wayne State University School of Medicine

Daniel Weiner, MD, MS
Nephrologist, Tufts Medical Center
Associate Medical Director
DCI Boston
Associate Professor of Medicine
Tufts University School of Medicine
Boston, MA

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, 2017

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

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

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2979
Measure Title: Standardized Transfusion Ratio for Dialysis Facilities
Measure Steward: Centers for Medicare & Medicaid Services
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.
Developer Rationale: Several changes in the ESRD system are likely to impact anemia management. These include identification of safety concerns associated with aggressive erythropoiesis-stimulating agent (ESA) use, expansion of the ESRD Prospective Payment System bundled payment, and the development of the ESRD Quality Incentive Program. There are concerns that these changes could result in underutilization of ESAs, with lower achieved hemoglobin values that may increase the frequency of red blood cell transfusion in the US chronic dialysis population.
Blood transfusion may be an indicator for underutilization of treatments to increase endogenous red blood cell production (e.g. ESA, iron). In addition, dialysis patients who are eligible for kidney transplant and are transfused risk the development of becoming sensitized to the donor pool thereby making transplant more difficult to accomplish. Blood transfusions carry a small risk of transmitting blood borne infections, development of a transfusion reaction, and using infusion centers or hospitals to transfuse patients is expensive, inconvenient, and could compromise future vascular access.
Monitoring the risk-adjusted transfusion rate at the dialysis facility level, relative to a national standard, allows for detection of treatment patterns in dialysis-related anemia management. This is of particular importance due to FDA guidance regarding minimizing the use of ESAs, and economic incentives to minimize ESA use introduced by Medicare's bundling of payment for ESAs. As providers use less ESAs in an effort to minimize the risks associated with aggressive anemia treatment it becomes more important to monitor for an overreliance on transfusions.

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.
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.
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.

Measure Type: Outcome

Data Source: Administrative claims, Electronic Clinical Data
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This measure calculates 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.
- The developer provides the following rationale for the measure:
    - Several changes in the ESRD system are likely to impact anemia management, including:
        - Identification of safety concerns associated with aggressive erythropoiesis-stimulating agent (ESA) use
        - Expansion of the ESRD Prospective Payment System bundled payment
        - Development of the ESRD Quality Incentive Program.
    - There are concerns that these changes could result in underutilization of ESAs, with lower achieved hemoglobin values that may increase the frequency of red blood cell transfusion in the US chronic dialysis population.
    - Blood transfusion may be an indicator for underutilization of treatments to increase endogenous red blood cell production (e.g. ESA, iron).
    - Dialysis patients who are eligible for kidney transplant and are transfused risk the development of becoming sensitized to the donor pool thereby making transplant more difficult to accomplish. Blood transfusions carry a small risk of transmitting blood borne infections, development of a transfusion reaction, and using infusion centers or hospitals to transfuse patients is expensive, inconvenient, and could compromise future vascular access.
    - Monitoring the risk-adjusted transfusion rate at the dialysis facility level, relative to a national standard, allows for detection of treatment patterns in dialysis-related anemia management. This is of particular importance due to FDA guidance regarding minimizing the use of ESAs, and economic incentives to minimize ESA use introduced by Medicare's bundling of payment for ESAs. As providers use less ESAs in an effort to minimize the risks associated with aggressive anemia treatment it becomes more important to monitor for an overreliance on transfusions.

Guidance from the Evidence Algorithm :
Measure assesses performance of a health outcome (Box 1)→ Relationship established between measured health outcome and at least one healthcare action (Box 2)→ PASS (eligible for PASS rating)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence?*

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Data for the measure are derived from an extensive national ESRD patient database, which is largely derived from CMS Consolidated Renal Operations in a Web-enabled Network (CROWN).
- Information on transfusions is obtained from Medicare Inpatient and Outpatient Claims Standard Analysis Files (SAFs).
- Standardized transfusion ratios vary across facilities. The data below show the distribution of STrR using Medicare claims data for 2011-2014.

Year	Facilities	Mean STrR	Standard Error	10th percentile	50th percentile	90th percentile
2011	5774	1.029	1.348	0.199	0.863	1.896
2012	5943	1.023	0.972	0.217	0.866	1.864
2013	6170	1.057	2.883	0.213	0.866	1.897
2014	6415	1.034	1.408	0.171	0.867	1.843

- The data below show the number of facilities, patients, total count of transfusions and total patient years at risk for each year. Also, an unadjusted or raw transfusion rates per year (defined as total transfusions divided by total patient years at risk).

Year	Facilities	Patients	Total transfusions	Total Patients Years at risk	Transfusion Rate per 100 patient years at risk*
2011	5774	387097	67428	227935.62	29.58
2012	5943	398769	74444	234847.09	31.70
2013	6170	415576	73122	241082.06	30.33
2014	6415	429241	69182	246710.49	28.04

*This analysis includes all facilities for the given year.

Disparities

- Analyses of the STrR by race, sex and ethnicity indicate relatively little variation and no disparities substantial to the measure among these groups.
- The data below shows the parameter estimates for the race, sex and ethnicity variables included in the model containing the other covariates listed in S.14.

	Estimate	Standard Error	p-value
Females	0.168	0.004	<.0001
Native American*	-0.075	0.023	<.0001
Asian*	-0.207	0.012	<.0001
Black*	-0.046	0.005	<.0001
Other Race*	0.090	0.045	0.044
Hispanic #	-0.181	0.007	<.0001

*White as reference

Non-Hispanic as reference

Questions for the Committee:

- *Would greater variation in care be seen if the measure were expressed as a rate rather than a ratio?*
- *How do you interpret the mean measure result in light of a rather large standard error?*
- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ High ☒ Moderate ☐ Low ☐ Insufficient
Committee pre-evaluation comments **Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data

Specifications:

- The numerator of this measure is: *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.*
- The denominator of this measure is: *Number of eligible red blood cell transfusion events that would be expected among patients at a facility during the reporting period, given the patient mix at the facility.*
- The measure has the following exclusions:
  - All transfusions associated with transplant hospitalization
  - Patients 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.
- The ICD-9 and ICD-10 codes have been included in the Data Dictionary Code Table.
- The calculation algorithm is stated in the appendix and appears straightforward.
- This outcome measure is risk adjusted, using a statistical risk model.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- The reliability of the STrR was assessed using data from ESRD dialysis patients during 2011-2014.

- Since the STrR is not a simple average, the developer estimated the inter-unit reliability (IUR) using a bootstrap approach, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by an one-way analysis of variance (ANOVA).
- The developer states 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.

Results of reliability testing :

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
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

- The STrR calculation only included facilities with at least 10 patient years at risk.
- 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, larger facilities have greater IUR.

Guidance from the Reliability Algorithm:

Submitted specifications precise, unambiguous and complete (Box 1)→ Empirical reliability testing conducted (Box 2)→ Testing conducted with measure score at entity level (Box 4) →Method described and appropriate (Box 5) → Level of certainty or confidence that performance measure score is reliable (Box 6): 8→9→10 eligible for MODERATE rating for larger sample sizes.

Questions for the Committee:

- o *Is the test sample adequate to generalize for widespread implementation?*
- o *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- 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.
    - Using this model, 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.
    - Using this model, 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.
- In May 2012 there was an assessment of the measure's face validity based on polling of a CMS Technical Expert Panel (TEP).

Validity testing results:

Association of STrR with other facility-level outcomes: The developer states the results from the Poisson model indicated that the STrR tertiles were significantly associated with both SMR and SHR.

For the 2014 SMR:

Tertiles of STrR	RR	95% CI	P
T1*: 0-<0.66	AS REFERENCE		
T2: 0.66-<1.15	1.06	1.04, 1.08	<0.001
T3: 1.15-<5.66	1.14	1.12, 1.16	<0.001

For 2014 SHR:

Tertiles of STrR	RR	95% CI	P
T1*: 0-<0.66	AS REFERENCE		
T2: 0.66-<1.15	1.11	1.10, 1.11	0.001
T3: 1.15-<5.66	1.29	1.29, 1.30	0.001

Association of STrR with facility-level intermediate anemia management outcome: The developer states the results from the Poisson model indicated that the % of patients with Hgb < 10 was significantly associated with the risks of transfusion.

Tertiles of % of patients with Hgb < 10	RR	95% CI	P
T1: 0-<9.5%	AS REFERENCE		
T2: 9.5%-<16.5%	1.15	1.13, 1.18	<0.001
T3: 16.5%-<85.3%	1.31	1.28, 1.33	<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.
- The overall measure demonstrates face validity based on the structured 2012 TEP vote.

Questions for the Committee:

- *Do you expect a positive correlation among the StrR, SMR, and SHR?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Transfusions associated with transplant hospitalization are excluded as they mark a transition of care from the dialysis facility to a transplant team.
- 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.
- 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 developers present 4 year data on frequency of exclusions:
  - 2011-2014 patient years: 20% excluded
  - 2011-2014 patients: 14% excluded

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary:

The developer provided the following information:

- The calculation of the STrR is a two-stage approach.
  - Stage 1: The model is first fitted to the national data with piecewise-constant baseline rates stratified by facility; transfusion rates are adjusted for patient age, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, and calendar year.
    - This model allows the baseline transfusion rates to vary between strata (facilities), but assumes that the regression coefficients are the same across all strata

        - This approach is robust to possible differences between facilities in the patient mix being treated. The regression parameter estimates from Stage 1 are used to compute the expected number of transfusions for each patient.
        - 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 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.
    - Table 10 lists results from the Stage 1 model that 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.
    - Stage two: The expected number of transfusions by facility is summed, then computing facility-specific STrRs as the ratio of observed / expected transfusions.
- 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.
    - 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.
        - 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.
- 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).

Questions for the Committee:

- *Is there a conceptual relationship between the SDS factor(s) and the measure focus?*
- *Does empirical analysis (as provided by the measure developer) show that the SDS factor(s) has a significant and unique effect on the outcome in question?*
- *Does the reliability and validity testing match the final measure specifications?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *How well do the SDS variables that were available and analyzed align with the conceptual description provided?*
- *Are these variables available and generally accessible for the measured patient population?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- In order to classify facilities as having transfusion rates that are better, no different or worse than the national average, a Z-score was first calculated using the estimate and standard error for each facility using the method of generalized estimating equations.
- 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).
- 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, a 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" and provide references for assessing the extent to which a given facility's outcomes are extreme in comparison with other facilities.The developer then used the mean and standard deviation from the empirical null distribution of the STrR z-scores to calculate the p-value for classifying facility performance.
- The following table shows how the facilities are flagged for the year 2014, based on the method described above.

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

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

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

- The developer states 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.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not needed.

2b7. Missing Data An analysis of missing data analysis was not provided on this measure.

Guidance from the algorithm:
Specifications consistent with evidence (Box 1) → Potential threats to validity addressed – meaningful differences may be an issue (Box 2) → empirical testing of measure score (Boxes 3 and 6) → appropriate method (Box 7) → moderate certainty (Box 8b) -- moderate

Preliminary rating for validity: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: None

2a2. Reliability Testing

Comments:

**LDOs have large IUR, while SDOs have relatively low IURs. Thus, reliability is limited in SDOs - - may be OK for LDOs.
Preliminary rating Reliability: Moderate
**Value of IUR indicates a moderate degree of reliability. When stratified by facility size, larger facilities have greater IUR.
**The sample is adequate.
**It is not clear why the developer chose the cutoff points for patient census between small, medium and large dialysis centers. These seem to differ from reliability testing measures for similar measures submitted by the same developer. Could the developer provide a rationale for the cutoff points and why they vary from measure to measure?
**IUR overall was 0.60 - 0.66, however, highly variable based on unit size and IUR for smaller units in 2014 was 0.30
**reliability demonstrated
**reliability testing based on at least 10 pt yrs at risk but the measure does not exclude facilities with lower number of pt yrs at risk so unclear if reliable with these facilities
IUR overall 0.6 to 0.66 but for small facilities only 0.3 to 0.4 and only around 0.5 for medium sized facilities -- is this good enough for these sized units?
**Reliability testing was adequate in scope and methods are sound. Only concern is that IUR are much lower for small facilities (<= 46 patients).
**No concerns

2b2. Validity Testing

Comments:

**Transfusion ratios appear to correlate with SMR and SHR. I am not sure what this tells us - - it si possible that healthier people have low mortality and hospitalizations and stable Hb levels - - maybe independent of our care.
I do not see this score as a measure of quality
**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.
The results from the Poisson model indicated that the STrR tertiles were significantly associated with both SMR and SHR.
**The correlation between StrR, SMR, and SHR is not surprising.
**The conceptual model for the ability to the dialysis unit to manage this measure would be through hemoglobin. It would be good to see a correlation for validity between this measure and facility hemoglobin.
Given that the TEP occurred immediately after the black box changes in 2012, and given the trends in data for transfusion described by the USRDS and the developer, 2012 may be too old to ascertain face validity.
**Validity was examined by assessing the association between STrR and the SHR or SMR; 2) Association between HgB < 10 and STrR; and 3)face validity. Further clarification of timeframes used in validity testing of 1) and 2) may aid in the interpretation of findings.

**Validity demonstrated
**face validity from TEP
regression models comparing SMR, SHR, anemia data with STrR
**Empirical validity testing was good - tertiles associated with SHR and SMR. Face validity also given.
**No concerns

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**I do not see a conceptual relationship between SDS and transfusion ratio
6.5% of facilities have a "worse than expected" transfusion ratio. Do these facilities have lower ESA and Fe use - - and does it matter? I am not convinced it matters.
I do not find meaningful information about quality.
Preliminary rating for validity: Low
**patient groups are not inappropriately excluded.
An appropriate risk adjustment is provided.
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
**Increased transfusion may reflect comoribid illness and lack of repsonsiveness to epo not accounted for by current exclusions. The number of "worse than expected" facilities is small.
**As noted above the lack for an exclusion related to GI bleeds is problematic and presents a threat to validity.
The one year look back for claims data seems to vary from the length of lookback periods for other claims based measures from the same developer. Could the developer explain this variability?
**Additional exclusions may be warranted - GIB, etc.
Risk adjustment necessary and appears appropriate
Measure distinguishes facility performance (6% worse than expected, 0.4% better than expected)
**exclusions are appropriate
**Meaningful differences: 93% as expected -- only 0.4% better and 6.5% worse -- does not seem as if measure shows difference if better performance exists
**Yes, an SDS factor analysis was presented. A few questions/concerns: 1) a conceptual relationship was not really discussed 2) Certain SDS factors (female sex, Hispanic) did have an impact but only 1.6% (91) facilities changed categories. It is a philosophical question whether this warrants SDS adjustment.A detailed explanation for many of these factors is provided but not Hispanic ethnicity.
Comorbidities for the base model to determine STr are not time updated with claims data.
**No concerns

Criterion 3. Feasibility

3. Feasibility is the 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.

- All data elements are in defined fields in electronic form and generated or collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: None

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- This measure is publically reported nationally in Dialysis Facility Compare (DFC).
- The measure has been finalized for use in End Stage Renal Disease Quality Incentive Program starting PY2018.

Improvement results: Given that the measure has only been publically reported for a short time, progress on improvement could not be evaluated.

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms: A potential unintended consequence of the STrR would be to create an incentive for dialysis facilities to target higher hemoglobin levels. The literature suggests that targeting to hemoglobin concentrations above 12 to 13 grams per deciliter is associated with elevated risk of cardiac events and related mortality. Transfusion avoidance is optimized with achieved hemoglobin in the 10 to 11 grams per deciliter range. Therefore, the developer believes the potential for unintended consequences is very low with appropriate provider anemia management practices.

Feedback : No feedback provided on QPS. During the 2012-2013 MAP review, MAP supported the direction of this measure for inclusion in the End-Stage Renal Disease Quality Reporting. The measure was titled "Risk-Adjusted Facility Level Transfusion Rate "STrR"" at the time of review. They stated the measure addresses an important concept, but establishment of guidelines for hemoglobin range is needed. Public comments support MAP's conclusion, noting that there is no solid evidence to fully support this measure.

Questions for the Committee:

- *The measure description states "This measure is calculated as a ratio, but can also be expressed as a rate." Is ratio or rate more usable for public reporting?*
- *How usable is this measure for identifying variations in care when more than 90% of facilities as "as expected"?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments:

**Has been a publicly reported metric.

Potential harm: withhold transfusion for patients with symptoms of anemia

Preliminary rating Usability and Use: Low

**This measure is publically reported nationally in Dialysis Facility Compare (DFC).

The measure has been finalized for use in End Stage Renal Disease Quality Incentive Program starting PY2018.

**Usability is high.

**Dialysis units do not have access to transfusion data outside the dialysis unit which represent the overwhelming majority of transfusions. Therefore there are issues with usability. CMS should provide dialysis units the data on a monthly basis from the six month lagged claims file.

The above, at, or below the national norm is concerning in this ratio given the known regional variability.

**Current uses include public reporting and payment

**Currently reported. However, context of this PM is affected by changing reimbursement and drug safety issues.

**used in DFC

planned for QIP in 2018

**Measure already approved for CMS QIP in PY 2018. Already an existing DFC measure.

One concern is that while there is dispersion of results across facilities, additional data given shows that ~ 93% of facilities are as expected or better. This raises question about is this measure already topped out?

Unintended consequence could be to target higher hgb (>12) but transfusion avoidance should be ok with hgb > 10. And there is an anemia measure (#months of hgb and ESA) in the QIP.

**Proposed to be used in 2018

Criterion 5: Related and Competing Measures

Related or competing measures

- None identified

Pre-meeting public and member comments

Comment by Joseph Vassalotti
Organization National Kidney Foundation
Comment #5702: The National Kidney Foundation (NKF) believes that a transfusion avoidance measure is important to protecting patients from unnecessary transfusions. Risks of red blood cell transfusions in dialysis patients include hyperkalemia, volume overload and antigen sensitization for a potential future kidney transplant. However, a transfusion avoidance measure should be stratified to appropriately capture blood

transfusions that could have been prevented by the dialysis facility and exclude other reasons for transfusions. To this end we appreciate the exclusions of certain patient populations that are likely to experience anemia and may require blood transfusions due to other comorbid conditions. NKF acknowledges tracking blood transfusion data are critical to understanding patient safety hazards. NKF also recognizes that since most blood transfusions are provided outside of the dialysis setting how transfusions are reported and submitted as claims to CMS may vary by hospital and by patient and this could cause variation in performance on the StR. NKF encourages CMS to explore ways to ensure hospitals appropriately report and standardize reporting on blood transfusions for dialysis patients.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5689: During the last project, this Standing Committee reviewed the STrR as #2699 and did not recommend it. As we discuss further in the section on Validity, we do not believe the new measure addresses the Committee's concerns about hospital- and physician-related factors. We comment on the specifications, reliability, validity (risk model), and harmonization issues.

SPECIFICATIONS. CMS has revised the measure specifications to more "conservatively" define transfusion events, such that all inpatient transfusion events must include, at a minimum, an appropriate ICD-9 Procedure Code or Value Code to be captured in the measure—inpatient transfusion events for claims that include only 038 or 039 revenue codes without an accompanying procedure or value code are not captured in the numerator. The specifications also specify a maximum of one event per day and that an event not be defined by the number of units of blood transfused.

KCP supports and appreciates the need to refine and tighten how transfusion events are counted and applauds CMS's intent in undertaking these revisions, but we do not believe the proposed solution is a valid representation of transfusion events. Importantly, there is no existing coding requirement that procedure or value codes be used, which means valid transfusion claims that include only revenue codes will be missed. KCP believes the proposed specification changes result in a measure with significant threats to validity.

Current transfusion coding practices clearly vary by hospital,[3] and hospital coding practices are beyond dialysis facilities' sphere of control. For example, we are aware of hospitals that exclusively use revenue codes and do not use the procedure or value codes. In-patients at this type of hospital will appear to have no transfusion events assigned to the dialysis facility, whereas those at a hospital that uses the procedure and/or value codes will have recorded events. Simply put, facilities within given catchment areas will be differentially affected by hospital coding variations, which clearly impact measure scoring. We are particularly concerned that the revisions, if implemented, will result in increased variability in performance across dialysis facilities wholly due to external factors and not performance. Facilities will appear to have "poor" performance because of higher than expected numbers of transfusions—and will expend time and resources to improve—when in fact the score is merely a reflection of coding practices.

[3]Weinhandl ED, Gilbertson DT, Collins AJ. Dialysis facility-level transfusion rates can be unreliable due to variability in hospital-level billing patterns for blood. Chronic Disease Research Group poster, ASN. 2014.

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5690: SPECIFICATIONS (cont.). Again, KCP strongly supports the need to refine how transfusion events are defined, and we urge the Standing Committee to recommend the developer continue considering alternative models to define transfusion events. Alternatively, the Committee could suggest that CMS consider

revising hospital transfusion coding rules to require that the ICD-9/ICD-10 procedure and value codes necessary for the validity of the proposed methodology be universally included in claims.

Additionally, the testing documentation notes that facilities with 10 or fewer patients were excluded, but we note the specifications do not state this. Again, KCP believes that a minimum size exclusion should be indicated and, as the developer's results document, and we discuss in the following section, reliability is poor even when the facility size is significantly greater than 10 patients.

The submission also indicates the minimum data requirement for the STrR is 10 patient-years at risk, which differs from the SHR, which uses 5 patient-years at risk. No justification or empirical analyses are offered to justify the selected threshold or the difference.

Finally, the STrR specifications indicate the measure can be expressed as a rate, but is calculated as a ratio. KCP prefers normalized rates or year-over-year improvement in rates instead of a standardized ratio. We believe comprehension, transparency, and utility to all stakeholders is superior with a scientifically valid rate methodology.

RELIABILITY. In addition to our concerns that the specifications pose a threat to the validity of the updated STrR, KCP also has concerns about the reliability testing for these revised specifications.

KCP again notes a reliability statistic of 0.70 is often considered as "good" reliability,though the characterization also depends on the analytic method. Reliability testing, overall, for the STrR yielded IURs of 0.60-0.66 across all facilities for each of 2011, 2012, 2013, and 2014. Such values indicate about 65% of the variation in a score can be attributed to between-facility differences (signal) and about 35% to within-facility differences (noise)—a moderate degree of reliability. However, when looking exclusively at small (defined as <=46) and medium (47-78) facilities, the IURs are substantially lower. Specifically, the IURs ranged from 0.30-0.41 and 0.50-0.56 for small and medium facilities, respectively, over the same time period. As noted earlier, KCP thus believes the specifications must specifically require a minimum sample as identified through the developer's empirical testing.

(cont.)

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5691: VALIDITY. In addition to KCP's concerns about the specifications and the threat to validity of variable capture of transfusion events depending on hospital coding practices, KCP has several concerns about the co-variates (or lack thereof) and risk model.

NQF did not endorse the STrR in 2015, in part because this Standing Committee raised concern that the measure did not adjust for hospital- and physician-related transfusion practices. Physicians independently, or following hospital protocols, make decisions about whether or not to transfuse a specific patient, so it is important to account for the variability these factors create. The revised measure does not incorporate these factors into the risk model, so KCP's concurrence with the Committee's original concern remains.

KCP notes that while the SMR and SHR have been revised to incorporate prevalent co-morbidities into their risk models, the STrR has not been so revised; only incident co-morbidities, derived from the Medical Evidence Form (CMS 2728), are considered. This approach means the STrR risk model only reflects those conditions present upon when the patient initiates dialysis; failure to appropriately account for prevalent co-morbidities is a threat to validity. In the harmonization section, we also note that CMS adjusts for 2728-derived co-morbidities for SHR and SMR differently than it does for the STrR. Finally, as we have noted before, we continue to be concerned about the validity of the 2728 as a data source and urge that the Committee

recommend that CMS assess this matter.

KCP notes that the validity testing yielded an overall c-statistic of 0.65. We are concerned the model will not adequately discriminate performance—particularly that smaller units might look worse than reality. We believe a minimum c-statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

(cont.)

Comment by Lisa McGonigal, MD, MPH
Organization Kidney Care Partners
Comment #5692: HARMONIZATION ISSUES. The new SMR and SHR risk models adjust for each incident co-morbidity (from the 2728) separately, instead of using a "co-morbidity index." The model also approaches diabetes as a single co-morbidity rather than four separate indicators (currently on insulin, on oral medications, without medications, diabetic retinopathy). The STrR has not been similarly revised. KCP believes the Standing Committee should recommend that the developer harmonize the STrR with the other measures so that each incident co-morbidity is examined separately (i.e., unbundled, as compared to the current measure) and diabetes is approached as a single co-morbidity (i.e., bundled, as compared to the current risk model).

The risk models for the groupings used for patient age and duration of ESRD differ among the SMR, SHR, and STrR. For example, the age groups for the SMR is n=3, but for the SHR and STrR the age groupings are the same, but n=6. Similarly, the number of groups for ESRD duration for the SMR (n=4) differs from that for the SHR (n=6). No justification or empirical analyses are offered to justify these differences.

There also are significant inconsistencies in how facility size is defined when assessing reliability for the SMR, SHR, and STrR. Specifically, for the SMR, the definitions were <=45, 46-85, >=86 for the 1-year reliability analyses, but were <=135, 136-305, and >=306 for the 4-year analyses. For the SHR, <=50, 51-87, and >=88 were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as <=46, 47-78, and >=79, respectively. We understand reliability for a given measure depends on sample size, but find the varying demarcations analytically troubling. We posit a more appropriate analytic approach would be to analyze reliability using consistent "bins" of size (i.e., small, medium, and large are consistently defined) and identify the facility size at which reliability for that particular measure can be confidently inferred—and then reflect the minimum size in the actual specifications.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
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: 4/15/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
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
☒ Health 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

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

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.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

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 ≥ 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 (β = 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.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

The 2013 Clinical TEP reviewed a suite of articles related to transfusions in ESRD patients.

1a.8.2. Provide the citation and summary for each piece of evidence.

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 subcriteria 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.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Several changes in the ESRD system are likely to impact anemia management. These include identification of safety concerns associated with aggressive erythropoiesis-stimulating agent (ESA) use, expansion of the ESRD Prospective Payment System bundled payment, and the development of the ESRD Quality Incentive Program. There are concerns that these changes could result in underutilization of ESAs, with lower achieved hemoglobin values that may increase the frequency of red blood cell transfusion in the US chronic dialysis population.
Blood transfusion may be an indicator for underutilization of treatments to increase endogenous red blood cell production (e.g. ESA, iron). In addition, dialysis patients who are eligible for kidney transplant and are transfused risk the development of becoming sensitized to the donor pool thereby making transplant more difficult to accomplish. Blood transfusions carry a small risk of transmitting blood borne infections, development of a transfusion reaction, and using infusion centers or hospitals to transfuse patients is expensive, inconvenient, and could compromise future vascular access.
Monitoring the risk-adjusted transfusion rate at the dialysis facility level, relative to a national standard, allows for detection of treatment patterns in dialysis-related anemia management. This is of particular importance due to FDA guidance regarding minimizing the use of ESAs, and economic incentives to minimize ESA use introduced by Medicare's bundling of payment for ESAs. As providers use less ESAs in an effort to minimize the risks associated with aggressive anemia treatment it becomes more important to monitor for an overreliance on transfusions.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) 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 2011-2014.

2011: 5774 facilities, 1.029 mean STrR, 1.348 Standard Error. Facility percentiles: 0.199 (10th), 0.494 (25th), 0.863 (50th), 1.329 (75th), 1.896 (90th).
2012: 5943 facilities, 1.023 mean STrR, 0.972 Standard Error. Facility percentiles: 0.217 (10th), 0.518 (25th), 0.866 (50th), 1.309 (75th), 1.864 (90th).

2013: 6170 facilities, 1.057 mean STrR, 2.883 Standard Error. Facility percentiles: 0.213 (10th), 0.517 (25th), 0.866 (50th), 1.321 (75th), 1.897 (90th)
2014: 6415 facilities, 1.034 mean STrR, 1.408 Standard Error. Facility percentiles: 0.171 (10th), 0.494 (25th), 0.867 (50th), 1.317 (75th), 1.843 (90th)

Data for the measure are derived from an extensive national ESRD patient database, which is largely derived from the CMS Consolidated Renal Operations in a Web-enabled Network (CROWN), which includes Renal Management Information System (REMIS), and the Standard Information Management System (SIMS) database, Medicare claims, the CMS Medical Evidence Form (Form CMS-2728), transplant data from the Organ Procurement and Transplant Network (OPTN), the Death Notification Form (Form CMS-2746), the Nursing Home Minimum Dataset, the Dialysis Facility Compare (DFC) and the Social Security Death Master File. Information on transfusions is obtained from Medicare Inpatient and Outpatient Claims Standard Analysis Files (SAFs).

The data below show the number of facilities, patients, total count of transfusions and total patient years at risk for each year. Also, we calculate unadjusted or raw transfusion rates per year (defined as total transfusions divided by total patient years at risk).
2011: 5774 facilities, 387097 patients, 67428 total transfusions, 227935.62 Total Patients Years at risk, 29.58 Raw Transfusion Rate per 100 patient years at risk*.
2012: 5943 facilities, 398769 patients, 74444 total transfusions, 234847.09 Total Patients Years at risk, 31.70 Raw Transfusion Rate per 100 patient years at risk*.
2013: 6170 facilities, 415576 patients, 73122 total transfusions, 241082.06 Total Patients Years at risk, 30.33 Raw Transfusion Rate per 100 patient years at risk*.
2014: 6415 facilities, 429241 patients, 69182 total transfusions, 246710.49 Total Patients Years at risk, 28.04 Raw Transfusion Rate per 100 patient years at risk*.

*This analysis includes all facilities for the given year.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) 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 as well as similar parameter estimates for the other covariates. The data below shows the parameter estimates for the race, sex and ethnicity variables included in the model containing the other covariates listed in S.14.

Females: 0.168 estimate, 0.004 standard error, <.0001 p-value.
Native American*: -0.075 estimate, 0.023 standard error, 0.001 p-value.
Asian*: -0.207 estimate, 0.012 standard error, <.0001 p-value.
Black*: -0.046 estimate, 0.005 standard error, <.0001 p-value.
Other Race*: 0.090 estimate, 0.045 standard error, 0.044 p-value.
Hispanic #: -0.181 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 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
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 to the present.
The risks associated with aggressive treatment of anemia of CKD with ESAs have been well documented in KDIGO Anemia Management Guidelines as well as in updated FDA package insert information for ESAs. In contrast, the effect of anemia management paradigms that target to lower hemoglobin levels, and generally use less ESA, on transfusion risk is less well defined. Several clinical interventional trials comparing higher vs. lower hemoglobin targets have shown higher transfusion rates in those patients randomized to lower hemoglobin targets. The importance of these observations is limited by lack of predefined criteria for use of blood transfusion in most studies.
It has been postulated that a national trend toward increased use of transfusions in dialysis patients would adversely affect the supply of blood available for acute injuries and surgical procedures. Lastly, greater exposure to human leukocyte antigens, present in transfused blood, may increase anti-HLA antibodies in kidney transplant candidates, resulting in reduced access to kidney transplantation.
The inverse relationship between achieved hemoglobin and transfusion events has been reported previously for Medicare dialysis patients (Ma 1999) and for non-dialysis CKD patients treated in the Veterans Administration system (Lawler 2010)
Unpublished analyses of Medicare Claims data presented at CMS Technical Expert Panel in May 2012 demonstrate an inverse association between achieved hemoglobin and subsequent transfusion risk using more recent data from 2008-2011.
In early 2012, a highly publicized USRDS study presented at the NKF Clinical meeting reported increased dialysis patient transfusion rates in 2011 compared to 2010.
UM-KECC and Arbor Research collaborators presented an analysis of transfusion events in Medicare dialysis patients from 2009-2011, observing increased transfusions in 2011, although the magnitude of change in transfusion rates was much lower than reported by the USRDS.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Lawler EV, Bradbury BD, Fonda JR, et al. "Transfusion burden among patients with chronic kidney disease and anemia." Clinical journal of the American Society of Nephrology : CJASN (2010) 5:667-72. PMID: 20299366
Ma JZ, Ebben J, Xia H, et al. "Hematocrit level and associated mortality in hemodialysis patients." Journal of the American Society of Nephrology : JASN (1999) 10:610-9. PMID: 10073612

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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_and_Risk_Model.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
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. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
One year

S.6. **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, 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.
Transfusion events in the inpatient setting are counted in the following way. The event is identified by the presence in a Medicare inpatient claim of the appropriate ICD-9 procedure codes (99.03, 99.04), or, value code (37). For inpatient transfusion events that are identified using specific ICD-9 procedure codes (99.03, 99.04), we identify a transfusion event for each transfusion procedure code with a corresponding unique date listed on the inpatient claim, thus allowing determination of multiple transfusion events on inpatient claims with multiple ICD-9 procedure codes present. For inpatient claims with value code (37), we count a single transfusion event regardless of the number of transfusion value codes reported, so that the number of discrete events counted is the same whether the claim value code indicates 1 unit of blood or multiple units of blood. This results in a more conservative estimate of blood transfusion events from inpatient claims with transfusion value codes.

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); or, value code (37). In outpatient claims we count a transfusion event for each HCPCS and corresponding unique revenue center date to determine the number of unique transfusion events.

Therefore multiple corresponding unique dates for revenue center codes will result in multiple transfusions events, while multiple HCPCS codes reported for the same revenue center date are counted as a single transfusion event, regardless of the number of units of blood recorded. For example, a HCPCS indicating 3 pints of blood reported for two different revenue center dates would equal two transfusion events, while a HCPCS indicating 3 pints of blood reported with the same revenue center date would be counted as a single transfusion event. Finally, outpatient claims with a transfusion related value code (37) is counted as one event.

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

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
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.10. **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.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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 Appendix, 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.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The denominator of the "STrR" uses expected transfusions calculated from a Cox model (Cox, 1972) as extended to handle repeated events (Lawless and Nadeau, 1995; Lin et al., 2000; Kalbfleisch and Prentice, 2002). For computational purposes, we adopt a model with piecewise constant baseline rates (e.g. Cook and Lawless, 2007) and computational methodology as developed in Liu, Schaubel and Kalbfleisch (2010). A stage 1 model is first fitted to the national data with piecewise-constant baseline rates stratified by facility; transfusion rates are adjusted for patient age, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, and calendar year. This model allows the baseline transfusion rates to vary between strata (facilities), but assumes that the regression coefficients are the same across all strata; this approach is robust to possible differences between facilities in the patient mix being treated. The linear predictor for each patient based on the regression coefficients in the stage 1 model is used to compute a risk adjustment factor that is then used as an offset in the stage 2 model to estimate the population baseline rate without stratifying facilities.

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).
Coefficients can be found in the attached excel file.

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.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; 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.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(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.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Dialysis Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2979_Testing_form.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:**

Measure Title: Standardized Transfusion Ratio for Dialysis Facilities

Date of Submission: 4/15/2016

Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *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.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/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.

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

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> 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.26*)**	**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 <u>measured entities</u> 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%

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

Table 2. 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

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 3. 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

The following table (Table 4) 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 4. 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

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 5. 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%

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

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 (%)

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. Our approach for calculating IUR is presented in the appendix.

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 6: 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
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

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.

_____

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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.

In May 2012 there was an assessment of the measure's face validity based on polling of a CMS Technical Expert Panel (TEP).

2b2.3. 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).

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).

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.

2b2.4. 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".

_____

2b3. EXCLUSIONS ANALYSIS

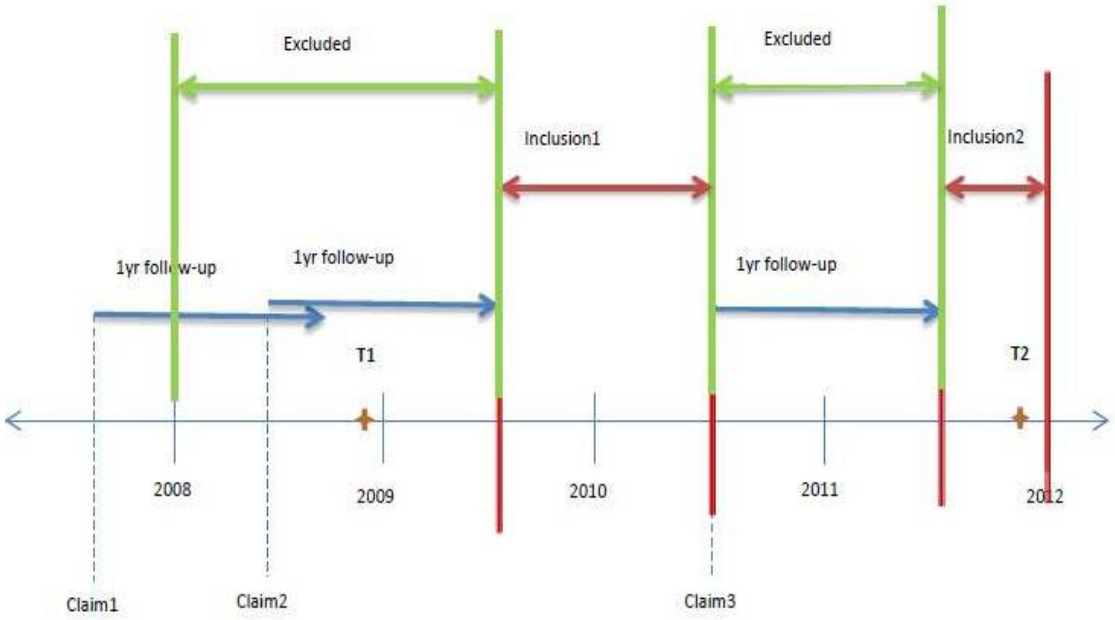
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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.

2b3.2. 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 7: 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%

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

Table 8: Number of patients and percent excluded each year.

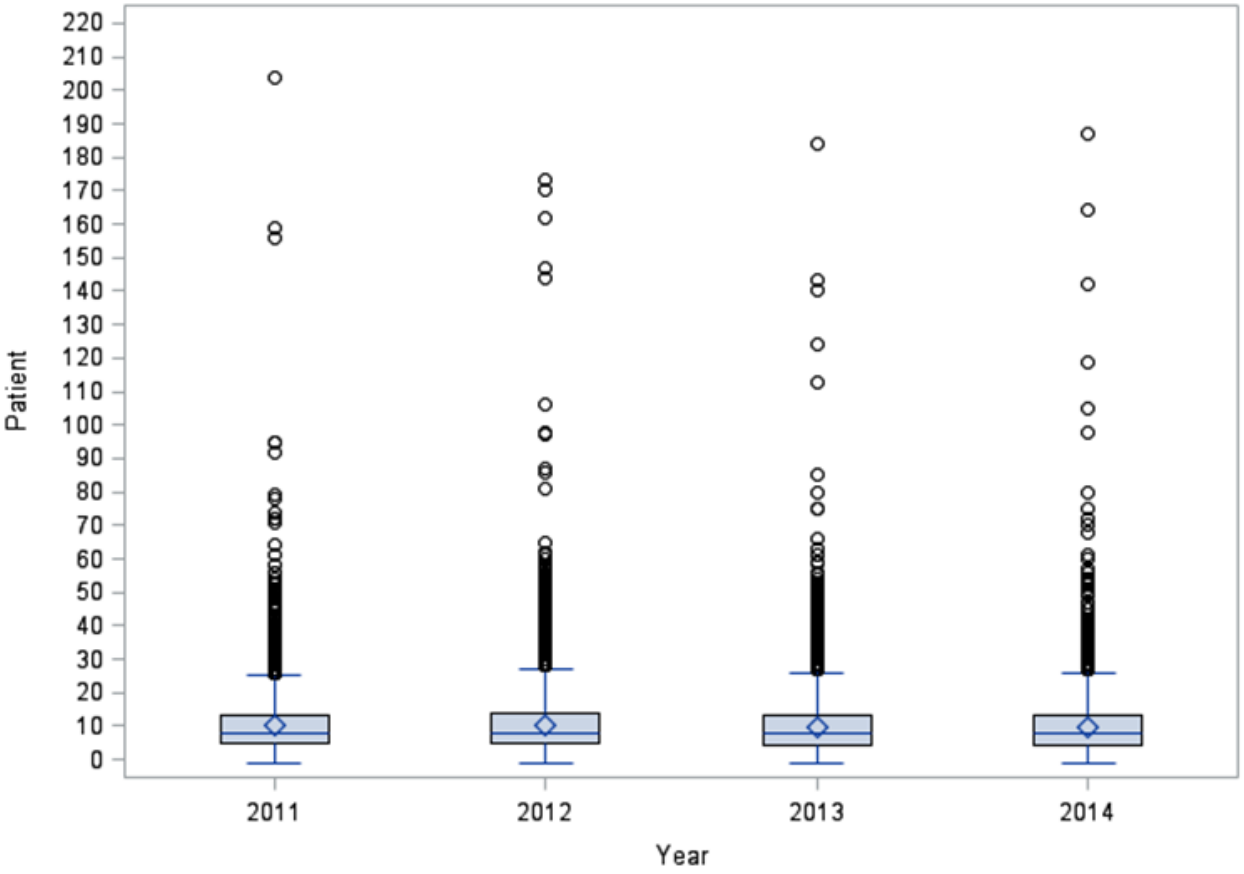
	Number of Patients		
Year	Before Exclusions	After Exclusions	Percent Excluded
2011	452134	387097	14.38%
2012	468592	398769	14.90%
2013	486644	415576	14.60%
2014	503016	429241	14.67%

2b3.3. 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



2b4. 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 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

We included all the standard patient characteristics that are included in the facility level modeling for primary outcomes. We sought input from clinicians and epidemiologists and incorporated claims based risk factors and covariate adjustments recommended by the Technical Expert Panel.

The denominator of the "STrR" is an estimate of the expected number of transfusions at the facility; accounting for each patient's follow-up time and risk factors. The expected number of transfusions is based on the recurrent event analog of Cox regression (Cox, 1972), as developed by Lawless and Nadeau (1995) and Lin et al. (2000); see also Kalbfleisch and Prentice (2002). For computational purposes, we adopt a model with piecewise constant baseline rates (e.g. Cook and Lawless, 2007) and computational methodology as developed in Liu, Schaubel and Kalbfleisch (2010).

The calculation of the STrR is a two-stage approach. At Stage 1, the model is first fitted to the national data with piecewise-constant baseline rates stratified by facility; transfusion rates are adjusted for patient age, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, and calendar year. This model allows the baseline transfusion rates to vary between strata (facilities), but assumes that the regression coefficients are the same across all strata; this approach is robust to possible differences between facilities in the patient mix being treated. The regression parameter estimates from Stage 1 are used to compute the expected number of transfusions for each patient. Stage two involves summing the expected number of transfusions by facility, then computing facility-specific STrRs as the ratio of observed / expected transfusions.

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 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, duration and cause of ESRD are also included:

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

In response to the requirements for NQF's Trial Period for the incorporation of sociodemographic factors into quality measures, we investigated several patient and zip code level indicators of SDS/SES (see list in 1.8). Sociodemographic factors included in the analysis were based on conceptual criteria and availability of data for the analyses. We were able to acquire individual area-level variables included in the Area Deprivation Index (ADI) developed by Singh and colleagues at the University of Wisconsin[1]. These testing results and interpretation are presented in the following sections.

2b4.4a. What were the statistical results of the analyses used to select risk 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.

[1] Singh, GK. Area deprivation and widening inequalities in US mortality, 1969–1998. Am J Public Health. 2003;93(7):1137–1143.

Table 9. Parameter estimates for covariates in STrR model.

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	-0.180	<.0001
60-74	Reference	
75+	0.035	<.0001
BMI		
Log BMI	-0.247	<.0001
BMI missing	0.024	0.190
Calendar year		
2011	Reference	
2012	0.068	<.0001
2013	0.027	<.0001
2014	-0.080	<.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	0.090	<.0001
5+ years	0.044	0.014
Age & diabetes as cause of ESRD interaction term		
0-14		
15-24	0.166	0.090
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

Covariate	Coefficient	P-value
Cancer	0.008	0.541
Diabetes	0.085	<.0001
Peripheral vascular disease	0.134	<.0001
Cerebrovascular disease	0.020	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

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

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

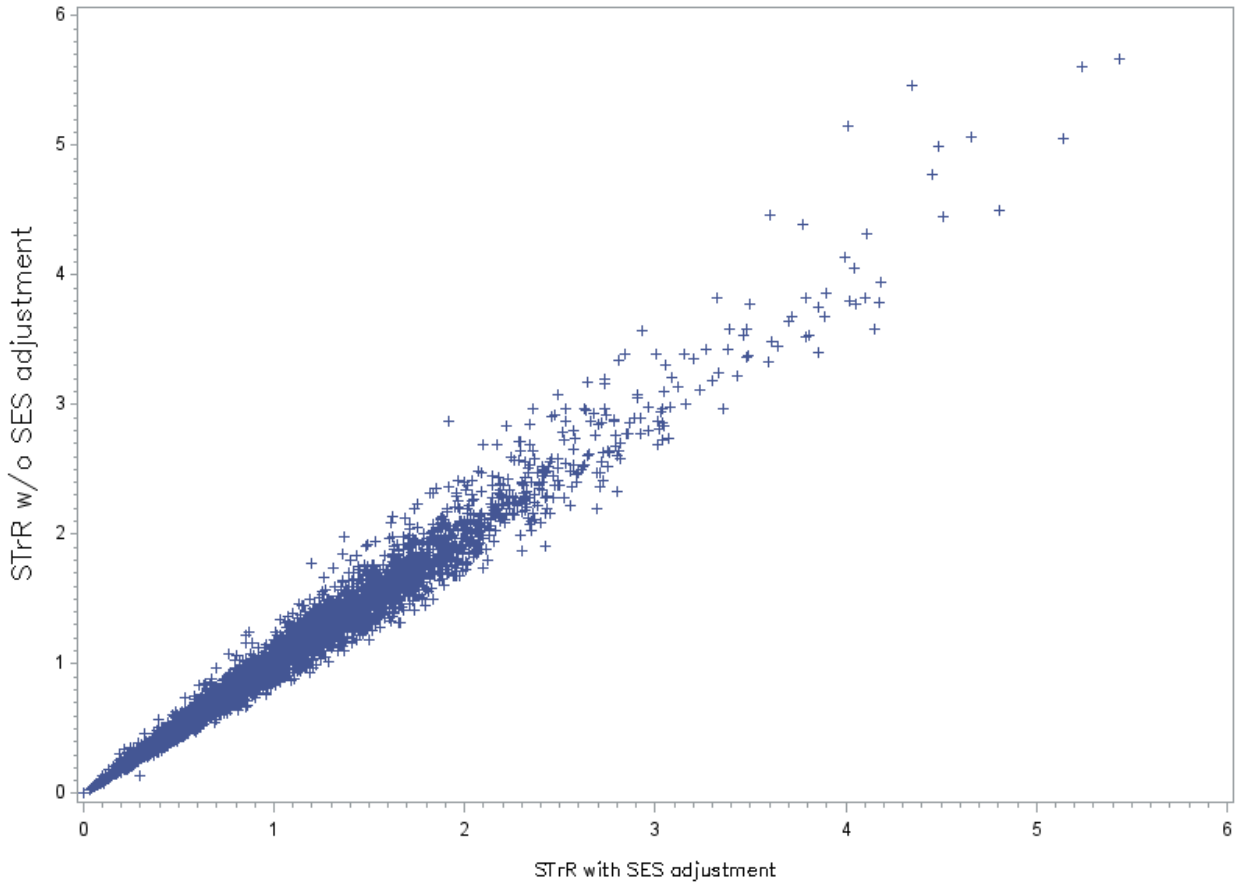
Covariate	Estimates	P-value	Hazard Ratio
Sex: Female	0.163	<.0001	1.177
Race			
White	ref		
Black	-0.048	<.0001	0.953
Asian/Pacific Islander	-0.180	<.0001	0.835
Native American	-0.044	0.058	0.957
Other	-0.031	0.114	0.970
Hispanic	-0.174	<.0001	0.840
Employment status			
Employed	ref		
Unemployed	0.119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

Covariate	Estimates	P-value	Hazard Ratio
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.

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



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

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

Table 11. 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	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).

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.

2b4.5. 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*)

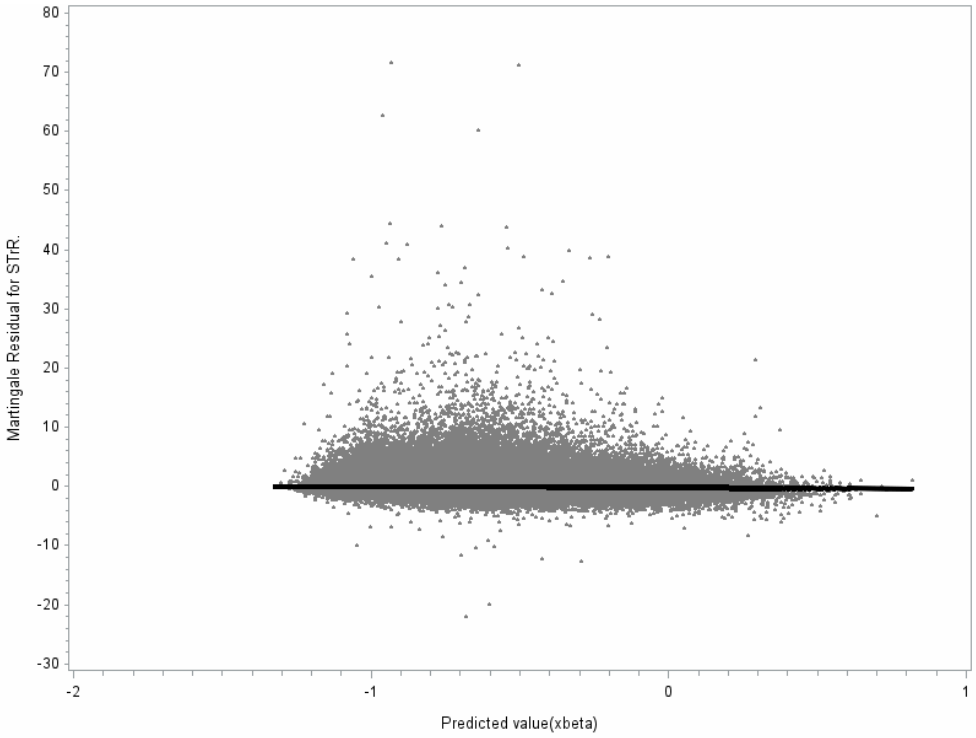
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

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



2b4.6. 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.

2b4.7. 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.

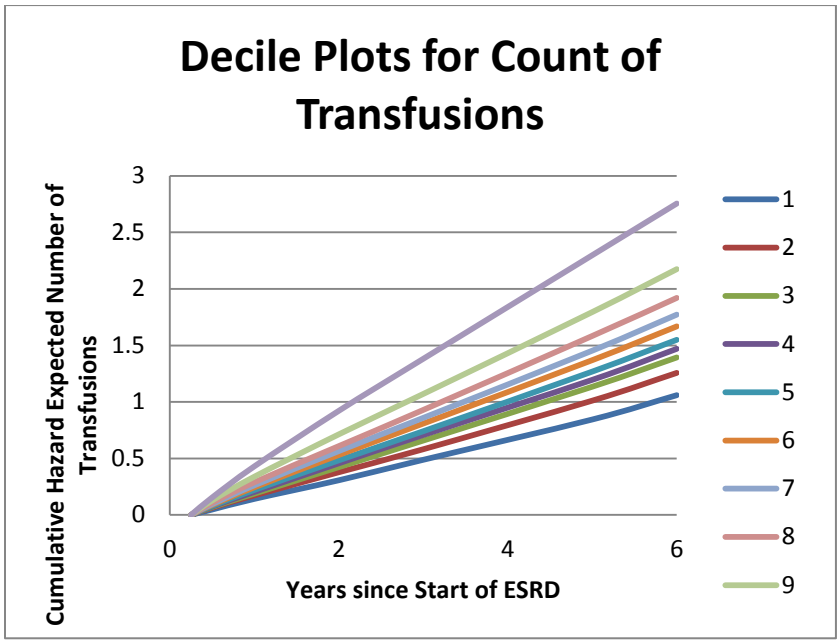
Table 12. Decile-specific observed and expected numbers of transfusions.

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

2b4.8. 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 4: Decile plots for count of transfusions.



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. 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.

2b4.11. 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

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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.

2b5.2. 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 13: Classification of Efron Empirical Null p-value for year 2014*.

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

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

2b5.3. 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.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

N/A

2b7.2. 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*)

N/A

2b7.3. 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*)

N/A

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)

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)*

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.**

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.**

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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

N/A

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.

Planned	Current Use (for current use provide URL)
	Public Reporting 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/

4a.1. For each CURRENT use, checked above, provide:

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

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 eligible for the measure, and have at least 11 patients (due to public reporting requirements). For the most recent DFC report, 5594 facilities were scored on STrR.
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 has been finalized for PY2018.
Geographic area: United States
Number of accountable entities: All Medicare-certified dialysis facilities who are eligible for the measure, and have at least 11 patients (due to public reporting requirements). Number of accountable entities: All Medicare-certified dialysis facilities who are eligible for the measure, and have at least 11 patients (due to public reporting requirements). For the most recent QIP report, 6048 facilities received reports.
Patients included: All patients who meet the requirements to be included in the measure from included facilities.

4a.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

4a.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

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **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.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
The STrR is intended to discourage reliance on unnecessary transfusions as an alternative anemia management response for patients with low achieved hemoglobin. However, a potential unintended consequence of the STrR would be to create an incentive for dialysis facilities to target higher hemoglobin levels. The literature suggests that targeting to hemoglobin concentrations above 12 to 13 grams per deciliter is associated with elevated risk of cardiac events and related mortality. Transfusion avoidance is optimized with achieved hemoglobin in the 10 to 11 grams per deciliter range. Therefore, we believe the potential for unintended consequences is very low with appropriate provider anemia management practices.

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

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 completely harmonized?

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.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, jmsto@med.umich.edu, 410-786-1158-
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:**