NATIONAL QUALITY FORUM
Driving measurable health improvements together

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 #: 3659

Measure Title: Standardized Fistula Rate for Incident Patients

Measure Steward: Centers for Medicare & Medicaid Services

Brief Description of Measure: Adjusted percentage of adult incident hemodialysis patient-months using an autogenous arteriovenous fistula (AVF) as the sole means of vascular access.

The Standardized Fistula Rate (SFR) for Incident Patients is based on the prior SFR (NQF #2977) that included both incident and prevalent patients. This measure was initially endorsed in 2016, but as part of measure maintenance review by the NQF Standing Committee in 2020, concerns were raised about the strength of evidence supporting the prior measure. Namely, recent updates to the KDOQI guidelines downgraded the evidence supporting fistula as the preferred access type and instead focus on catheter avoidance and developing an individualized ESKD Lifeplan. However, the guidelines do suggest that under favorable circumstances an AV fistula is preferred to an AV graft in incident patients due to fewer long-term vascular access events. Given that over 80% of incident dialysis patients begin treatment with a tunneled catheter, and that 12 months after dialysis initiation AV fistula rates exceed 60%, the incident SFR was developed to focus on the subset of dialysis patients that the evidence suggests may benefit the most during a time of intense vascular access creation. Specifically, blood stream infection rates are the lowest in incident patients with AV fistula compared to long-term catheters. Therefore the goal of this new measure is to evaluate facility performance in increasing fistula use in the incident population in order to reduce the heightened risks patients face due to bacteremia and infection related hospitalizations.

Developer Rationale: The Standardized Fistula Rate (SFR) for Incident Patients is based on the prior SFR (NQF #2977) that included both incident and prevalent patients. That measure was initially endorsed in 2016, but as part of a measure maintenance review by the NQF Standing Committee in 2020, concerns were raised about the strength of evidence supporting the prior measure. Namely, recent updates to the KDOQI guidelines downgraded the evidence supporting fistula as the preferred access type and instead focused on catheter avoidance and developing an individualized ESKD Lifeplan. However, the guidelines do suggest that under favorable circumstances an AV fistula is preferred to an AV graft in incident patients due to fewer long-term vascular access events. Given that over 80% of incident dialysis patients begin treatment with a tunneled catheter, and that 12 months after dialysis initiation AV fistula rates exceed 60%, the incident SFR was developed to focus on the subset of dialysis patients that the evidence suggests may benefit the most during a time of intense vascular access creation. Specifically, blood stream infection rates are the lowest in incident patients with AV fistula compared to long-term catheters. Therefore, the goal of this new measure is to evaluate facility performance in increasing fistula use in the incident population in order to reduce the heightened risks patients face due to bacteremia and infection-related hospitalizations.

By focusing on the incident population, we have addressed two other concerns that were raised by the NQF Standing Committee in 2020 with regards to NQF #2977 that included both incident and prevalent patients.

- Exhausted Vascular Access Options: One difficulty with the prior SFR measure was the inability to account for patients that had extensive dialysis exposure and had multiple failed vascular accesses such that they were deemed catheter dependent. While there was widespread agreement that these patients should be excluded from a fistula measure, there has been no way to operationalize that exclusion criteria and no consensus was reached by our 2015 TEP as to how best to do so. This measure, by focusing on patients in their first 12 months of dialysis, avoids the problem of exhausted vascular access since this is typically not encountered in such a short time span.
- Performance Gap: During the consideration of NQF #2977, the performance gap was considered to be negligible since the national average for the standardized fistula rate was at 63% (even though there were, and continue to be, significant disparities as well as variation between providers). By focusing on incident patients where the catheter rate is dramatically higher than seen in prevalent patients, there is more room for improvement in placing AVF in patients who are deemed appropriate candidates. In addition, as is noted in Section 1b.02 below, there is marked variation between facilities in AVF creation during the first year of dialysis.

With a focus on the incident ESRD dialysis population, this measure addresses an important gap in quality outcomes, namely that of catheter-related blood stream infections. 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 approach is supported by the recent KDOQI updated guidelines which stress a patient centered approach to vascular access.

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. 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 (incident SFR, long-term catheter rate) reflects consensus best practice, and supports maintenance of the gains in vascular access success achieved over the last decade.

Numerator Statement: The numerator is the adjusted count of adult incident patient-months using an AVF as the sole means of vascular access as of the last hemodialysis treatment session of the month.

sp.14. Denominator Statement: All patient-months for 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) and became ESRD within the prior 12 months for the entire reporting month at the same facility.

sp.16. Denominator Exclusions:

Exclusions that are implicit in the denominator definition include:

- Patient-months after 12 months of starting ESRD
- Pediatric patients (<18 years old)
- Patients-months 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

The denominator is defined at the patient level not facility level. The reason this rule is applied is to comport with how measures are implemented for public reporting. Due to small cell size and potentially identifiable data, facilities with <11 patients do not receive a score.

As stated in the measure description and rationale, this is a measure of incident patients only. Dialysis patients in their first 12 months of ESRD are more likely to be using a catheter for vascular access and in turn are at higher risk for CVC related infections. The measure focus is on the first 12 months of dialysis since this is the most active time of vascular access creation and where the potential benefit is greatest relative to treatment with a CVC.

Patient attribution to facilities is already described – see SP15: "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."

When a patient is not treated in a single facility for a span of 30 days (for instance, if there were two facility transfers within 30 days of each other), we do not attribute that patient to any facility for that month. Therefore, transient treatment at a facility due to either travel or a temporary clinical condition do not impact the fistula rate of that facility.

Patients with a catheter (of any duration) AND one or more of the limited life expectancy exclusions are excluded from the denominator.

Measure Type: Outcome: Intermediate Clinical Outcome

Data Source: Claims, Registry Data

Level of Analysis: Facility

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a ***structure, process or intermediate outcome*** measure are 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. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a new intermediate clinical outcome measure at the facility level of analysis that focuses on facility percentage of patients in their first year of dialysis with an arteriovenous (AV) fistula; specifically, the measure evaluates facility performance in increasing fistula use in the incident population in order to reduce the heightened risks patients face due to bacteremia and infection related hospitalizations.
- The developer provides a logic model that depicts improvement in blood stream infections due to increased use of arteriovenous (AV) fistulae. The logic model presents a path for facilities to identify and evaluate new dialysis patients with a tunneled catheter for placement of an AV fistula.

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

Summary:

- The developer noted that a convened 2015 Technical Expert Panel (TEP) agreed that AVFs are the preferred access for most dialysis patients.
- In addition, the developer noted The National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI) Vascular Access Guidelines suggest:
    - A functioning AVF is preferred to an arteriovenous graft (AVG), due to fewer long-term vascular access events such as thrombosis and loss of primary patency. (Conditional Recommendation, Low Quality of Evidence)
    - Most patients starting dialysis with a central venous catheter should convert to either an AVF or AVG, to reduce their risk of infection/bacteremia, infection-related hospitalizations, and adverse consequences. (Conditional Recommendation, Very Low-Moderate Quality of Evidence)
- The developer cited that there is 1.78 increased odds of starting dialysis with an AVF versus a 0.51 odds of starting dialysis with a central venous catheter if kidney disease patient education is provided. The developer added that creation of a vascular access coordinator program decreased rates of central venous catheter rates from 45% to 8% and that there is variation in vascular access choice per surgeon, with some surgeons more than twice as likely to create an AVF as others.

Exception to evidence

- N/A

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 – The measure does not assess an outcome → Box 3 The measure assesses an intermediate outcome → Box 4 → The measure indicates a systematic review/summary of quantity, quality, and consistency (QQC) → Box 5 The systematic review indicates high quality evidence that benefits clearly outweigh undesirable effects

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

1b. Gap in Care/Opportunity for Improvement and Disparities

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

- The developer presented an analysis of descriptive statistics of standardized fistula rates from 2018-2019.

- In 2019, the mean value was 41.4 percent. The interquartile range of 17 percentage points 49.9 percent 75th, 32.9 percent 25th), with the bottom quartile of dialysis group practices having 19.7 percent of incident patients using an AVF and the top quartile of dialysis group practices having 163.2 percent of incident patients using an AVF.

Disparities

- The developer provided descriptive statistics of Standardized Fistula Rate (SFR) percentage data by race, ethnicity, sex, employment status, insurance status and area deprivation index (ADI) and determined that there are clinically meaningful differences in mean percentage of SFR based on sex, race, ethnicity, and SES as measure by the ADI.
  - Mean SFR was higher for males (46.2 percent) than females (35.1 percent),
  - Mean SFR was highest for those identified as "Other" (45.1 percent), followed by those who identify as White (43.4 percent) and those who identify as Black (36.4 percent).
  - Hispanics (45.1 percent) had a higher mean SFR than non-Hispanics (40.8 percent)
  - Non-dual eligible persons (41.9 percent) presented a slightly higher SFR than dual-eligible persons (40.8 percent).
  - The lower ADIs within the quintile of 0 – 33.3 presented a higher mean SFR (43.3 percent) than ADI values within the quintile of 76.6 – 100 (39 percent).
- Based on the variation among the descriptive statistics, the developer suggests potential disparities in AVF use among the incident ESRD dialysis population.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

1a. Evidence

- I am not seeing an evidence review. The developers note the failure of the committee to endorse a measure 2977.
- The SFR standardized fistula rate for incident patients is based on the prior SFR #2977 that included both incident and prevalent patients. Measure was initially endorsed in 2016 but as part of the measure maintenance review by the NQF Standing Committee in 2020 concerns were raised about the strength of the evidence supporting the prior measure. Recent updates to the KDOQI guidelines downgraded evidence supporting fistula as the preferred access type and instead focus on catheter avoidance and develop an individualized ESKD lifeplan. 80% of incident dialysis patients begin treatment with a tunneled catheter and that 12 months after dialysis initiation AV fistula rates exceed 60%, the incident SFR was developed to focus on the subset of dialysis patients hat the evidence suggests may benefit the most during a time of intense vascular access creation.
- Evidence submitted includes 2015 UM-KECC TEP, KDOQI Vascular Access Guidelines (published 2020), and review of peer-reviewed research. KDOQI guideline cited includes the following: "KDOQI suggests that most incident HD patients starting dialysis with a CVC should convert to either an AVF or AVG, if possible, to reduce their risk of infection/bacteremia, infection-related hospitalizations, and adverse consequences. (Conditional Recommendation, Very Low-Moderate Quality of Evidence)."
- Evidence is strong for the association between AVF use for HD and lower adverse outcomes.
- Data provided showing several patient benefits of AVF vs catheter.

- The evidence presented is tangential to the intermediate outcome being measured. The all-patient iteration of the Standardized Fistula Rate Measure (SFR, previously NQF 2977) lost NQF endorsement in 2020 secondary to KDOQI's downgrading of the evidence supporting fistulas as the preferred access type, in favor of catheter avoidance and individualized ESKD Lifeplans. To support the premise for this new, incident-only measure, the developer (CMS) now counters that the same guidelines do suggest that under favorable circumstances an AV fistula is preferred to an AV graft specifically in incident patients due to fewer long-term vascular access events (e.g., thrombosis, loss of primary patency, interventions) and because "blood stream infection rates are the lowest in incident patients with AV fistula compared to long-term catheters." However, the KDOQI guideline explicitly indicates there is inadequate evidence to make a recommendation on choice of AV fistula vs AV graft for incident vascular access based on associations with infections; thus, here again, the KDOQI statement focuses on catheter reduction and takes no stance on the superiority of fistulas over grafts in this regard. CMS also indicates that NQF 3659 was developed to focus on the subset of dialysis patients that evidence suggests may benefit the most during a time of intense vascular access creation, noting that while greater than 80% of incident dialysis patients begin treatment with a tunneled catheter, AV fistula rates exceed 60% by twelve months after dialysis initiation. It should be noted that this Standing Committee also rejected the prior, all-patient SFR because we believed the measures was effectively "topped out" at 64% for all patients for whom an AV fistula is clinically appropriate. As the new measure defines an incident patient as one who began maintenance hemodialysis within the prior twelve months, CMS's rationale here is flawed. Rather than supporting the premise of the measure, fistula rates climbing from less than 20% at dialysis initiation to greater than 60% within twelve months supports that dialysis facilities are already placing fistulas in nearly all clinically appropriate incident patients, once under their care, such that by the end of the first year of dialysis the population approaches that "topped out" AV fistula rate identified by this Standing Committee. Thus merely narrowing the target population of the prior, all-patient iteration of the Standardized Fistula Rate Measure does not effectively address the issues that led to its loss of NQF endorsement in 2020. Catheter avoidance remains the appropriate focus for vascular access in both the incident and prevalent dialysis populations; the Standardized Fistula Rate for Incident Patients is an unnecessary solution to a problem already being effectively addressed by the existing, endorsed Long-term Catheter Rate measure (NQF 2978).
- I am not aware of any new investigations that alter this measure's evidence base.
- This is a new intermediate outcome measure. The evidence provided by the developer appears to support the need for this measure in creating quality care.
- I feel this would help patients; however, I would need more information. I am not aware of other studies.
- High
- Evidence demonstrates that patients with AVF have less bloodstream infections than those dialyzed with central venous catheters which directly will reduce morbidity and mortality.
- Strong evidence, especially for incident patients.
- Low- #3659 Facility-level standardized fistula rate for incident patients is a modification of the prior NQF Standardize Fistula Rate measure (NQF 2977). When last reviewed (2020) that measure was not endorsed and as it appeared to be effectively "topped out" at 64%. The current NQF measure has changed the target population from incident and prevalent patients to only patients who have begun maintenance hemodialysis within the prior 12 months. The premise of the measure is that this modification in the target population to "incident only" patients will significantly improve the quality of care for HD patients and lead to a reduction in catheter use and an increase in AVF use. No evidence is presented to support this construct. Instead, the data presented suggest that currently during the first 12 months of dialysis, fistulae use increases to over 60% and only 4.7 % of facilities performed "worse than expected." The evidence to support this as a measure that will substantially improve

facility quality of care is weak, and the potential for unintended consequences exist. (ie repetitive trials of AV fistula creation in patients in which an AV graft may be preferable).

- True
- High
- The KDOQI Practice Guidelines have not changed in regards to low quality of evidence. Patient education, VA coordinators, and surgeon variations vary in scope. How is that accounted for?
- All newer studies support the measure.
- The data supports improvement in AVF rate but it is not clear how much impact can be had on incident AVF rate.
- The measure focuses on facility percentage of patients in their first year of dialysis with an AVF. AVF's are preferred access for most HD patients and the developer cited KDOQI guidelines, 2015 technical expert panel as well as peer reviewed literature. The developer also noted that most patients start HD with CVC. Thus, they noted a relationship to outcomes of decreased rate of risks to patients due to bacteremia and infection related hospitalizations. Thus, the evidence does support the measured focus and I would rate it high.

1b. Gap in Care/Opportunity for Improvement and Disparities

- 4.67% of facilities performed worse than expected. Disparities are evident.
- Focusing on incident patients where the catheter rate is dramatically higher than seen in prevalent patients, there is more room for improvement in placing AVF in patients who are deemed appropriate candidates. There is marked variation between facilities in AVF creation during the first year of dialysis. With focus on the incident ESRD dialysis population the measure addresses an important gap in quality outcomes namely that of catheter related blood stream infections.
- Performance gap analysis included examination of 6664 facilities (2018 – 2019): median SFR 41.6% (lower quartile 32.9%, upper quartile 49.9%). Gap and variation in performance demonstrated. Differences in SFR performance based on sex, race, ethnicity, employment, dual eligibility, and ADI.
- There appears to be gaps between facilities, potentially some disparities amongst demographic groups.
- Mean current standardized fistula rate for index pts at 44% with significant room for improvement and data on effect of sex, race, ethnicity, and SES on proportion of pts with fistula.
- As noted in Question 3, CMS presents evidence that greater than 80% of incident dialysis patients begin treatment with a tunneled catheter, but AV fistula rates exceed 60% by twelve months after dialysis initiation, approaching the "topped out" value of 64% for all patients previously identified by this Standing Committee.
- Current data on the performance measure was provided. Population subgroup information was displayed. Disparities are demonstrated in SFR and ADI by gender and race.
- Interesting disparities data. That allow signifies a gap in performance and a need to measure.
- I believe this would help patients; however, I would need more information.
- Moderate
- Since looking at incident patients, more likely to identify performance gap as newer ESKD patients may be less likely to have an AVF created prior to initiation.
- Significant gap, worse for Black patients.
- Moderate - some 80 between male and female black and Caucasian. Her ADI seem to have a higher standardize fistula rate than did individuals with higher ADI value.
- True
- For incident patients, yes.

- Evidence demonstrates disparities.
- Current performance data was provided. A moderate gap in care and moderate disparities still exist. The data provided showed disparities for subgroups without an insurance provider. Yes, it demonstrates disparities in care.
- The data show a gap in current practice vs optimal practice but do not show that the measure will lead us to an improved place.
- Performance gap data was provided where it was evident that there was a disparity of AVF use among different ESRD population. This gap does warrant a national performance measure. Rate as high.

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: Dave Nerenz; Matt Austin; Zhenqiu Lin; Joseph Kunisch; Patrick Romano; Daniel Deutscher; John Bott; Ron Walters; Eugene Nuccio; Joseph Hyder (Combined Methods Panel Review)

- The SMP passed on Reliability with a score of: H-3; M-4; L-1; I-2
- The SMP passed on Validity with a score of: H-1; M-7; L-2; I-0

2a. Reliability: Specifications and Testing

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

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.

Specifications:

- Measure specifications are clear and precise.

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
    - Reliability testing was conducted at the accountable entity level using the inter-unit reliability (IUR) with a bootstrap approach, and Profile IUR (PIUR).
    - The developer calculated a IUR value of 0.705, which indicates that 70.5 percent of the variation in the Incident SFR can be attributed to between-facility differences in performance (signal) and 29.5 percent to the within-facility variation (noise).
    - The developer also calculated a PIUR of 0.970. They noted that this value is higher compared to the IUR, indicating the existence of outlier facilities.
    - Facilities with <11 eligible patients were excluded from this calculation.

SMP Summary:

- SMP raised concerns with the use of patient-months as the unit of counting and analysis for both numerator and denominator. SMP members shared that the ability to count one patient up to twelve times in the measure for an entity for a given year creates questions about the independence of the observations that go into calculating the performance rate.

- SMP expressed that reliability statistics may be overestimated if the observations for a given patient are highly correlated with each other.
- SMP expressed additional concern with the measure's ability to identify variation in performance with over 92 percent of facilities classified as "average" / "as expected."

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*
- *The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

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

2b. Validity: Validity testing; exclusions; Risk-Adjustment; Meaningful Differences; comparability; Missing Data

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.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Validity Testing

- Validity testing conducted at the Accountable Entity Level:
    - The developer assessed validity by using a Poisson regression model to examine the association between facility level quintiles of performance scores (during 2018-2019 performance period) and to produce a Standardized Mortality Ratio
    - The developer found that the relative risk of mortality increased as the performance measure quintile decreased from the reference group (Q5) with the highest risk in quintile
        - Results indicated quintile 4, RR=1.02 (95% CI: 1.00, 1.04; p<0.001)
    - Results indicated quintile 3, RR=1.06 (95% CI: 1.04, 1.08; p<0.001)
    - The developer assessed 2018-2019 Standardized Hospitalization Ratio (SHR, NQF #1463). The developer found that the relative risk of hospitalization increased as the performance measure quintile decreased from the reference group (Q5) with the highest risk in quintile 1.
        - Results indicated quintile 4, RR=1.06 (95% CI: 1.05, 1.06; p<0.001)
    - The developer assessed the First Year Standardized Mortality Ratio (SMR) and found that the relative risk of mortality increased as the performance measure quintile decreased from the reference group (Q5) with the highest risk in quintile 1.
        - Results indicated quintile 4, RR=1.08 (95% CI: 1.03, 1.14; p=0.002), quintile 3, RR=1.11 (95% CI: 1.05, 1.16; p<0.001)
    - The developer found that the hospitalization rate decreased as the performance measure quintile increased. Hospitalization rates for quintiles 1 to 5 were 1.06, 0.99, 0.95, 0.93, and 0.87 patient-years respectively (trend test p<0.001).

Exclusions

- The developer identifies the following exclusions, as applied to the denominator:
    - Patients with a catheter that have limited life expectancy, which the developer defines 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 standardized fistula rate with and without the patient-month exclusions.
    - The developer's calculations revealed that the percentage of patients excluded at each facility is not evenly distributed across facilities. Therefore, the developer posits that the exclusion criteria are necessary.
    - The developer added that the exclusion criteria take into account that some facilities treat a higher proportion of patients with limited life expectancy based on the unequal distribution across facilities.
    - The developer's exclusion analysis resulted in a Pearson Correlation Coefficient of 0.996 (p-value <0.0001) between Incident SFRs with and without the exclusion.
        - Based on these findings, the developer suggests that the overall impact of the exclusion on the measure's validity is not substantial since the two are highly correlated.

Risk-Adjustment

- The developer used a multivariate logistic regression risk adjustment model to calculate the incident SFR for facility as an estimate of what the facilities percentage of AVF would equal if the facility's patient mix was equal to the nation as a whole.
- The model uses age, BMI at incidence, nursing home status in the prior 12 months (3 levels), nephrologist's care prior to ESRD, Diabetes as primary case of ESRD, Comorbidities at ESRD incidence (8 categories), indicator for missing 2728 form, indicator for at least one incident comorbidity, and a Medicare indicator.
- C-statistic = 0.748; The Hosmer-Lemeshow test statistic based on deciles of risk is 9.7 with p-value=0.29.
    - The developer advises that the c-statistic and risk decile plot show that the model provides an overall good fit to the data. There is good separation among all 10 groups by risk scores, and the ordering is as predicted by the model.
- The developer tested socioeconomic status SES/SDS factors, including sex, race and ethnicity, Medicare-Medicaid dual eligibility and ADI as social risk factors in the risk adjustment model.
    - Black race, female sex, unemployment at ESRD incidence, and dual-eligible status were all associated with lower odds of having an AVF indicating that patient-level, but not area-level, variables for SDS/SES have some impact on expected performance of the incident SFR.
- The developer noted that patient-level SDS/SES variables are not included as adjustments in the final 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.

Meaningful Differences

- Developer adjusted for case mix and expected variation. This adjustment resulted in:
    - 2.88% of dialysis group practices performed better than expected
    - 4.67% performed worse than expected
    - 92.45% performed as expected
- Developer provides data to support statistically significant differences in performance across facilities based on their adjusted proportion of patient months with a fistula in use.

Missing Data

- The developer identifies patients-months with missing BMI on the CMS 2728 is 0.01% and missing CMS 2728 for Incident SFR is 1.11%. The developer indicated negligible impact of missing data for these elements due to the low frequency of patients with missing data/small percentages.
- The developer identifies patient months with missing vascular access type in the denominator for the Incident SFR is counted in the denominator and is not counted in the numerator, as it is classified as a catheter. The developer included that 3.6% of overall patient months are missing vascular access type.
- The developer states that 32.47 % is the percentage of patient months that the developer was unable to identify. Developer acknowledges this as a general limitation of reliance on FFS Medicare claims and insufficient claims history.

Comparability

- The measure only uses one set of specifications for this measure.

SMP Summary:

- The SMP shared concern with the risk model, specifically that the C-statistics and calibration were based on development data only, with no external validation.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

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

Committee Pre-evaluation Comments:

2a1. Reliability – Specifications

- No concerns.
- Measurements specifications are clean and precise. Reliability testing conducted at the accountable entity level using inter-unit reliability with a bootstrap approach. Conducted at the Accountable Entity Level during IUR. IUR value calculated at 0.705 70.5 percent of the variation in the incident SFR can be attributed to between facility differences in performance and 29.5 to the within facility variation. SMP expressed that reliability statistics may be overestimated if the observations for a given patient are highly correlated with each other. Expressed additional concern with the measure ability to identify variation in performance over 92% of facilities classified as average/as expected.
- Data specifications are clear.
- No concerns.
- Data elements are defined in satisfactory fashion and no concerns about potential for consistent implementation of measure.
- No concerns with the specifications. Data elements and logic/calculation algorithm are clearly defined. Believe the measure could be consistently implemented. It's unclear from the information provided whether the descriptive statistics and disparities data presented in Section 1b (Performance Gap) are for all patients or are limited to incident patients; request clarity from the developer on this issue.

- Elements of the data are clear. All reliability components are precisely described.
- SMP passed on Reliability. I do not have any concerns. Specs are clear.
- I'm not sure this could be tracked accurately.
- Moderate
- No concerns.
- Clearly defined.
- Moderate, agree with comment by SMP that use of Pt -months in calculations means one patient can be counted 12 times for a given year for an entity and raises questions about the independence of the observations going to calculating the performance.
- True
- No concerns.
- How they collect the data is not consistent. They state claims and registry in most areas. In S29 they provide an extensive list of where the data is derived but not discussed extensively.
- Measure is reliable.
- None
- I do not have concerns about the measure being implemented consistently and I believe the science is there for this measure to be reliable. I understand the need to measure patient-months.

2a1. Reliability – Testing

- IUR .705 and PIUR .97. No concerns.
- No
- IUR 0.705 and PIUR 0.970
- No concerns
- IUR of 0.7 with higher PIUR, so expected standards of reliability seem to have been met.
- Yes. While the overall IUR across all facilities is acceptable at 0.705, stratification of reliability scores by facility size was not detailed. Because of this, it's impossible to determine how widely reliability varies across the spectrum of facility sizes. As has often been the case with other CMS standardized measures, reliability for small facilities might be substantially lower than the overall IUR, effectively rendering the metric meaningless for use in performance measurement in this group of providers. Request CMS provide data demonstrating reliability for all facilities by detailing IURs by facility size.
- There are no reliability concerns.
- There is some discussion if I interpret this right about variability among facilities in reporting. the value seems high especially if the specs are said to be clear. Maybe I'm not interpreting that correctly.
- Yes.
- No
- No concerns.
- No
- See above
- Appropriate
- No concerns.
- No issue with reliability testing.
- No
- None
- I do not have any concerns about the reliability of the measure.

2b1. Validity – Testing

- Some of the quintiles had very comparable results depending on the comparator.
- No
- Validity testing included examination of SMR, SHR, first-year SMR, all-cause hospitalization and vascular access related infection hospitalization rate.
- No concerns.
- Seems to be satisfactory presentation of evidence that lower SFR related to higher mortality and hospitalization rates.
- No additional concerns beyond the SMP's concern that the risk model C-statistics and calibration were based on development data only, with no external validation.
- There are no significant concerns with validity testing.
- The missing data summary is of some concerns. it seems this would affect the validity of the measure. Interested to hear what others think. It looks like the SMP did not have concern.
- Yes
- No
- No concerns.
- No
- The data are clear: Both fistula and AV graft are superior to catheter. The developer demonstrated that the relative risk of mortality increases the performance measure decreased from the reference group.
- Appropriate
- No
- Face and empirical testing were performed with established outcome measures.
- No concerns
- None
- I believe the risk adjustments and exclusions are appropriate. Given that the developed advised that the C statistic and risk decile plot showed a good fit to the data- I do not have any concerns.

2b2-2b3. Threats to Validity (Exclusions, Risk Adjustment)

- Risk adjustments were typical
- Risk adjustment – developer used a multivariate logistic regression risk adjustment model to calculate the incident SFR for facility as an estimate of what the facilities percentage of AVF avf would equal if the facilities patient mix was equal to the nation as a whole. The model uses age, BMI at incidence nursing home status in the prior 12 months nephrologists care prior to ESRD, diabetes as primary case of ESRD. All variables at start of care.
- Risk-adjustment includes age, BMI at incidence, NH residence (last 12 months), nephrology care prior to ESRD, diabetes as cause of ESRD, incident co-morbidities, and Medicare indicator (6 months FFS or 1 month MA in prior 12 months).
- Exclusions are appropriate, but they also depend on accurate and not-missing information.
- No risk appreciated.
- No concerns with exclusions. The risk model appears to fit well, with a c-statistic of 0.748; however, the SMP's concerns (and unclear consensus) on the lack of inclusion of social risk variables in the final model are noted. A discussion among Standing Committee members would be helpful to further elucidate the appropriate approach to this issue.
- There is an acceptable risk adjustment strategy implemented in the measure.

- It appears the developer took into account SES and SDS factors and ADI in risk adjustment
- I would need more information.
- No
- No concern
- No significant threats.
- N/A
- True
- My only concern remains that patient choice is not ever considered for this measure. Exclusions include malignancy, short life expectancy, etc. However, informed patients have the right to choose not to have a fistula constructed, and this is never considered in this metric.
- Concern with 32% of patient months unable to identify. Claims data is not always accurate! no external validation.
- Exclusions are consistent. Risk adjustments appropriately developed.
- Adjustments are appropriate.
- Exclusions are consistent as well as risk adjustment. No concerns.

2b4-2b7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)

- 32% of patient months had indeterminate life expectancy exclusion conditions.
- Calculations revealed that the percentage of patients excluded at each facility is not evenly distributed across facilities. Therefore exclusion criteria is necessary.
- 32.95% of patient months missing information on limited life expectancy exclusion conditions.
- Missing data may be significant. If risk adjusting, missing co-morbidities may also have significant impact.
- 92.5% of facilities now perform as expected and nearly 3% perform better than expected, so would need to see national rate of SFR start moving upward to likely help distinguish between facilities now in the broad swathe of doing as expected (which may not be good).
- Scores differentiated as “as expected,” “better than expected,” and “worse than expected.” No concerns with approach. Missing data unclear. Request additional information/clarity around the developer’s note that it was unable to identify 32.47% of patient-months. While this may be a general limitation of reliance on FFS Medicare claims and insufficient claims history, it raises concerns about the potential impact of these missing patient-months on measure validity.
- Missing data in the numerator may pose a threat to validity.
- Possibly
- Yes
- No
- Missing data can be a threat to validity of this measure as it will not reflect actual meaningful data
- Meaningful differences identified.
- The risk model, C-statistics and calibration were based on development data only, with no external validation as noted by the SMP. The number of patients lost through Medicare managed care models may affect the Validity of the data.
- Yes
- No concerns
- No concerns.
- Minimal threat to validity. Comparable performance scores between studies were met. No missing data.

- Valid
- The developer provided data to support meaningful difference. Missing data incident is low. I do have some concerns that 32.47% is the percentage of patient months that the developer was unable to identify- but as noted this is a limitation of reliance on Medicare claims.

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.

- The developer indicates that the data elements are generated or collected and used by healthcare personnel during the provision of care and are coded by someone other than person obtaining original information.
- The developer states that the measure relies on data elements that are defined in a combination of electronic sources and cites no difficulties in the collection of data.

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:

3. Feasibility

- No data given
- Data elements are generated or collected and used by healthcare personnel during the provision of care. Further, data elements are coded by someone other than the person obtaining original information. The measure relies on data elements that are defined in a combination of electronic sources.
- No concerns about feasibility.
- Seemingly feasible.
- Seems very feasible
- No concerns with feasibility for this measure.
- Feasibility is exceptional for this measure.
- No concerns noted.
- I would need more information.
- High
- No concerns.
- All data routinely generated.
- Moderate
- Appropriate

- High feasibility
- All the data can be obtained. There are limitations with inaccurate and missing data.
- Feasibility is acceptable. No concerns about data collection.
- Feasible
- Data elements are routinely generated by healthcare personnel. No concerns for feasibility.

Criterion 4: Use and Usability

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

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

4a.1. 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.

Current uses of the measure

Publicly reported? ☐ Yes ☒ No

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

Planned use in an accountability program? ☒ Yes ☐ No ☐ NA

Accountability program details

- The developer plans to use the measure in public reporting and in a quality payment program.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer reports that facility-level results have not been disseminated to those being measured as part of the development process.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*
- *How have performance results, data, and assistance with interpretation been provided to those being measured or other users during development or implementation?*
- *Has an effective process been implemented for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.?*

Preliminary rating for Use: ☒ Pass ☐ No Pass

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

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

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. The developer advises that CMS currently anticipates implementation of this measure. Once implemented facility performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in incident fistula rates.

4b2. Benefits vs. harms. 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).

Unexpected findings (positive or negative) during implementation

- Developer did not report any unexpected findings as the measure is not implemented yet

Potential harms

- Developer did not report any potential harms as the measure is not implemented yet

***Questions for the Committee**:*

- *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:

4a. Use

- New measure
- Yes, those being measure has been given performance data as well as assistance with interpreting results and data 2. Those being measure and other users have been given an opportunity to provide feedback on measure performance or implementation; 3. Feedback has been considered when changes are incorporated into the measure.
- Potential future use includes QIP and DFC.
- Assuming facility level reporting.
- Plan for implementation seems apparent.
- This is a new measure, not yet being publicly reported. It is intended for use in the ESRD QIP, thus results will be disclosed and available to the broader public if the measure is endorsed and implemented. The developer reports that results have not been disseminated to those being measured as part of the development process.
- There is an acceptable plan to use this measure in an accountability program.

- Facility level results have not been shared. I would assume improvements/enhancements could be learned if the information was shared. Currently the measure is not reported publicly or in any programs.
- I would need more information.
- Moderate
- Not publicly reported or used in accountability program but planned to be used in accountability program. Feedback is essential for programs to improve. Need to develop method for feedback and opportunity to provide feedback.
- Not yet implemented
- Westchester Healthcare Corporation
- Appropriate
- No concerns
- Not currently disseminated publicly with no formal feedback.
- The measure is widely used and accepted.
- Not reported publicly; role of feedback unclear.
- The measures are not publicly reported but can be used to make healthcare facilities accountable.

4b. Usability

- Not answered
- The measure is not yet implemented so improvement could not be evaluated. Benefits of the performance measure in facilitating progress toward achieving high quality efficient health care for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations.
- Measure revised to address/minimize unintended consequences.
- Despite risk adjustments (and the stated exclusions), there is still the likelihood that this measure may push providers to get AVFs in patients where it would be less appropriate and therefore constitute some degree of harm.
- Rationale for measure seems reasonably and cogently presented with no apparent harms/adverse consequences from measure implementation expected.
- The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. Once implemented, facility performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in incident fistula rates. The developer did not provide an assessment of benefits vs. harms. A concern with this measure, however, is the potential for focused pursuit of AV fistula placement to meet performance expectations. This is inconsistent with current guidelines, which instead support a goal of catheter reduction and individualized care.
- There were no harms detected in this measure.
- Not aware of any unintended consequences.
- I would hope this would not have unintended consequences, but I would need more information.
- Moderate
- Unintended consequences is that dialysis units may not be willing to accept patients who do not have an established AVF; may result in cherry picking, particularly if no grace period to work on getting AVF. Some patients may present for nephrology care late and therefore not have the time to prepare for AVF access.
- Not yet implemented
- Low see above

- Appropriate
- I remain concerned with the potential harm to coerce patients who choose no fistula surgery to have unwanted procedures.
- The developers do provide rationale for improvement in incidence patients. No unintended consequences noted.
- The measure is usable and will improve lives of CKD5 patients.
- Not clear
- In the future the measure can be used to further quality and accountability of the facility. There are not unintended consequences.

Criterion 5: Related and Competing Measures

Related measures

- NQF #2594 Optimal End Stage Renal Disease (ESRD) Starts

Harmonization

- Developer provides explanation on distinctions of measure 2594 and asserts that there is a fundamental difference in the measure target populations, setting, and intent that cannot be harmonized.
    - Measure #2594 is not directed toward dialysis facilities
    - Measure #2594 setting focus addresses a different provider type which falls outside the purview of measures evaluating dialysis facility performance on fistula use

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- Not answered
- No
- Not if this measure is not at a dialysis-level of reporting
- N/A
- No concerns.
- To my knowledge, there are no related or competing measures.
- Measure 2594 is pointed out as a similar measure, but nuances are explained that support the inclusion of measure 3659 in a measure set.
- I am unaware of competing measures.
- 2594 if applied to dialysis providers
- N/A
- #2594 related but different targets for improvement. These are complementary efforts.
- No
- True
- None
- None
- No competing measures.
- Yes. 2594

- The measure does not appear to be competing as this addresses dialysis facility.

Public and NQF Member Comments (Submitted as of June 7, 2022)

Member Expression of Support

- Of the one NQF member who has submitted an expression of support, none expressed "support" and one expressed "do not support" for the measure.

Comments

Comment 1 by: David White, American Society of Nephrology; Submitted by David White

TO: National Quality Forum Renal Standing Committee

FR: Tod Ibrahim, Executive Vice President, the American Society of Nephrology

DA: June 7, 2022

RE: Public Comment: Spring 2022 Renal Measures

Dear Members of the National Quality Forum Renal Standing Committee On behalf of the more than 37,000,000 Americans living with kidney diseases and the 21,000 nephrologists, scientists, and other kidney health care professionals who are members of the American Society of Nephrology (ASN), thank you for the opportunity to offer commentary on the five proposed transplantation, vascular access, and modality education measures put forth by the Centers for Medicare and Medicaid Services (CMS)/University of Michigan Kidney Epidemiology and Cost Center (UM-KECC): • Facility-Level Standardized Modality Switch Ratio for Incident Dialysis Patients (SMoSR) • Facility-Level Standardized Fistula Rate for Incident Patients (ISFR) • Practitioner/Group-Level First Year Standard Waitlist Ratio (FYSWR) • Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted (PPPW) • Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW) Based on our review, ASN is concerned by several aspects of the measures and offers comment on all five measures submitted to NQF: • Focus on incident maintenance dialysis populations with "stand alone" measures that are independent of measures targeting patients in other stages of kidney diseases such as non-dialysis advanced chronic kidney disease and prevalent dialysis. This siloed focus disadvantages kidney care providers who have provided high quality care for people with advanced CKD, including referral for home dialysis and pre-emptive transplantation and penalizes dialysis providers who assume care of individuals with insufficient care prior to dialysis initiation • Reliance on CMS-2728 data (End Stage Renal Disease Medical Evidence Report Medicare Entitlement and/or Patient Registration) for any risk adjustment including transplant measures • Attribution of measures to dialysis facilities • Lack of adjustment for variables that are critical for patient equity, such as social determinants of health • Focus on dialysis unit-specific measures, without consideration of advanced CKD care and nephrologist-led care Below are comments about the specific measures: Facility-Level Standardized Fistula Rate for Incident Patients (ISFR) ASN agrees that vascular access is an important clinical consideration for patients and supports that hypothesis that some facilities are better than other facilities at optimizing the longevity of hemodialysis fistulas and grafts as well as at facilitating creation of fistulas and grafts. ASN also continues its support of CMS's Long-Term Catheter Rate Measure (NQF #2978) in the ESRD QIP to maintain prevalent central venous catheter use at a small portion of the dialysis population. However, ASN does not believe that narrowing the target population of the prior, all-patient iteration of the Standardized Fistula Rate Measure (SFR, previously NQF #2977) to incident dialysis patients makes for an appropriate metric or that this change addresses the issues that led to its loss of NQF endorsement in 2020. Inherently, the proposed fistula measure is unchanged from the prevalent measure, applying the existing measure to an incident population. ASN does believe attributing performance on this measure to the dialysis facility is appropriate. As a nephrologists' society, ASN considers optimizing vascular access

among incident dialysis patients an appropriate focus for a measure for physicians and physician groups, but the proposed measure is misdirected at dialysis facilities. A well-thought-out vascular access plan is patient-centered, and clinician led. Dialysis facilities who meet patients for the first time should not be primarily responsible for vascular access plans. Rather, this should be done under the direction of the patient's whole kidney care team, in which the patient and their nephrologist work closely with the providers placing access, such as the surgeon or interventionalist. Of note, there are patients for whom timely AVF placement is not feasible and AV graft (AVG) is a reasonable, safer alternative to a catheter. AVG placement should be considered in the numerator. Finally, this measure encourages dialysis facilities to cherry pick patients with existing arteriovenous fistulas, potentially marginalizing patients with other types of access. This is not patient-centered and is not equitable. ASN appreciates the opportunity to provide comments on the five proposed transplantation, vascular access, and modality education measures under consideration. To discuss the contents of this memorandum, please contact ASN Regulatory and Quality Officer David L. White at dwhite@asn-online.org or call (202) 640-4635.

Comment 2 by: Lisa McGonigal, Kidney Care Partners; Submitted by Lisa McGonigal, Kidney Care Partners

Facility-Level Standardized Fistula Rate for Incident Patients (NQF 3659, CMS): KCP does not support the Standardized Fistula Rate for Incident Patients (Incident SFR) Measure. KCP maintains that vascular access is one of the most important clinical considerations for patients making decisions about dialysis facilities, and we continue our strong support of CMS's Long-Term Catheter Rate Measure (NQF 2978) in the ESRD QIP to reduce catheter use. However, we do not believe that merely narrowing the target population of the prior, all-patient iteration of the Standardized Fistula Rate Measure (SFR, previously NQF 2977) effectively addresses the issues that led to its loss of NQF endorsement in 2020. We note that the SFR's loss of NQF endorsement was precipitated by KDOQI's then-recent downgrading of the evidence supporting fistulas as the preferred access type, in favor of catheter avoidance and individualized ESKD Lifeplans. To support the premise for this new, incident-only measure, CMS now counters that the same guidelines do suggest that under favorable circumstances an AV fistula is preferred to an AV graft in incident patients due to fewer long-term vascular access events (e.g., thrombosis, loss of primary patency, interventions) and because "blood stream infection rates are the lowest in incident patients with AV fistula compared to long-term catheters." We note, however, that the KDOQI guideline explicitly indicates there is inadequate evidence to make a recommendation on choice of AV fistula vs AV graft for incident vascular access based on associations with infections; thus, here again, the KDOQI statement focuses on catheter reduction and takes no stance on the superiority of fistulas over grafts in this regard. CMS also indicates that the Incident SFR was developed to focus on the subset of dialysis patients that evidence suggests may benefit the most during a time of intense vascular access creation, noting that while greater than 80% of incident dialysis patients begin treatment with a tunneled catheter, AV fistula rates exceed 60% by twelve months after dialysis initiation. Here we note that NQF's Renal Standing Committee also rejected the prior SFR because they believed the measures was effectively "topped out" at 64% for all patients for whom an AV fistula is clinically appropriate. As the new measure defines an incident patient as one who began maintenance hemodialysis within the prior twelve months, we believe CMS's logic here is flawed. Rather than supporting the premise of the measure, fistula rates climbing from less than 20% at dialysis initiation to greater than 60% within twelve months supports that dialysis facilities are already placing fistulas in nearly all clinically appropriate new patients, once under their care, such that by the end of the first year of dialysis the population approaches that "topped out" AV fistula rate identified by NQF. We also note that stratification of reliability scores by facility size was not detailed; we are thus unable to discern how widely reliability varies across the spectrum of facility sizes. We are concerned that the reliability for small facilities might be substantially lower than the overall IUR, as has often been the case with other CMS standardized measures. Without evidence to the contrary, KCP is thus concerned the Incident SFR reliability may be unacceptably low for small facilities, effectively rendering the metric meaningless for use in performance measurement in this group of providers. KCP believes it is incumbent on CMS to demonstrate reliability for all facilities by providing data by facility

size. Taking all of the above into consideration, we do not believe limiting the SFR population to incident patients effectively addresses the previously identified issues with the original measure. We maintain that catheter avoidance is the appropriate focus for vascular access in both the incident and prevalent dialysis populations, and we believe the Standardized Fistula Rate for Incident Patients is an unnecessary solution to a problem already being effectively addressed by the existing vascular access measure.

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☒ No**

   Submission document: Items sp.01-sp.30

2. **Briefly summarize any concerns about the measure specifications.**

   Reviewer 1: I am concerned about the use of "patient-months" as the unit of counting and analysis for both numerator and denominator. The specifications are very clear, but the ability to count one patient up to twelve times in the measure for an entity for a given year seems to create questions about the independence of the observations that go into calculating the performance rate - they seem to be not independent. This may or may not matter significantly in assessment of reliability and validity, but it does seem worth discussing within the subgroup and staff.

   Reviewer 2: The form indicates that the data source used to calculate this measure is "Claims", but then in sp.17, they indicate that the treatment modality is derived from a combination of Medicare dialysis claims, Medical Evidence Form, and data from CROWNWeb (is that a registry?).

   Reviewer 3: There are a few areas that should be made more clear: 1. Denominator: Sp14 denominator statement refers to "all patient-months for patients at least 18 years old", however, subsequently the application says "the denominator is defined at patient level not facility level" (page 2 and other places). 2. The unit of analysis seems to be at patient-month level, given than most patients contribute many patient-months, should dependency among patient-months from the same patient be accounted for? 3. Is this a measure specified for one-year or two-year data? Most testing results were based on 2018-2019 (2-year) data. 4. For the risk model, there seemed to be no validation of the risk model. C statistics and calibration were based on development data only, no external validation.

   Reviewer 4: While the measure stewards indicate a higher incidence of blood stream infections related other long-term catheters or a lower incidence with AV fistulas. Why not use rate for BSI as an outcome versus rates of a specific access device.

   Reviewer 7: No concerns. A minor comment; add to the brief description under measure type/data source (page 2) 'Registry' as detailed under SP.28.

   Reviewer 8: No concerns.

   Reviewer 9: I think as I understand it, the access type the on the last day of a given month is what contributes to being eligible or not for inclusion. Denominator is at the patient level but must have been treated at the same facility for the entire month to be assigned to that facility for the reporting month.

   Reviewer 11: Some of the Reliability testing items seem to be missing (e.g., Type of Measure) from provided documentation. Denominator requirement for ESRD designation within past 12 months may cause denominator population to vary widely and introduce too much noise to the measure value. No patient can be in the measure for more than one year based on this requirement—even if the patient is

still be cared for by the facility. Also, inconsistency in denominator definition ("patients at least 18 years old") in statement sp.14 vs stated target population (sp.06) of "Elderly (Age >= 65)".

RELIABILITY: TESTING

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☒ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **eCQM (HQMF) implemented in EHRs** ☐ **Abstracted from Electronic Health Records**

☐ **Abstracted from Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry**

☒ **Enrollment Data** ☒ **Other (please specify)**

Reviewer 5: Renal Management Information System, CROWNWeb facility-reported clinical and administrative data, Medicare enrollment database

Level of Analysis:

☐ **Group/Practice** ☐ **Individual Clinician** ☒ **Hospital/facility/agency** ☐ **Health Plan**

☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**

☐ **Integrated Delivery System** ☐ **Other (please specify)**

Submission document: Questions 2a.01-09

3. **Reliability testing level**

   ☒ **Accountable-Entity Level** ☐ **Patient/Encounter Level** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**

   ☒ **Yes** ☒ **No**

5. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical *VALIDITY* testing** of **patient-level data** conducted?

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Question 2a.10

   Reviewer 1: The method involves calculation of IUR and PIUR statistics, as noted above, with patient-months as the unit of counting and analysis.

   Reviewer 2: Appropriate method. Signal-to-noise analysis.

   Reviewer 3: Both IUR and PIUR methods used by the developers are acceptable for reliability testing.

   Reviewer 4: The IUR is 0.705, which indicates that 70.5% of the variation in the Incident SFR can be attributed to between facility differences in performance (signal) and 29.5% to the within-facility variation (noise). The PIUR is 0.970 which is higher compared to the IUR, indicating the existence of outlier facilities that can be identified by the measure.

Reviewer 5: IUR and PIUR were calculated using conventional ANOVA-based methods.

Reviewer 7: Methods are well described and appropriate.

Reviewer 8: The use of ANOVA test appears to be a reasonable method to assess reliability.

Reviewer 9: Inter-unit reliability (IUR) was utilized to assess the proportion of variation attributable to "between facilities". PIUR, the profile IUR, was used to address that a small IUR may be misrepresented by extreme outcomes. The difference between the IUR and PIUR indicated the ability to identify outliers. Special note to the reliability statistic only being applied to facilities with at least 11 attributable patients over a two-year period.

Reviewer 11: "Accountable Entry Level (e.g., signal-to-noise analysis)." Not clear.

Reviewer 12: OK; SNR (IUR and PIUR for outlier ID); at least 11 pts per 2 yrs.

7. **Assess the results of reliability testing**

Submission document: Question 2a.11

Reviewer 1: Reliability is strong at both IUR and PIUR levels, with the caveat noted earlier about possible non-independence of observations. That would tend to exaggerate reliability calculations.

Reviewer 2: IUR of 0.705; PIUR of 0.970; both suggest a high degree of reliability.

Reviewer 3: The IUR is 0.705 and the PIUR is 0.97. Both are acceptable results. However, it is not clear if not accounting for lack of independence among patient-months should be a concern when interpreting these results.

Reviewer 4: Demonstrated reliability.

Reviewer 5: Reliability is moderate across all facilities, and high for identifying outliers. IUR=0.705, PIUR=0.970 (after exclusion of facilities with <11 patients).

Reviewer 7: Results are indicative of acceptable levels of reliability at the facility level (0.705).

Reviewer 8: The ANOVA test results in an IUR of 0.705. Thus, I would assess the measure performs well in this reliability test.

Reviewer 9: AN IUR of .705 was obtained in this population with a PIUR of .970. Conclusions were that 70.5% of the variation in attributable to between-facility differences in the performance measure and that 29.5 % was within-facility variation. The PIUR result concluded that outlier facilities could be identified though not by the IUR.

Reviewer 11: Based on response to 2a.05, the average number of patients/facilities was about 13. They are only publicly reporting facilities with n>=11. A large percentage of facilities will not have publicly reported data. "The reliability calculation only included facilities with at least 11 patients during the two-year period." Based on their exclusion criteria (ESRD within the past 12 months), there would be no overlap in the patient populations on which reliability is calculated. The IUR results (0.705) are suspect.

Reviewer 12: OK

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? **NOTE:** If multiple methods used, at least one must be appropriate.

Submission document: Question 2a.10-12

☒ **Yes**

☒ **No**

☒ **Not applicable**

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Question 2a.10-12

   ☒ **Yes**

   ☐ **No**

   ☒ **Not applicable** (patient/encounter level testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and ***all*** testing results):

    ☒ **High** (NOTE: Can be HIGH **only if** accountable-entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has **not** been conducted)

    ☒ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of *OVERALL RATING OF RELIABILITY* and any concerns you may have with the approach to demonstrating reliability.**

    Reviewer 1: IUR and PIUR statistics are both relatively high, but we should discuss the issue of independence of observations.

    Reviewer 2: It was unclear in the NQF measure document which data source(s) is(are) actually needed to calculate this measure. The data source listed (claims) was not the only data source used (claims + registry?) to calculate actual performance on the measure.

    Reviewer 3: The developers should address the concern with not accounting for lack of independence among patient-months from the same patient.

    Reviewer 5: No concerns.

    Reviewer 7: See above

    Reviewer 8: See response to #7: The ANOVA test results in an IUR of 0.705. Thus, I would assess the measure performs well in this reliability test.

    Reviewer 9: Moderate IUR with further "explanation" by the PIUR and the difference between them. Note again, for facilities with at least 11 attributable patients.

    Reviewer 11: My concern is based on the changing population specified by the denominator of the measure. If the measure is computed (reported) annually, there is no overlap in the patient population—even if a portion of the original patient population is still under the care of the facility. A better measure of reliability might be change in decile ranking from one year to next.

    Reviewer 12: Ok

VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

    ☒ **Accountable-Entity Level** ☐ **Patient or Encounter-Level** ☐ **Both**

13. **Was the method described and appropriate for assessing the accuracy of ALL *critical data elements*? NOTE** that data element validation from the literature is acceptable.

    Submission document*:* Questions 2b.01-02.

    ☒ **Yes**

    ☐ **No**

    ☒ **Not applicable** (patient/encounter level testing was not performed)

14. **Method of establishing validity at the *accountable-entity level:***

    NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

    Submission document: Questions 2b.01-02

    ☒ **Face validity**

    ☒ **Empirical validity testing at the accountable-entity level**

    ☐ **N/A (accountable-entity level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Question 2b.02

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (accountable-entity level testing was not performed)

16. **Assess the method(s) for establishing validity**

    Submission document: Question 2b.02

    Reviewer 1: Both face validity and empirical validity testing were performed, and both were done using reasonable methods.

    Reviewer 2: Used Poisson regression models to measure association between a facility's performance on the measure (in quintiles) as compared to SMR, SHR and first year SMR. Hypothesized negative relationships with these outcomes.

    Reviewer 3: The developers assessed the measure validity by the examining the associations of this measure score with several related quality of care measures. In particular, the associations with first year standardized mortality ratio and vascular access related infection hospitalization rate are in line with the conceptual rationale of this proposed measure.

    Reviewer 4: The results of the Poisson regression and trend test suggest that lower fistula use is associated with higher risk of mortality and hospitalization (measured by the respective standardized mortality, standardized hospitalization, and first year standardized mortality ratios), as well as all-cause and vascular access infection related hospitalization (measured by the hospitalization rates), as compared to facilities with higher standardized fistula rates.

    Reviewer 5: Construct validity established through facility-level association with established outcome measures, including SMR, SHR, first year SMR, all-cause hospitalization rate, and vascular access infection related hospitalization rate.

Reviewer 7: Methods are well described and appropriate.

Reviewer 8: Use of a Poisson regression to identify potential correlations with several existing measure ratings seems like an appropriate validity test. The hypothesized relationships between the proposed measure and the three stated measures seems logical.

Reviewer 9: Validity was tested by association with other quality metrics: SMR (standardized mortality rate), SHR (standardized hospitalization rate), and First Year SMR across quintiles. Cutoff points were 30.8%, 38.3%, 44.6%, 52.1%, and 99%. The distribution of SFR of each facility and a standard error of the distribution were used to derive a test statistic modified by Efron to reflect national random variation amongst facilities. With this, 2.88% (192) of facilities performed better than expected, 92.45% (6,161) performed as expected, and 4.67% (311) performed worse than expected.

Reviewer 11: Comparing quintile relative risk to hospitalization & mortality.

Reviewer 12: 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

17. **Assess the results(s) for establishing validity**

Submission document: Questions 2b.03-04

Reviewer 1: Validity is adequate - face validity is strong, and empirical validity is at least reasonable, with correlations with other measures in the predicted directions.
Reviewer 2: Found relative risk of mortality/hospitalization increased as performance on measure decreased.
Reviewer 3: The results particularly those based on first year standardized mortality ratio and vascular access related infection hospitalization rate provide support to the validity of this measure.
Reviewer 4: Yes
Reviewer 5: Results closely mirror expectations. Better performance on this measure is associated with lower SMR, lower SHR, lower all-cause hospitalization rate, and lower vascular access infection related hospitalization rate. The effect is particularly strong (RR=1.53) for first year SMR, as expected.
Reviewer 7: Results are satisfactory with expected trends (negative associations) identified between fistula use and risks of mortality and hospitalization.
Reviewer 8: The correlations are in the hypothesized direction stated by the measure submitter. In general, the correlations are modest.
Reviewer 9: The test sample is adequate for implementation. 92+% of facilities perform as expected by the measure. It identified 4.87% or 311 who need improvement.
Reviewer 11: Not impressive. The Relative Risk (RR) values go from Q4: 1.02; Q3: 1.06; Q2: 1.08; Q1: 1.13. Not much difference. The quintile values for hospitalization were in the correct direction but the measure (hospitalization rate expressed in “patient-years”} is confusing. And inclusion of “vascular access related infection” shows very little difference between Q2-Q5 (one hundredth difference at each quintile).
Reviewer 17: Above

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

Submission document: Questions 2b.15-18.

Reviewer 1: None

Reviewer 2: None. High correlation in results with and without exclusions.

Reviewer 3: No concerns.

Reviewer 4: Exclusions acceptable

Reviewer 5: Exclusions for hospice care, metastatic cancer, ESLD, and coma/anoxic brain injury, are parsimonious and appropriate. The exclusions cannot be fully implemented for non-Medicare patients (due to lack of claims data), but these conditions are likely to be much less common in non-Medicare patients.

Reviewer 7: No concerns.

Reviewer 8: No concerns.

Reviewer 9: Missing data was accounted for and provided. The major contributor was the absence of available life expectancy exclusion criteria in 33%. An explanation is provided.

Reviewer 11: See previous comments about ESRD within the previous 12 months.

Reviewer 12: N

19. **Risk Adjustment**

Submission Document: Questions 2b.19-32

19a. **Risk-adjustment method**

☐ None ☒ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

Reviewer 12: Multivariate logistic regression model. The adjustment is made for age, BMI at ESRD incidence, nursing home status in the prior 12 months, nephrologist's care prior to ESRD, diabetes as primary cause of ESRD, and comorbidities at ESRD incidence

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☒ No

19d.**Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☒ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☒ Yes ☐ No

19d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☒ No

19e**. Assess the risk-adjustment approach**

Reviewer 1: It would be nice to have a response option above rather than clear "yes" or "no". Leaving SES factors out of a model for this particular process measure is probably OK, particularly as the developer shows that inclusion/exclusion of these factors doesn't make much difference. However, the standing committee should address this issue from its own perspective - since some social and demographic factors

do have a statistical association with fistula use, is it acceptable to leave them out of an adjustment model?

Reviewer 2: Includes 16 risk factors. Multivariate regression model. C-statistic of 0.748. Risk factors informed by TEP and literature.

Reviewer 3: For a new measure based on a risk model, it is desirable to provide validation of the risk model, not just rely on the development sample.

Reviewer 4: I felt the results from the SDS/SES warranted including these variables for risk adjustment. I believe their rationale for not including the risk factors; Adjusting for these factors could have the unintended effect of masking or reinforcing disparities in vascular access outcomes" was not supported by evidence

Reviewer 5: This is a process measure. Accordingly, we should be very skeptical of risk factors that do not represent contraindications to AVF. Just because a factor is associated with fistula use does not mean that it should be included in risk-adjustment, as it may be in the same causal pathway with quality. Diabetes as primary cause of ESRD is associated with more fistula use; why? Obesity is associated with more fistula use; why? Medicare coverage is associated with more fistula use; why? Many comorbid conditions are associated with SLIGHTLY less fistula use, but again the conceptual rationale is weak.

Reviewer 7: This is a complex measure to risk adjust, given that some patients might have unintended consequences when subjected to fistula surgery. The discussion regarding the inclusion or exclusion of clinical factors considering both statistical and TEP considerations is appreciated. I also appreciate the discussion about the decision not to include social risk factors (although Medicare was included and could be considered as such). However, I do not agree with the use of impact (or lack of) on performance as a good reason to exclude social risk factors, as this criterion was not used for clinical factors. If this criterion is used, it should be used for all variables considered.

Reviewer 8: The approach to consider & institute risk adjustment is logical. I see no issues with the development of the risk model.

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Questions 2b.05-07

Reviewer 1: The measure seems designed to identify extreme outliers and not to make any distinctions among facilities in the broad middle of the performance distribution.

Reviewer 2: None. 93% of facilities were classified as "as expected" with 7% of facilities either better/worse.

Reviewer 3: No concern.

Reviewer 5: Number of statistical outliers is small, but differences are certainly meaningful.

Reviewer 7: Given the analyses conducted, the measure could be close to being topped out, as only 4.7% of facilities performed worse than expected. However, given the importance of the measure, this is not a major concern also given the reliability results at the accountable entity level which support the ability of the measure to identify differences in performance.

Reviewer 8: Modest concern with the measure to identify variation in performance. In response to 2b.06, we see over 92% of facilities are “average” / “as expected”.

Reviewer 9: None

Reviewer 12: N

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Questions 2b.11-14.
Reviewer 1: None
Reviewer 2: Not applicable.
Reviewer 3: No concern.
Reviewer 5: Not applicable.
Reviewer 7: Not applicable
Reviewer 8: No concerns based on responses provided. However, measure submitter did not answer question 2b.12 regarding differing data sources.
Reviewer 9: None
Reviewer 12: N
22. **Please describe any concerns you have regarding missing data.**

Submission document: Questions 2b.08-10.

Reviewer 1: None

Reviewer 2: There does seem to be an issue with missing data for some of the exclusions but given the high correlation of results with and without exclusions, this missingness has limited implications.

Reviewer 3: No concern.

Reviewer 5: Missing Medicare claims data limits application of denominator exclusions, but this problem is well addressed by developers. A more serious problem is the missing CMS-2728. Not submitting a CMS-2728 form seems to be the best way for facilities to justify low fistula use rates. I cannot conceive of any rationale for adjusting for missing data in a way that encourages missingness.

Reviewer 7: No concerns

Reviewer 8: No concerns

Reviewer 9: None

Reviewer 11: Minimal missing data.

Reviewer 12: N

23. **Briefly explain rationale for rating of *OVERALL RATING OF VALIDITY* and any concerns you may have with the developers' approach to demonstrating validity.**

Reviewer 1: Face validity is good, empirical validity is OK, and the decisions about risk adjustment are reasonable, even if I would have preferred something else.

Reviewer 2: Used appropriate methods to test empiric validity of the measure score. Results from testing matched hypothesized relationships. R-A model was well developed.

Reviewer 3: The empirical validity testing results, particularly those based on first year standardized mortality ratio and vascular access related infection hospitalization rate, provide evidence to support the validity of this measure.

Reviewer 4: Overall demonstrated validity but I felt the measure was weaker without the risk adjustment.

Reviewer 5: Major concern is the risk-adjustment model and inappropriate adjustment for factors that do not represent contraindications to AVF. Why would we encourage or facilitate discrimination against certain types of patients by adjusting for these factors? In particular, not submitting a CMS-2728 form seems to be the best way for facilities to justify low fistula use rates.

Reviewer 7: See above

Reviewer 8: Response to question 17: The correlations are in the hypothesized direction stated by the measure submitter. In general, the correlations are modest.

Reviewer 9: Does correlate with other meaningful measures. Does discriminate for those with more than 11 patients per two-year period and identifies about 5% poor performers.

Reviewer 11: How risk adjusted values are used to modify measure score is not clear. Is it used to establish “better” or “worse” than expected?

Reviewer 12: Adequate.

ADDITIONAL RECOMMENDATIONS

24. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Reviewer 1: The use of patient-months as the unit of analysis in reliability calculations is my main concern.
Reviewer 2: I am still a bit confused on which data sources are actually needed to calculate this measure. The form only lists "Claims", but all of the testing documentation seems to reflect multiple data sources.
Reviewer 4: Significant email discussion regarding over inflation of reliability based on the methods used. I hope we have a chance to discuss as a group.
Reviewer 5: Inappropriate risk-adjustment model that increases bias by adjusting for factors in the quality causal pathway, including missing data.
Reviewer 9: Discussion of IUR and PIUR. Adjustment in test statistic according to Efron. Meaningfulness of identifying 5% of under-performers.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

2021 Submission:

Updated evidence information here.

2018 Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]

Several observational studies have demonstrated an association between type of vascular access used for hemodialysis and blood stream infections. Arteriovenous fistulae (AVF) are associated with significantly lower risk of bacteremia compared to long term catheters. Arteriovenous grafts (AVG) have been found to have a risk of blood stream infections that is higher than AVF but lower than catheters.

The measure focus is the facility percentage of patients in their first year of dialysis with an AV Fistula.

Facility actions lead to improvement in blood stream infections as follows:

Measure AV Fistula Rate > Assess value > Identify patients who are new to dialysis and are using a tunneled catheter > Evaluation for a surgical access by a qualified dialysis vascular access provider; consider an AV Fistula when appropriate > Increase Fistula Rate > Lower patient blood stream infection.

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

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

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

Evidence is based on the 2015 Vascular Access TEP, 2020 KDOQI guidelines, and a literate review of studies published from 2020 through February 2020

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

N/A

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

N/A

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation and include the definition of the grade.

[Response Begins]

N/A

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

N/A

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

N/A

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

N/A

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

N/A

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

N/A

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

N/A

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

N/A

[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]

Three main sources of input were used to provide the evidence base for this measure:

1. In 2015, UM-KECC held a Technical Expert Panel (TEP) to seek input for development of a standardized fistula measure[1]. The TEP agreed that AVF are the preferred access for most patients, and that AVG were still preferred relative to a vascular catheter. The TEP recommended that the AVF measure should be adjusted for conditions where an AVG may be an acceptable alternative such as: older age, diabetes, vascular disease, and BMI. Of note, three of our TEP members went on to author/edit the revised KDOQI Vascular Access Guidelines that were published in 2020
2. KDOQI Vascular Access Guidelines[2]: In general, the evidence for the guidelines has been rated as either low or moderate, with many of the guidelines relying on expert opinion. The evidence review team focused on 16 studies and noted that bloodstream infections were significantly lower among patients who started HD with an AV fistula or AV graft versus a catheter. The workgroup refrained from recommending AV fistula on the basis of lower mortality compared to catheter use, and instead relied on the evidence indicating lower blood stream infections. The two guidelines that are most directly relevant to this measure are the following:
    1. KDOQI suggests that if sufficient time and patient circumstances are favorable for a mature, usable AVF, such a functioning AVF is preferred to an AVG in incident HD patients due to fewer longterm vascular access events (eg, thrombosis, loss of primary patency, interventions) associated with unassisted AVF use. (Conditional Recommendation, Low Quality of Evidence)

2. KDOQI suggests that most incident HD patients starting dialysis with a CVC should convert to either an AVF or AVG, if possible, to reduce their risk of infection/bacteremia, infection-related hospitalizations, and adverse consequences. (Conditional Recommendation, Very Low-Moderate Quality of Evidence)
3. From the peer reviewed literature presented in the revised guidelines, the core evidence that the Workgroup was most compelled by centered around the lower rates of bloodstream infection associated with AVF compared to CVCs. This measure relies on those studies that highlight lower infection risk with AVF that have withstood the enhanced scrutiny of the evidence review team, who noted many vascular access studies were observational in nature and thus had a potential risk of bias.

3. Recent peer-reviewed literature continues to highlight benefits of AVF over CVCs particularly for incident patients who are in their first year of dialysis. Notable recent findings include:

- AVF associated with lower risk of access-related hospitalizations: One study using USRDS data that focused on elderly patients who started dialysis with a catheter and had an AVF or AVG created within the first 6 months of dialysis noted that AVF creation was associated with a lower risk of access-related hospitalization[3].
- AVF has lower rates of blood stream infection and sepsis compared to AVG or CVC: In a study of 2352 incident dialysis patients, after adjusting for confounders, AVF use was associated with 61% lower risk of blood stream infections compared with CVC or AVG use8. In a separate study, based on the vascular access used at initiation of dialysis, patients with AVG (HR 1.35) and CVC (HR 1.80) were more likely to develop sepsis (both P < .001)6. Additionally, in patients who developed sepsis, mortality at 1 year after sepsis was 21% higher in AVG and nearly doubled in CVC when compared to AVF. A third study of patients over the age of 67 who start dialysis with a catheter and went on to have either an AVF or AVG placed in the first 6 months reports that rates of all-cause infection-related hospitalization (RR 0.93, P=0.01) and bacteremia/septicemia-related hospitalization (RR 0.90; P=0.02) were lower in the AVF group versus AVG group10.
- AVF have lower maintenance interventional requirements compared to AVG: Using USRDS data and accounting for patient characteristics, one study[4] reported that during maturation of the AV access, interventions for both AVFs and AVGs were relatively common and similar between the two types of access. However, once successfully matured, AVFs had lower maintenance interventional requirements.
- Catheter dependence after AVF or AVG placement among elderly incident dialysis patients is complex: for many younger hemodialysis patients, creation of an AVF, compared with an AVG, is associated with longer initial catheter dependence, but then longer access survival and lower long-term catheter dependence. In patients ≥ 67 years of age, similar increased catheter dependence was found at 1 and 3 months after AVF creation, compared to AVG, but lower catheter dependence at 12 and 36 months[5]. However, creation of AVF in the older population was associated with greater cumulative catheter-dependent days (80 vs 55 days per person-year) after 3 years of follow up. From a cost perspective, Hall and colleagues report[11] that based on Markov models of hypothetical patients starting dialysis with a CVC, the AVF option was cost effective compared with continued catheter use for all age and life expectancy groups, except for 85–89-year-olds in the lowest life expectancy quartile. The AVF option was more cost effective than the AVG option for all quartiles of life expectancy among the 65- to 69-year-old age group. For older age groups, differences in cost-effectiveness between the strategies were attenuated, and the AVF option tended to only be cost effective in patients with life expectancy >2 years. These findings highlights that not all elderly patients will realize the benefit of catheter independence from AVF creation and specific patient characteristics and shared decision making remain critical in appropriate vascular access selection.
- AVF is associated with higher health-related quality of life (HRQOL) and less depression compared to CVC in the first year of dialysis: A prospective cohort study of 1461 patients who initiated dialysis reported that patients with an AVF had higher KDQOL-36 scores and lower Beck Depression Inventory scores at 3 months and 12 months after the initiation of dialysis compared to those with CVC[7]. Furthermore, in a survey conducted by the American Association of Kidney Patients[9], satisfaction with current vascular access was 90% with AVF, 79% with AVG, and 67% with CVC. The factors most frequently reported as important in influencing the selection of vascular access modality included infection risk (87%), physician recommendation (84%), vascular access durability (78%), risk of complications involving surgery (76%), and impact on daily activities (73%). As we navigate vascular access decisions that embrace shared decision making and respect patient choice, these two studies highlight that the majority of patients who choose an AVF are satisfied with that decision and may enjoy better health-related quality of life.

In summary, the recently revised KDOQI guidelines for vascular access continue to support AV fistula as the preferred vascular access for most patients on dialysis, although with less emphasis than in prior

iterations. Long-term catheters are still viewed as the least desirable vascular access, primarily due to the increased risk of blood-stream infections, with increased recognition of certain patient characteristics and scenarios where this access type may be the most appropriate. Given that over 80% of new hemodialysis patients start with a CVC, the additional studies noted above that were published after the updated KDOQI guidelines suggest that attempts to create AVF are still warranted.

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]

In addition to the above evidence supporting the measure, there are three major healthcare processes associated with achieving AVF creation:

1. Patient education interventions: Providing kidney disease education is associated with 1.78 increased odds of starting dialysis with an AVF and a 0.51 odds of starting dialysis with a CVC alone[12]. Patient education can enhance motivation and potentially lead to improved health outcomes[13].
2. Vascular Access Coordinator/Program: In one study[14], an organized dialysis access program resulted in a 82% decrease in the number of central venous catheter days which lead to a concurrent reduction in central line-associated bloodstream infection and deaths. As a result of creating an access program, central venous catheter rates decreased from an average rate of 45% to 8%.
3. Surgeon Selection: Several studies have suggested that there is significant variation in likelihood of AVF, as opposed to AVG, creation based on the vascular access surgeon. Using a national claims database to identify patients initiating hemodialysis with a CVC, and adjusting for demographic and comorbid conditions, the individual surgeon identifier had the greatest magnitude of effect on access type (AVF or AVG) created, with some surgeons more than twice as likely to create AVF as other surgeons[15]. Thus, surgeon selection by the dialysis facility is an important component in efforts to maximize creation of AVF in otherwise eligible patients.

Comorbidity adjustment : One frequently cited barrier to successful AVF creation has been the burden of comorbidities at the dialysis facility level. A recent study[16] noted that after adjustment for facility-level comorbidity burden, only small differences in facility rates of AVF use were seen except in the extremes of high or low levels of comorbidity burden. This suggests that dialysis facilities with a relatively high patient comorbidity burden can achieve similar fistula rates as facilities with healthier patients if the above care processes are employed.

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]

We searched Pubmed using the following search string: “Arteriovenous fistula OR venous catheter AND dialysis AND published January 1, 2017 – February 1, 2022”. This returned 43 citations that were reviewed independently by two faculty knowledgeable about dialysis vascular access and the following citations were selected for consideration.

[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

References

1. UM-KECC Vascular Access Technical Expert Panel Summary Report 2015 available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/ESRD-Vascular-Access-TEP-Summary-Report.pdf
2. Lok CE, Huber TS, Lee T, et al; KDOQI Vascular Access Guideline Work Group. KDOQI clinical practice guideline for vascular access: 2019 update. Am J Kidney Dis. 2020;75(4)(suppl 2):S1-S164. https://www.ajkd.org/article/S0272-6386(19)31137-0/fulltext

3. Lyu B, Chan MR, Yevzlin AS, Gardezi A, Astor BC. Arteriovenous Access Type and Risk of Mortality, Hospitalization, and Sepsis Among Elderly Hemodialysis Patients: A Target Trial Emulation Approach. Am J Kidney Dis. 2022 Jan;79(1):69-78. doi: 10.1053/j.ajkd.2021.03.030. Epub 2021 Jun 9
4. Woodside KJ, Repeck KJ, Mukhopadhyay P, Schaubel DE, Shahinian VB, Saran R, Pisoni RL. Arteriovenous Vascular Access-Related Procedural Burden Among Incident Hemodialysis Patients in the United States. Am J Kidney Dis. 2021 Sep;78(3):369-379.e1. doi: 10.1053/j.ajkd.2021.01.019. Epub 2021 Apr 20.
5. Lyu B, Chan MR, Yevzlin AS, Astor BC. Catheter Dependence After Arteriovenous Fistula or Graft Placement Among Elderly Patients on Hemodialysis. Am J Kidney Dis. 2021 Sep;78(3):399-408.e1. doi: 10.1053/j.ajkd.2020.12.019. Epub 2021 Feb 12.
6. Locham S, Naazie I, Canner J, Siracuse J, Al-Nouri O, Malas M. Incidence and risk factors of sepsis in hemodialysis patients in the United States. J Vasc Surg. 2021 Mar;73(3):1016-1021.e3. doi: 10.1016/j.jvs.2020.06.126. Epub 2020 Jul 21.
7. Kim DH, Park JI, Lee JP, Kim YL et al. The effects of vascular access types on the survival and quality of life and depression in the incident hemodialysis patients. Ren Fail. 2020 Nov;42(1):30-39. doi: 10.1080/0886022X.2019.1702558
8. Kazakova SV, Baggs J, Apata IW, Yi SH, Jernigan JA, Nguyen D, Patel PR. Vascular Access and Risk of Bloodstream Infection Among Older Incident Hemodialysis Patients. Kidney Med. 2020 Mar 16;2(3):276-285. doi: 10.1016/j.xkme.2019.12.010 eCollection 2020 May-Jun.
9. Balamuthusamy S, Miller LE, Clynes D, Kahle E, Knight RA, Conway PT. American Association of Kidney Patients survey of patient preferences for hemodialysis vascular access. J Vasc Access. 2020 Mar;21(2):230-236. doi: 10.1177/1129729819870962. Epub 2019 Aug 29
10. Lee T, Thamer M, Zhang Q, Zhang Y, Allon M. Vascular Access Type and Clinical Outcomes among Elderly Patients on Hemodialysis. Clin J Am Soc Nephrol. 2017 Nov 7;12(11):1823-1830. doi: 10.2215/CJN.01410217. Epub 2017 Aug 10.
11. Hall RK, Myers ER, Rosas SE, O'Hare AM, Colón-Emeric CS. Choice of Hemodialysis Access in Older Adults: A Cost-Effectiveness Analysis. Clin J Am Soc Nephrol. 2017 Jun 7;12(6):947-954. doi: 10.2215/CJN.11631116. Epub 2017 May 18.
12. Ruchi R, Bozorghehri S, Chamarthi G et al. Provision of Kidney Disease Education Service Is Associated with Improved Vascular Access Outcomes among US Incident Hemodialysis Patients. Kidney360 January 2022, 3 (1) 91-98; DOI:https://doi.org/10.34067/KID.0004502021
13. Lacson E Jr., Wang W, DeVries C, et al. Effects of a nationwide predialysis educational program on modality choice, vascular access, and patient outcomes. Am J Kidney Dis 58: 235–242, 2011 https://doi.org/10.1053/j.ajkd.2011.04.015
14. Rosenberry PM, Niederhaus SV, Schweitzer EJ, Leeser DB. Decreasing dialysis catheter rates by creating a multidisciplinary dialysis access program. J Vasc Access. 2018 Nov;19(6):569-572. doi: 10.1177/1129729818762977. Epub 2018 Mar 26.
15. Copeland TP, Lawrence PF, Woo K. Surgeon Factors Have a Larger Effect on Vascular Access Type and Outcomes than Patient Factors. J Surg Res. 2021 Sep;265:33-41. doi: 10.1016/j.jss.2021.02.046. Epub 2021 Apr 18.
16. Dahlerus C, Kim S, Chen S, Segal JH. Arteriovenous Fistula Use in the United States and Dialysis Facility-Level Comorbidity Burden. Am J Kidney Dis. 2019 Nov 22:S0272-6386(19)31031-5. doi: 10.105/j.ajkd.2019.08.023.

[Response Ends]

1b. Gap in Care/Opportunity for Improvement and Disparities

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

The Standardized Fistula Rate (SFR) for Incident Patients is based on the prior SFR (NQF #2977) that included both incident and prevalent patients. That measure was initially endorsed in 2016, but as part of a measure maintenance review by the NQF Standing Committee in 2020, concerns were raised about the strength of evidence supporting the prior measure. Namely, recent updates to the KDOQI guidelines downgraded the evidence supporting fistula as the preferred access type and instead focused on catheter avoidance and developing an individualized ESKD Lifeplan. However, the guidelines do suggest that under favorable circumstances an AV fistula is preferred to an AV graft in incident patients due to fewer long-term vascular access events. Given that over 80% of incident dialysis patients begin treatment with a tunneled catheter, and that 12 months after dialysis initiation AV fistula rates exceed 60%, the incident SFR was developed to focus on the subset of dialysis patients that the evidence suggests may benefit the most during a time of intense vascular access creation. Specifically, blood stream infection rates are the lowest in incident patients with AV fistula compared to long-term catheters. Therefore, the goal of this new measure is to evaluate facility performance in increasing fistula use in the incident population in order to reduce the heightened risks patients face due to bacteremia and infection-related hospitalizations.

By focusing on the incident population, we have addressed two other concerns that were raised by the NQF Standing Committee in 2020 with regards to NQF #2977 that included both incident and prevalent patients.

- Exhausted Vascular Access Options: One difficulty with the prior SFR measure was the inability to account for patients that had extensive dialysis exposure and had multiple failed vascular accesses such that they were deemed catheter dependent. While there was widespread agreement that these patients should be excluded from a fistula measure, there has been no way to operationalize that exclusion criteria and no consensus was reached by our 2015 TEP as to how best to do so. This measure, by focusing on patients in their first 12 months of dialysis, avoids the problem of exhausted vascular access since this is typically not encountered in such a short time span.
- Performance Gap: During the consideration of NQF #2977, the performance gap was considered to be negligible since the national average for the standardized fistula rate was at 63% (even though there were, and continue to be, significant disparities as well as variation between providers). By focusing on incident patients where the catheter rate is dramatically higher than seen in prevalent patients, there is more room for improvement in placing AVF in patients who are deemed appropriate candidates. In addition, as is noted in Section 1b.02 below, there is marked variation between facilities in AVF creation during the first year of dialysis.

With a focus on the incident ESRD dialysis population, this measure addresses an important gap in quality outcomes, namely that of catheter-related blood stream infections. 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 approach is supported by the recent KDOQI updated guidelines which stress a patient centered approach to vascular access.

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. 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 (incident SFR, long-term catheter rate) reflects consensus best practice, and supports maintenance of the gains in vascular access success achieved over the last decade.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

N of facilities=6664; N of patient-months=1,864,647.

Table 1. Descriptive statistics of SFR (%), overall and by decile, 2018-2019

*	Mean	Std Dev	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Overall	*	*	*	*	*	*	*
*	41.4	12.7	0.0	32.9	41.6	49.9	99.0
Decile	*	*	*	*	*	*	*
1	19.7	5.4	0.0	25.8	21.3	17.0	23.8
2	28.9	1.6	25.8	31.4	29.1	27.6	30.2
3	33.4	1.1	31.4	35.2	33.5	32.5	34.4
4	36.9	0.9	35.2	38.4	37.0	36.1	37.6
5	40.0	0.9	38.4	41.5	40.0	39.2	40.7
6	43.0	0.9	41.5	44.6	43.0	42.3	43.8
7	46.2	1.0	44.6	47.8	46.2	45.4	47.0
8	49.6	1.1	47.8	51.6	49.6	48.8	50.5
9	54.2	1.7	51.6	57.2	54.1	52.8	55.6
10	63.2	5.3	57.2	98.9	61.8	59.3	65.9

*cells intentionally left blank

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

We provide descriptives of SFR % by race, ethnicity, sex, insurance status and the area deprivation index (higher values mean higher SES deprivation). There are clinically meaningful differences in mean percentage of SFR based on sex, race, ethnicity, and SES as measure by the ADI. These suggest potential disparities in AVF use among the incident ESRD dialysis population.

N of facilities=6664; N of patient-months=1,864,647.

Table 2.Descriptive statistics of SFR (%), by demographic characteristics, 2018-2019

*cells intentionally left blank

*	Mean	Std Dev	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Sex	*	*	*	*	*	*	*
Female	35.1	11.7	1.3	27.3	34.7	42.5	98.6
Male	46.2	12.7	2.1	38	46.4	54.7	99.1
Race	*	*	*	*	*	*	*
White	43.4	12.4	1.9	35.3	43.5	51.6	99
Black	36.4	11.7	1.4	28.5	36.1	43.9	84.7
Other race	45.1	12	6.5	37.2	45.2	53.2	99.1
Ethnicity	*	*	*	*	*	*	*
Hispanic	45.1	12.4	3.6	37	45.3	53.4	99.1
Non-Hispanic	40.8	12.1	1.7	32.8	40.7	48.7	98.9
Employment Status	*	*	*	*	*	*	*
Employed	46.9	12.7	2.1	38.6	47.1	55.4	99.2
Unemployed	38.9	12	1.5	30.9	38.6	46.6	98.8
Other	41.2	12.2	1.7	33.2	41.1	49.1	98.9
Medicare Coverage	*	*	*	*	*	*	*
Dual eligible	40.8	12.1	1.7	32.8	40.7	48.7	98.8
Non dual eligible	41.9	12.3	1.8	33.8	41.9	50	98.9
ADI (quintiles)	*	*	*	*	*	*	*
0-33.3	43.3	12.1	1.9	35.3	43.2	51.5	86.1
33.3-51.1	42.2	12.2	1.8	34.2	42.2	50.4	98.9
51.1-64.3	41.3	12.1	1.7	33.3	41.3	49.3	87.3
64.3-76.6	40.5	12.1	1.7	32.4	40.4	48.4	98.9
76.6-100	39	12	1.6	31	38.8	47	86.4

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

N/A

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

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

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Standardized Fistula Rate for Incident Patients

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

Adjusted percentage of adult incident hemodialysis patient-months using an autogenous arteriovenous fistula (AVF) as the sole means of vascular access.

The Standardized Fistula Rate (SFR) for Incident Patients is based on the prior SFR (NQF #2977) that included both incident and prevalent patients. This measure was initially endorsed in 2016, but as part of measure maintenance review by the NQF Standing Committee in 2020, concerns were raised about the strength of evidence supporting the prior measure. Namely, recent updates to the KDOQI guidelines downgraded the evidence supporting fistula as the preferred access type and instead focus on catheter avoidance and developing an individualized ESKD Lifeplan. However, the guidelines do suggest that under favorable circumstances an AV fistula is preferred to an AV graft in incident patients due to fewer long-term vascular access events. Given that over 80% of incident dialysis patients begin treatment with a tunneled catheter, and that 12 months after dialysis initiation AV fistula rates exceed 60%, the incident SFR was developed to focus on the subset of dialysis patients that the evidence suggests may benefit the most during a time of intense vascular access creation. Specifically, blood stream infection rates are the lowest in incident patients with AV fistula compared to long-term catheters. Therefore the goal of this new measure is to evaluate facility performance in increasing fistula use in the incident population in order to reduce the heightened risks patients face due to bacteremia and infection related hospitalizations.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Renal: End Stage Renal Disease (ESRD)

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Care Coordination

Safety: Complications

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Elderly (Age >= 65)

Populations at Risk: Individuals with multiple chronic conditions

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Facility

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Outpatient Services

[Response Ends]

sp.09. 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. If no URL is available, indicate "none available".

[Response Begins]

N/A

[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: Incident SFR Data Dictionary Code Table.xlsx

sp.12. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

The numerator is the adjusted count of adult incident patient-months using an AVF as the sole means of vascular access as of the last hemodialysis treatment session of the month.

[Response Ends]

sp.13. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

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. 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 in the Medical Evidence Form for the corresponding comorbidity, we assume this patient did not have the comorbidity at the start of chronic dialysis. The same methodology is applied to the comorbidity exclusions and the hospice exclusion in the preceding 12 months of Medicare claims.

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

All patient-months for 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) and became ESRD within the prior 12 months for the entire reporting month at the same facility.

[Response Ends]

sp.15. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

For each patient, we identify the dialysis provider at each month using a combination of data from CROWNWeb, Medicare-paid dialysis claims, and the Medical Evidence Form (Form CMS-2728). 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 ESRD and beginning chronic dialysis (in-center hemodialysis or home hemodialysis) within the prior 12 months, 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 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.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Exclusions that are implicit in the denominator definition include:

- Patient-months after 12 months of starting ESRD
- Pediatric patients (<18 years old)
- Patients-months 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

The denominator is defined at the patient level not facility level. The reason this rule is applied is to comport with how measures are implemented for public reporting. Due to small cell size and potentially identifiable data, facilities with <11 patients do not receive a score.

As stated in the measure description and rationale, this is a measure of incident patients only. Dialysis patients in their first 12 months of ESRD are more likely to be using a catheter for vascular access and in turn are at higher risk for CVC related infections. The measure focus is on the first 12 months of dialysis since this is the most active time of vascular access creation and where the potential benefit is greatest relative to treatment with a CVC.

Patient attribution to facilities is already described – see SP15: "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."

When a patient is not treated in a single facility for a span of 30 days (for instance, if there were two facility transfers within 30 days of each other), we do not attribute that patient to any facility for that month. Therefore, transient treatment at a facility due to either travel or a temporary clinical condition do not impact the fistula rate of that facility.

Patients with a catheter (of any duration) AND one or more of the limited life expectancy exclusions are excluded from the denominator.

[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

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-10 diagnostic codes used to identify these comorbidities is included in the attached data dictionary code table (excel file).

[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

N/A

[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

Statistical risk model

[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

See attached flowchart.

[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]

N/A

[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

Registry Data

[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

Data are derived from an extensive national ESRD patient database, which is primarily based on 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 and patient tracking data), the Medicare Enrollment Database (EDB), and Medicare claims data. In addition the database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), and data from the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC) (which includes Provider and Survey and Certification data from Automated Survey Processing Environment (ASPEN)), and the Dialysis Facility Compare (DFC).

The database is comprehensive for Medicare patients not enrolled in Medicare Advantage. Medicare Advantage patients are included in all sources but their Medicare payment records are limited to inpatient claims. Non-Medicare patients are included in all sources except for the Medicare payment records. Tracking by dialysis provider and treatment modality is available for all patients including those with only partial or no Medicare coverage. Past-year comorbidity data are obtained from multiple claim types (inpatient, outpatient, home health, hospice, skilled nursing facility claims).

CROWNWeb is the data source for establishing the vascular access type used to determine the numerator.

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- 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.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

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

2a. Reliability testing 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 instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing 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 instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

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).

2b3. 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 social risk 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.

2b4. 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.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. 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 non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

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).

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. The degree of consensus and any areas of disagreement must be provided/discussed.

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.

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

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

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.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

2021 Submission:

Updated testing information here.

2018 Submission:

Testing from the previous submission here.

2a. Reliability

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Claims

Registry Data

[Response Ends]

2a.02. 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).

[Response Begins]

National CROWNWeb data from January 2018-December 2019 and Medicare claims data from January 2017 – December 2019

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

01-01-2017 to 12-31-2019

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*

- *Population: Population*

[Response Begins]

Facility

[Response Ends]

2a.05. List the measured entities 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.

[Response Begins]

Patients on both home and in-center hemodialysis during the last HD treatment of the month from January 2018-December 2019 and starting chronic dialysis within the prior 12 months of the reporting month were included in the analyses. The number of facilities per month ranged from 6,355-6,659 and the total number of patients per month ranged from 76,012- 79,823.

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 reporting period for the measure, meaning facilities with <11 patients would have their results suppressed. Patients at those facilities are still included in the modeling of the measure unless they otherwise meet the patient-level exclusion criteria. We have applied this restriction to all the reliability and validity testing reported here.

Patients at those facilities with <11 attributed patients are still included in our modeling and are not excluded. The exclusion criteria are patient level as described in Sp.16.

For public reporting facilities with <11 patients would have their results suppressed.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

There were a total of 1,871,951 eligible patient-months. Among those patient-months over the whole reporting period, the average age was 64.4 years, 41.8% were female, 64.0% were white, 28.6% were black, 7.4 % reported race as “other”, 16.9% were Hispanic and 51.0% had type II diabetes as the primary cause of ESRD.

[Response Ends]

2a.07. 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.

[Response Begins]

N/A

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

Patient level:

- Employment status 6 months prior to ESRD
- Sex
- Race
- Ethnicity
- Medicare dual eligible
- ZIP code level – Area Deprivation Index (ADI) from Census data (2011-2015*). Based on patient zip-code.

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

*University of Wisconsin School of Medicine and Public Health. 2015 Area Deprivation Index v3.0. Downloaded from https://www.neighborhoodatlas.medicine.wisc.edu/

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing 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.

[Response Begins]

We used January 2019 – December 2019 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 Incident 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 Incident 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 Incident 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 Incident 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, $\bar{T} = \sum n_i T_i \Big/ \sum n_i$, and $n' = \frac{1}{N-1}(\sum n_i - \sum n_i^2 \Big/ \sum n_i)$ is approximately the average facility size (number of patients per facility). Thus, the IUR = $\sigma_b^2 \Big/ (\sigma_b^2 + \sigma_{t,w}^2)$ can be estimated by $(s_t^2 - s_{t,w}^2) \Big/ s_t^2$.

To assess more directly the value of the measure in identifying providers with extreme outcomes, we also computed an additional metric, termed the profile IUR (PIUR). This was to address the challenge that the IUR could be small with many providers having outcomes around the national norm, even though the measure may still be able to identify facilities with extreme outcomes. The PIUR, based on the measure's ability to consistently flag extreme providers, was computed with a two-step approach: first, we evaluated the ability of a measure to consistently profile facilities with extreme outcomes; second, we mapped this reflagging ability to an IUR value computed by assuming no outlier facilities. This value was defined to be the PIUR. The difference between the PIUR and the IUR indicates the extent to which the measure identifies outliers.

The reliability calculation only included facilities with at least 11 patients during the two-year period.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

The IUR is 0.705, which indicates that 70.5% of the variation in the Incident SFR can be attributed to between-facility differences in performance (signal) and 29.5% to the within-facility variation (noise).

The PIUR is 0.970 which is higher compared to the IUR, indicating the existence of outlier facilities that can be identified by the measure but were not captured by the IUR.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

The result of IUR and PIUR testing suggests a high degree of reliability.

[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

[Response Ends]

2b.02. 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.

[Response Begins]

Validity was assessed using Poisson regression models to measure the association between facility level quintiles of performance scores and the 2018-2019 Standardized Mortality Ratio (SMR, NQF 0369), 2018-2019 Standardized Hospitalization Ratio (SHR, NQF 1463), and 2018 First Year Standardized Mortality Ratio (SMR) respectively. Facility-level performance scores were divided into quintiles (Q1 to Q5), and the relative risk (RR) for SMR (and SHR and first year SMR, separately) was calculated for each quintile, using Q5 as the reference group. A RR>1.0 would indicate a higher relative risk of mortality or hospitalization, compared to the lowest performance score quintiles.

For the all-cause hospitalization rate and vascular access infection related hospitalization rate, we used linear regression to test the association between the SFR quintiles and the 2018-2019 all-cause hospitalization rate, and 2018-2019 vascular access related infection hospitalization rate, respectively. For all-cause hospitalization and vascular access related infection hospitalization, the respective rate was calculated for each quintile and a trend test of the rates was performed.

- SMR: We expect a negative association with SMR since successfully creating an AVF is generally seen as representing a robust process to coordinate care outside of the dialysis facility. AVFs are also associated with lower risk of infection which may reduce the risk of a life-threatening infection or other poor outcomes that place patients at higher risk of mortality. Higher standardized fistula rates will be negatively associated with SMR.
- SHR: We expect a negative association with SHR since successfully creating an AVF is generally seen as representing a robust process to coordinate care outside of the dialysis facility, and potentially reduces the risk for patients at such facilities going to hospital due to infections or other acute clinical events. Higher standardized fistula rates will be negatively associated with SHR.
- First Year SMR: We expect a negative association with the first year SMR as many incident patients begin with a catheter, and therefore face higher risk for infection compared to patients with an AVF. AVFs are associated with lower risk of infection which may reduce the risk of a life-threatening infection or other poor outcomes that

place patients at higher risk of mortality particularly in their first year of dialysis. Higher standardized fistula rates will be negatively associated with the first year SMR.

- All-cause hospitalization rate: We expect a negative association between all-cause hospitalization rates and higher AVF rates given the known risk of infection and other complications related to long-term catheter dependence, particularly in incident patients.
- Vascular access related infection hospitalization rate: We expect a negative association between access related infection hospitalizations and AVF rates because of the higher rates of catheter in patients in the first year of dialysis, which creates a higher risk of a catheter related infection.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

Cut-points for the quintiles of the performance scores were defined as follows:

Q1: 0% - <30.8%

Q2: 30.8% - <38.3%

Q3: 38.3% - <44.6%

Q4: 44.6 - <52.1%

Q5: 52.1% - <99.0% 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 2018-2019 SMR, the relative risk of mortality increased as the performance measure quintile decreased from the reference group (Q5)with the highest risk in quintile 1. For quintile 4, RR=1.02 (95% CI: 1.00, 1.04; p<0.001), quintile 3, RR=1.06 (95% CI: 1.04, 1.08; p<0.001), quintile 2, RR=1.08 (95% CI: 1.06, 1.10; p<0.001), and quintile 1, RR=1.13 (95% CI: 1.11, 1.15; p<0.001).

Similarly for 2018-2019 SHR, the relative risk of hospitalization increased as the performance measure quintile decreased from the reference group (Q5) with the highest risk in quintile 1. For quintile 4, RR=1.06 (95% CI: 1.05, 1.06; p<0.001), quintile 3, RR=1.07 (95% CI: 1.06, 1.07; p<0.001), quintile 2, RR=1.11 (95% CI: 1.10, 1.12; p<0.001), and quintile 1, RR=1.15 (95% CI: 1.14, 1.15; p<0.001).

For the 2018 first year SMR, the relative risk of mortality increased as the performance measure quintile decreased from the reference group (Q5)with the highest risk in quintile 1. For quintile 4, RR=1.08 (95% CI: 1.03, 1.14; p=0.002), quintile 3, RR=1.11 (95% CI: 1.05, 1.16; p<0.001), quintile 2, RR=1.17 (95% CI: 1.12, 1.23; p<0.001), and quintile 1, RR=1.53 (95% CI: 1.46, 1.60; p<0.001).

For the 2018-2019 all-cause hospitalization, the hospitalization rate decreased as the performance measure quintile increased. Hospitalization rates for quintiles 1 to 5 are 1.06, 0.99, 0.95, 0.93, and 0.87 patient-years respectively (trend test p<0.001).

For the 2018-2019 vascular access related infection hospitalization, the hospitalization rate decreased as the performance measure quintile increased. Hospitalization rates for quintiles 1 to 5 are 0.22, 0.18, 0.17, 0.16, and 0.15 respectively (trend test p<0.001).

[Response Ends]

2b.04. Provide 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?)

[Response Begins]

The results of the Poisson regression and trend test suggest that lower fistula use is associated with higher risk of mortality and hospitalization (measured by the respective standardized mortality, standardized hospitalization, and first year standardized mortality ratios), as well as all-cause and vascular access infection related hospitalization (measured by the hospitalization rates), as compared to facilities with higher standardized fistula rates.

[Response Ends]

2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

Differences in measure performance were evaluated separately for each facility, where the incident 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 Incident 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 for the incident ESRD population.

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

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

[Response Begins]

Table 7. Proportion of facilities with statistically significant differences (p-values < 0.05), using 2018-2019 data

Category	Number of facilities	Percent of facilities
Better than expected	192	2.88
As expected	6161	92.45
Worse than expected	311	4.67

[Response Ends]

2b.07. Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

For the 2018-2019 Incident SFR, 6161 (92%) facilities have achieved expected performance, 311 (4.7%) facilities have performed worse than expected (lower fistula rate), 192 (2.9%) 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.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

The Standardized Fistula Rate measure includes all patients regardless of Medicare coverage. The measure is based on data from CROWNWeb (representative of all ESRD dialysis patients) and Medicare claims. Missing data for vascular access type occurs rarely. We report the frequency of missing for the data below (patient-month level).

- Missing CMS 2728 – All ESRD patients are required to have a CMS 2728 form submitted to CMS regardless of the patient's Medicare status. The 2728 certifies the patient has ESRD. The 2728 data are in CROWNWeb.
- Missing BMI from the 2728 – see above. This is part of the required fields in the 2728.
- The overall percentage of patient months with missing vascular access type – In CROWNWeb, "Access Type IDs" were reported for each patient at each month. A vascular access type is considered missing if no access type ID was found at the given month.
- Patient months where we are unable to determine presence of comorbidities for the limited life expectancy exclusions. For each month we search patients' Medicare claims for the presence of any comorbidity exclusion conditions in the past 12 months. If no claims were found, we consider the months as "unable to determine the presence of comorbidities".

Incident SFR uses CROWNWeb clinical data along with other CMS administrative data for several important components of the measure calculation. These include comorbidities at ESRD incidence for risk adjustment from the CMS 2728, and

claims-based conditions for the limited life expectancy exclusions. The source of vascular access type and several clinical risk factors are from CROWNWeb (registry that includes all ESRD dialysis patients).

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

Summary findings:

1. Patient-months with missing CMS 2728 is 1.11% and missing BMI on the CMS 2728 is 0.01% from the measure.
2. 3.6% of overall patient months are missing vascular access type.

Table 8. Frequency of missing data elements, 2018-2019 data

Data Element	Missing (n)	Total (n)	Missing (%)
Patient months with missing CMS 2728	20,038	1,864,647	1.07%
Patient months without BMI reported on 2728	235	1,844,609	0.01%
Vascular access type	57,801	1,864,647	3.1%
Patient months where we are unable to determine the presence of limited life expectancy exclusion conditions*	614,367	1,864,647	32.95%

[Response Ends]

2b.10. Provide 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 non-responders), and how the specified handling of missing data minimizes bias.

In other words, 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 was conducted, justify the selected approach for missing data.

[Response Begins]

There is a very low frequency of patients with missing BMI from the CMS form 2728, and missing CMS form 2728 for Incident SFR. Missing CMS 2728 form was adjusted through inclusion of a missing indicator in the logistic regression model, and missing BMI was included as BMI 30+ category (the group with the highest frequency). Given such a small percent of missing (0.01% for BMI and 1.1% for CMS 2728 from), the impact of missing data for these elements is negligible.

Failure to report vascular access type indicates facilities are not appropriately monitoring or reporting vascular access outcomes as required. Reporting months with missing vascular access values are not excluded from this measure. We count patient months with missing vascular access type in the denominator for the Incident SFR, and missing is treated as a "catheter" and therefore does not count in the numerator for the Incident SFR. Since these patient months are not excluded from the measure, bias from missing vascular access type is not a consideration for the Incident SFR.

The percentage of patient months we are unable to identify the comorbidity exclusions is 32.47%. We acknowledge this is a general limitation of relying on FFS Medicare claims for ascertaining comorbidities for the incident population in particular who may not yet have a sufficient claims history. However, as shown in the exclusion analysis, Incident SFR with and without the exclusions applied are highly correlated. This suggests the unavailability of claims for non-Medicare patients to identify exclusions does not bias Incident SFR performance scores.

The 2.64% in 2b.16 refers to patient-months that are excluded based on a Medicare claim with one or more of the limited life expectancy exclusion conditions. In 2b.10 we are reporting the percentage of patient-months for which we cannot find a Medicare claim in the prior 12 months. That number is 32.95% of patient-months. These exclusion comorbidities are treated as missing (not excluded) due to patients not being Medicare, having Medicare coverage that is coincident with ESRD and thus no prior claims, or being Medicare Advantage patients that do not have inpatient claims. These patient-months are NOT excluded. The impact of these missing exclusion comorbidities is small. As shown in the exclusion analysis, Incident SFR with and without the exclusions applied are highly correlated. This suggests the unavailability of claims for MA or non-Medicare patients to identify exclusions does not bias Incident SFR performance scores.

We have edited sections 2b.08 – 2b.10 to provide greater clarity for the results.

[Response Ends]

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

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. 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. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

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?

[Response Begins]

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.

[Response Ends]

2b.17. Provide 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.

[Response Begins]

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

Table 1: Percent of patient-months at risk excluded, 2018-2019 data

Year	Before Exclusion	After Exclusion	Percent
2018-2019	1,915,119	1,864,647	2.64%

Table 2: Number and percent of unique patients excluded, 2018 data

Year	Before Exclusion	After Exclusion	Percent
2018-2019	276,606	270,702	2.13%

Table 3: Distribution of performance scores before and after the exclusion, 2018-2019 data

Standardized Fistula Rate	N	Mean	Std Dev	Minimum	Maximum	Median	Lower Quartile	Upper Quartile
Before exclusion	6664	40.369	12.459	0	94.425	40.472	31.966	48.619
After exclusion	6664	41.445	12.656	0	99.02	41.563	32.932	49.945

Figure 1: Scatterplot –Incident SFR with and without measure exclusions, 2018-2019 data

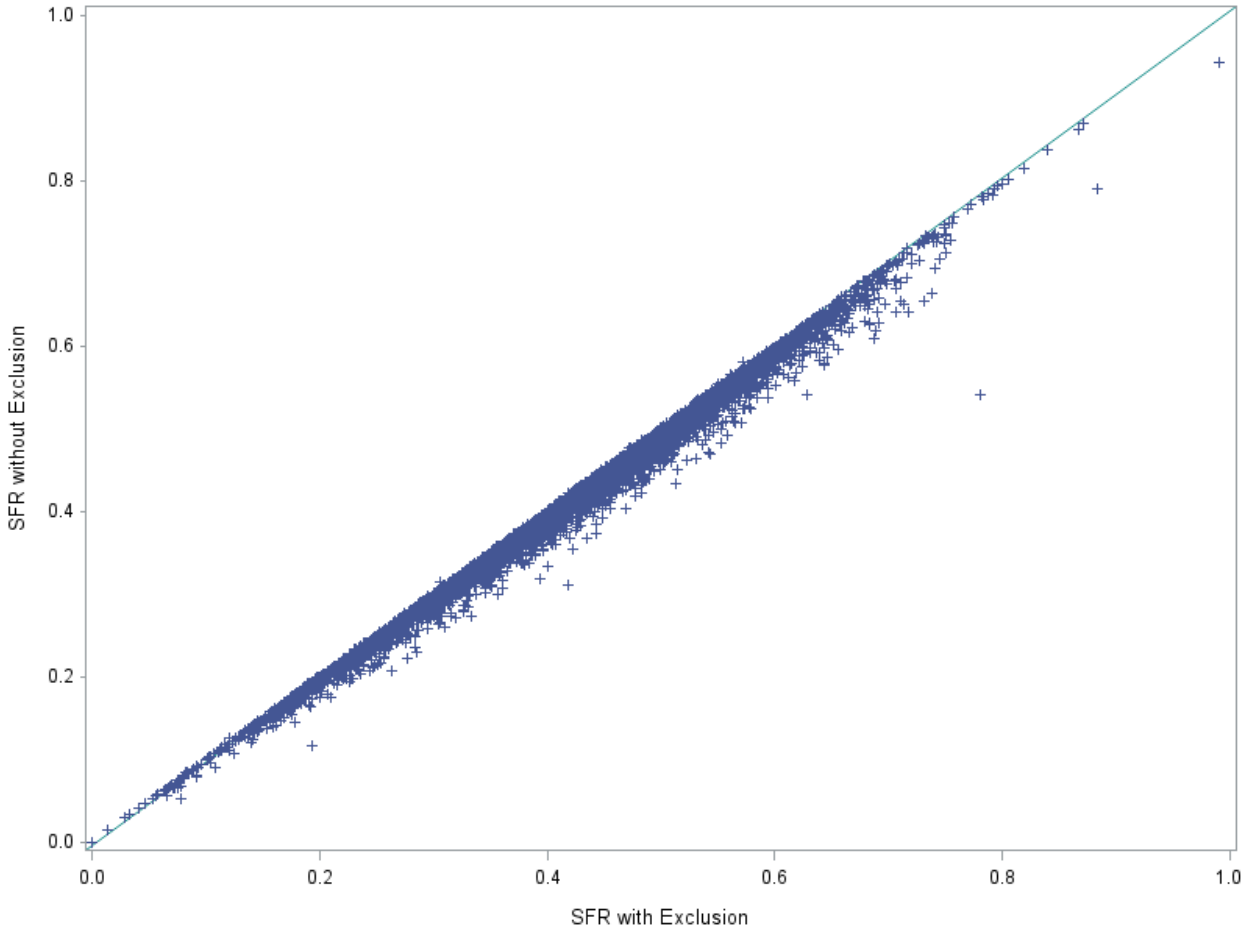
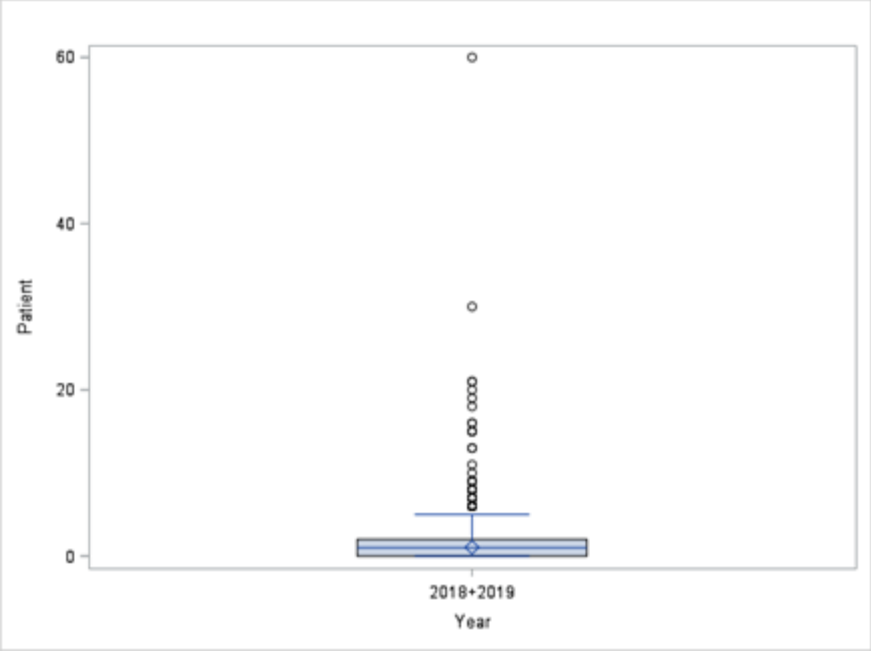


Figure 2. Distribution of Excluded Patients at facility level for 2018-2019



[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

[Response Begins]

Using 2018-2019 data, we show 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, figure 2). Due to the unequal distribution across facilities, the exclusion criteria take into account that some facilities treat a higher proportion of patients with limited life expectancy. Additionally, our results shown in both the scatterplot (Figure 1) as well as the Pearson Correlation Coefficient of 0.996 (p-value <0.0001) between Incident SFRs with and without the exclusion suggest that the overall impact of the exclusion on the measure's validity is not substantial since the two are highly correlated.

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

16 risk factors

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

The Incident 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 Incident 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 the risk factors are the same across all facilities. The common risk factor 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
    - 18-<25
    - 25-<59
    - 60-<75
    - 75+
- BMI at incidence
    - <18.5
    - 18.5-25
    - 25-30
    - >=30
- Nursing home status in the prior 12 months
    - No nursing home care: 0 days
    - Short term nursing home: 1-89 days
    - Long term nursing home: ≥90 days
- Nephrologist's care prior to ESRD (CMS-2728 form)
- Diabetes as primary cause of ESRD
- Comorbidities at ESRD incidence (CMS-2728 form)
    - Diabetes
    - Congestive heart failure
    - Other heart diseases
    - Peripheral vascular disease
    - Cerebrovascular disease
    - Chronic obstructive pulmonary disease
    - Drug dependence
    - Inability to ambulate/transfer at ESRD incidence
- Incident comorbidities: Incident comorbidity information was obtained from the CMS-2728. The covariates for comorbidities included in the final model take a value of 1 if there was any evidence of the condition from the CMS-2728, otherwise 0. We also use two binary indicators:
    - Indicator for missing a CMS-2728 form: this would signal an inability to capture incident comorbidities.
    - Indicator for having at least one of the ESRD incident comorbid conditions listed above: this would signal the possibility that a patient has a CMS-2728 form, but the comorbidity section could have been left blank.
- Medicare Indicator:
    - Medicare coverage for at least 6 months or more in the prior 12 months OR, Medicare Advantage coverage for one or more months in the prior 12 months. We used 1 month of MA because ESRD patients were restricted from enrolling in MA plans (prior to 2021). They already had to be Medicare

beneficiaries in an MA plan prior to becoming ESRD therefore they are considered as having a sufficient Medicare coverage history during this period.

[Response Ends]

2b.21. 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 (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

N/A

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature

Internal data analysis

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

Although there have been significant gains over the past two decades in the proportion of dialysis patients that have an AV fistula, these improvements remain markedly lower among incident ESRD patients. Approximately 75% of patients who are starting long-term dialysis do so with a tunneled catheter. While it is generally recognized that an AV graft or even a catheter is an appropriate vascular access for some incident patients who are starting hemodialysis, the majority will be good candidates for AV fistula creation. Given that there is variation in the burden of comorbidities between different facilities that impact the likelihood of having an AV fistula created, adjusting for these factors when calculating an AV fistula rate implicitly recognizes that some patients are more likely to have AV grafts. Ultimately, evaluation and selection of the clinical and patient risk factors for this measure was informed by the final recommendations made by a Vascular Access Technical Expert Panel. 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. Because this measure is restricted to incident ESRD patients, only incident comorbidities are used for adjustment. 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 ESRD incidence, nursing home status in the prior 12 months, nephrologist's care prior to ESRD, diabetes as primary cause of ESRD, and comorbidities at ESRD incidence. 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 Incident SFR were selected based on several considerations. We included a set of patient characteristics, including demographics, comorbidities at ESRD incidence, and other characteristics. Factors considered appropriate were then investigated with statistical models to determine if they were related to AVF use. Factors were also evaluated for face validity before being included.

Incident comorbidity information was obtained from the CMS-2728. The covariates for comorbidities included in the final model take a value of 1 if there was any evidence of the condition from the CMS-2728, otherwise 0. We also use two binary indicators: missing a CMS-2728 form; and an indicator for if at least one of the comorbidities were present. Additionally, we include a binary indicator to define patients with 6 months or more of Medicare coverage or one or more months of Medicare Advantage coverage, in the past 12 months. This was included because we observed an association between younger age and MA coverage status as a predictor of AVF use. The adjustment for nursing home status is measured by two binary variables. One for nursing homes stays of less than 90 days (short term), and the other to indicate nursing home stays of 90 days or more (long term) in the reporting period. Table 4 shows that almost all of the comorbidities had a statistically significant association with AVF use.

For the SES/SDS testing we investigated several patient and zip code level data elements (see description in 2a.8). Sociodemographic factors included in the analysis were based on conceptual criteria and empirically demonstrated patient-level findings in the literature which have shown differences in fistula use exist among racial minorities, women and the poor (Arya 2020; Shah 2018, Lin 2018). In addition, the particular patient and area level variables chosen were based on availability of data for the analyses. We use the publicly available Area Deprivation Index (ADI) originally developed by Singh and colleagues at the University of Wisconsin. We applied the updated ADI based on 2011-2015 census data. The ADI reflects a set of SES characteristics, including measures of income, education, and employment status, measured at the ZIP code level.

Arya et al. Racial and Sex Disparities in Catheter Use and Dialysis Access in the United States Medicare Population. JASN 2020. 31: ccc–ccc, 2020. doi: https://doi.org/10.1681/ASN.2019030274.

Shah et al. Gender and Racial Disparities in Initial Hemodialysis Access and Outcomes in Incident End-Stage Renal Disease Patients. Am J Nephrol 2018;48:4–14.

Lin et al. Health Insurance in the First 3 Months of Hemodialysis and Early Vascular Access. Clin J Am Soc Nephrol 13: ccc–ccc, 2018. doi: https://doi.org/10.2215/CJN.06660518.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

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 (Diabetes not as primary cause of ESRD, and at least one of the comorbidities), all main effects are statistically significant at the ≤ 0.05 level.

Table 4. Model Coefficients and Odds Ratios, using 2018-2019 data

Covariate	Coefficient	Odds Ratio	P-value
Age	**	**	**
18-<25	-0.326	0.722	0.002
25-<59	0.044	1.045	0.036
60-<75	reference	**	**
75+	-0.140	0.869	<0.001
BMI	**	**	**
Underweight (< 18.5)	-0.306	0.737	<0.001
Normal (18.5 - 24.9)	reference	**	**
Overweight (25-29.9)	0.167	1.181	<0.001
Obese (30+)	0.206	1.228	<0.001
Nursing home during the prior 12 months	**	**	**
No nursing home care (0 day)	reference	**	**
Short-term nursing home care (<90 days)	-0.621	0.537	<0.001
Long-term nursing home care (90+ days)	-0.600	0.549	<0.001
Nephrologist's Care prior to ESRD*	0.560	1.750	<0.001
Primary Cause of ESRD	**	**	**
Diabetes	0.112	1.119	<0.001
Other	reference	**	**
Comorbidities*	**	**	**
Diabetes (NOT as primary cause of ESRD)	-0.061	0.941	0.060
Congestive Heart Failure	-0.235	0.790	<0.001
Other Heart Diseases	-0.088	0.916	<0.001
Peripheral Vascular Disease	-0.070	0.933	0.028
Cerebrovascular Disease	-0.085	0.918	0.007
Chronic Obstructive Pulmonary Disease	-0.155	0.856	<0.001
Drug Dependence	-0.369	0.691	<0.001
Inability to ambulate/transfer	-0.491	0.612	<0.001
At least 6 months of Medicare covered months OR at least 1 month of MA covered months in prior 12 months	0.547	1.728	<0.001
Missing a CMS-2728 form	-1.487	0.226	<0.001
At least one of the comorbidities listed above	-0.063	0.939	0.068

*** "No" was used as the reference.**

**** This cell is intentionally left blank.**

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

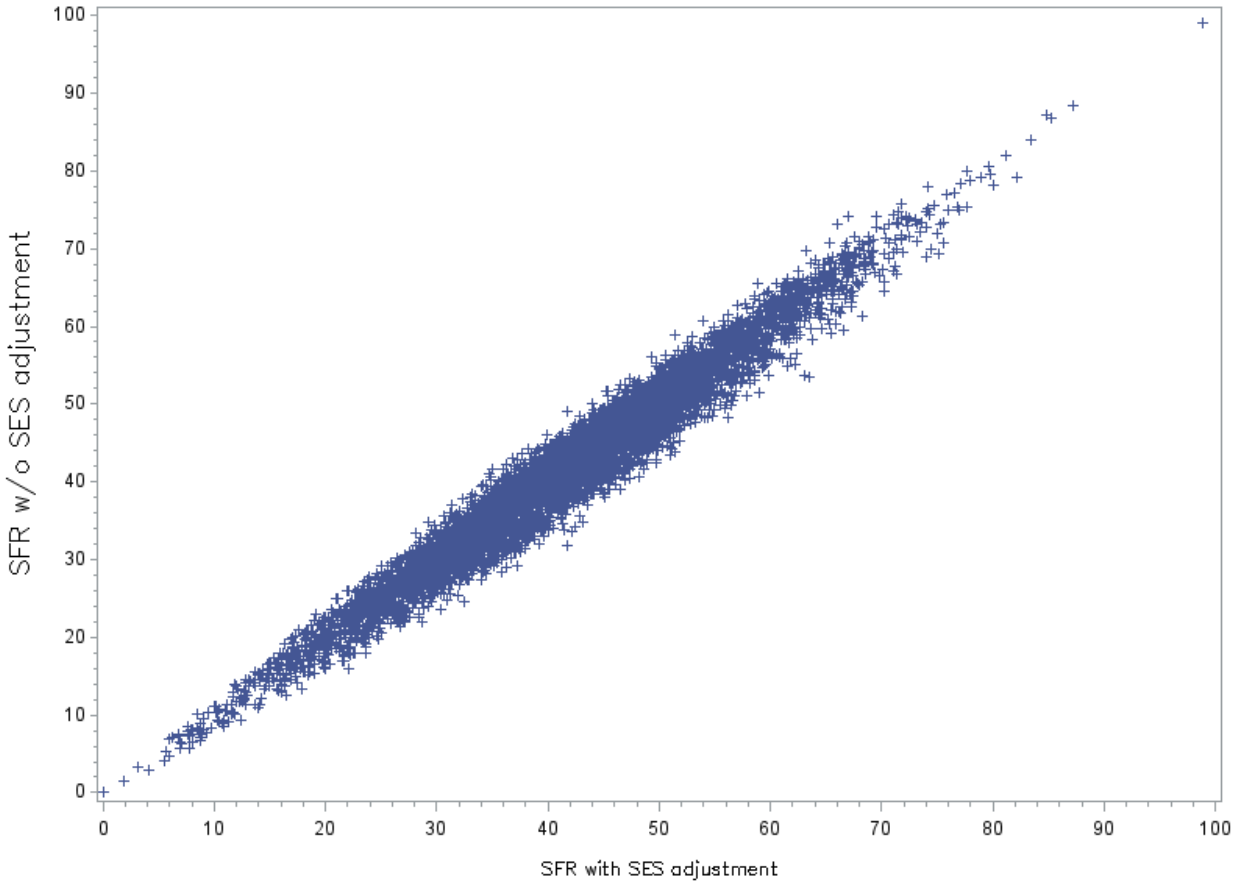
The table below shows the regression coefficient estimates and odds ratio for patient and area level SDS/SES variables based on a logistic regression model for AV fistula use that included all these variables along with all the other clinical covariates used for adjustment in the Incident SFR. Here we only report results for the SDS/SES factors.

Table 5. Coefficients and odds ratios for SDS/SES variables, using 2018-2019 data

Variable	Estimate	Odds Ratio	P-value
Sex	*	*	*
Female	-0.506	0.603	<0.001
Male	Reference	*	*
Ethnicity	*	*	*
Hispanic	0.100	1.105	0.019
Non-Hispanic	Reference	*	*
Race	*	*	*
White	Reference	*	*
Black	-0.241	0.786	<0.001
Other	0.122	1.130	0.030
Employment Status (2728)	*	*	*
Employed	Reference	*	*
Unemployed	-0.188	0.829	<0.001
Other	-0.214	0.807	<0.001
Medicare Coverage	*	*	*
Dual eligible	-0.021	0.979	0.462
Non dual eligible	Reference	*	*
ADI (zipcode_level)	*	*	*
National percentile ADI score	-0.001	0.999	0.288

*This cell is intentionally left blank.

Figure 3. Correlation between Incident SFRs with and without SDS/SES adjustment, Data year 2018-2019



The standard and SDS/SES-adjusted Incident SFRs were highly correlated at 0.98 (p<.001).

Table 6. Comparison of performances with vs. without adjusting for SDS/SES factors, using 2018-2019 data

*	*	**Incident SFR with SDS/SES**	**Incident SFR with SDS/SES**	**Incident SFR with SDS/SES**	**Incident SFR with SDS/SES**
*	*	**Better than expected**	**As expected**	**Worse than expected**	**Total**
Incident SFR w/o SDS/SES	**Better than expected**	161 (2.4%)	31 (0.5%)	0	192
Incident SFR w/o SDS/SES	**As expected**	43 (0.7%)	6083 (91.3%)	35 (0.5%)	6161

*	*	**Incident SFR with SDS/SES**	**Incident SFR with SDS/SES**	**Incident SFR with SDS/SES**	**Incident SFR with SDS/SES**
Incident SFR w/o SDS/SES	**Worse than expected**	0	36 (0.5%)	275 (4.1%)	311
Incident SFR w/o SDS/SES	**Total**	204	6150	310	6664

*This cell is intentionally left blank.

After adjustment for SDS/SES, 145 facilities (2.2%) changed performance categories. 66 (1.0%) facilities were downgraded, and 79 (1.2%) facilities were upgraded.

Black race, female sex, unemployment at ESRD incidence, and dual-eligible status were all associated with lower odds of having an AVF indicating that patient-level, but not area-level, variables for SDS/SES have some impact on expected performance of the incident SFR. Furthermore, we observed that adjustment for SDS/SES only minimally shifted facility performance, with slightly more facilities declining in performance ranking with SDS/SES adjustment than improved.

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. Adjusting for these factors could have the unintended effect of masking or reinforcing disparities in vascular access outcomes. 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].

Finally, area-level factors are not included as an adjustment due to the absence of clinically meaningful or statistically observed differences on the fistula rate with these adjustments.

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.

[Response Ends]

2b.26. 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

See sections 2b.20 and 2b.27-2b.29, which describe the statistical methods used to develop and validate the model. Risk factors were selected for the final model based on the magnitude of the coefficients, evaluation of their statistical and clinical significance, and the model c-statistic.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

The C-statistic was 0.748. This indicates that the model correctly ordered 75% of the pairs of patient-months that were discordant with respect to the response variate.

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

The Hosmer-Lemeshow test statistic based on deciles of risk is 9.7 with p-value=0.29. The c-statistic and risk decile plot show that the model provides an overall good fit to the data.

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

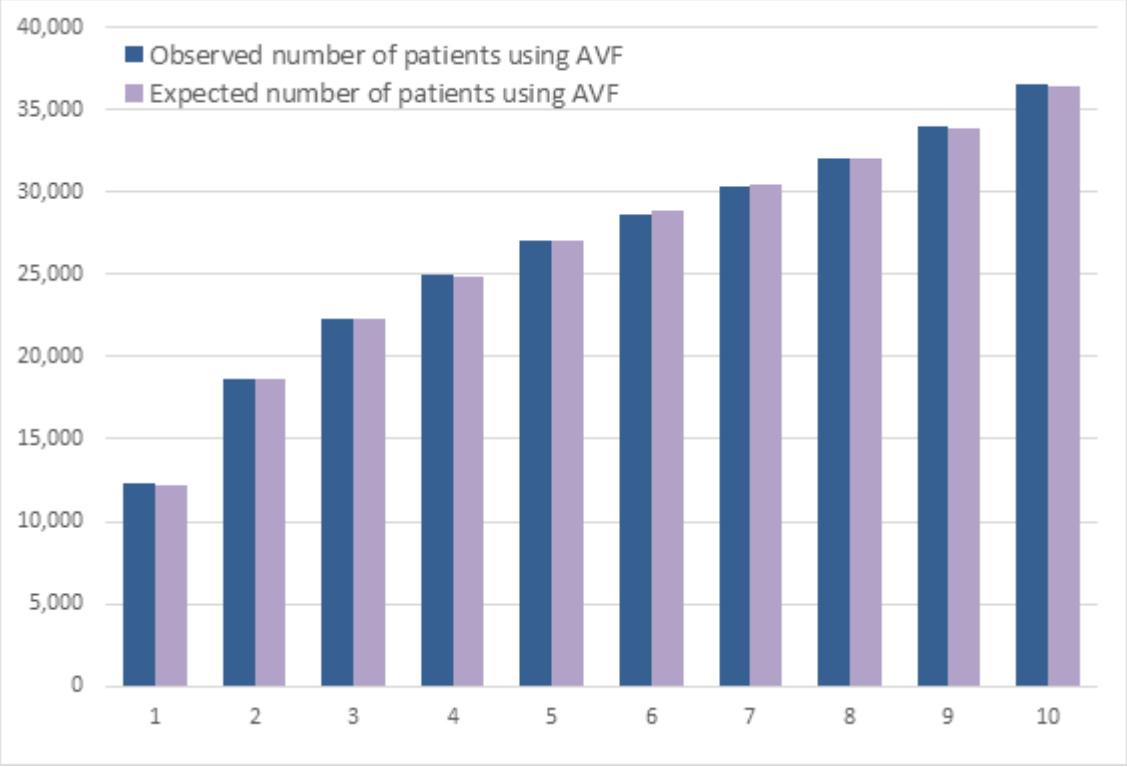


Figure 4: Decile plots for the number of patients using AVF, using 2018-2019 data

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

The decile plot (Figure 4) shows that the risk factors in the model are discriminating well between patients.

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

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).

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

N/A

[Response Ends]

Criterion 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.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in a combination of electronic sources

[Response Ends]

3.03. 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 data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

None Identified

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]

N/A

[Response Ends]

Criterion 4: Use and Usability

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.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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 demonstrating performance improvement.

4a. Use

4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]

Not in use

[Not in use Please Explain]

Not currently in use

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Payment Program

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

The measure is undergoing initial endorsement review.

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

[Response Begins]

CMS will determine if/when to report this measure in a public reporting/payment program. Potential applications for the measure include the ESRD Quality Incentive Program (ESRD QIP) or the Dialysis Facility Care Compare website.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

Facility level results have not been disseminated to those being measured as part of the development process for the new incident SFR.

Results of the previous iteration of this measure (SFR, formerly NQF #2977) are currently reported on Dialysis Facility Care Compare, and the ERSD Quality Incentive Program. All Medicare-certified dialysis facilities are eligible for reporting in both programs (approximately 7,000 dialysis facilities). Each program has a helpdesk and supporting documentation available to assist with interpretation of the measure results.

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

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

Facility level results have not been disseminated to those being measured as part of the development process for the new incident SFR.

How results were provided for SFR:

For DFCC, the results are first reported to facilities via a closed preview period, where facilities can review their data prior to each of the quarterly updates of the public facing Dialysis Facility Care Compare website. These preview reports are posted on dialysisdata.org, where facilities can also find a detailed Guide to the Quarterly Dialysis Facility Care Compare

Reports and other supporting documentation. Facilities can submit comments/questions about their results at any time and can request patient lists for their facilities during the specified preview periods.

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

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

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

iSFR is not yet implemented, and results have not been disseminated.

How feedback is gathered for SFR:

For DFC, feedback can be provided any time through contacting the dialysisdata.org helpdesk. Preview periods allow for specific times for facilities review and comment on measure calculations and provide an opportunity to request a patient list.

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

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

Feedback on SFR:

Comments received during DFC preview periods tend to be technical nature, asking for clarification on how the SFR is calculated for particular facilities, including questions about patient assignment and requests for confirmation of patient vascular access type in a specific month. UM-KECC investigates all inquiries received about specific patients and works with facilities to ensure that they understand their measure results and that data discrepancies are resolved.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

For SFR: Since the SFR was first proposed in the ESRD QIP PY 2021 proposed rule, several commenters requested that this measure account for situations for which the patient has elected not to have a fistula (patient choice/preference). CMS also received comments about facilities possibly being doubly penalized if they have low fistula rates, and high catheter rates, and also do not get credit for grafts. Comments were also received on additional clinical risk adjustors and exclusions, including exhausting all other vascular access options.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

iSFR is a refinement of its predecessor, SFR, based on stakeholder feedback. By limiting SFR to the first 12 months of dialysis, iSFR avoids the prior issue of exhausted vascular access that was raised by numerous stakeholders. We have been unable to incorporate patient choice into the measure since there is no validated measurement of patient choice with regards to vascular access.

[Response Ends]

4b. Usability

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. 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.

[Response Begins]

See Importance to Measure and Report for data on performance gap and disparities.

The incident SFR is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of this measure after endorsement review. Once implemented, facility performance on this measure can be evaluated to determine if the measure has supported and detected quality improvement in incident fistula rates, as well as whether there are changes in disparities among groups that typically have lower fistula rates, suggesting potential racial, ethnic, and gender disparities.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

None, as the measure is not yet implemented

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

None, as the measure is not yet implemented

[Response Ends]

Criterion 5: Related and 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.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

2594: Optimal End Stage Renal Disease (ESRD) Starts

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

N/A

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

No

[Response Ends]

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

[Response Begins]

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.

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

There are no competing measures.

[Response Ends]