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 #: 2594

Measure Title: Optimal End Stage Renal Disease (ESRD) Starts

Measure Steward: The Permanente Federation

Brief Description of Measure: Optimal End Stage Renal Disease (ESRD) Starts is the percentage of new adult ESRD patients during the measurement period who experience a planned start of renal replacement therapy by receiving a preemptive kidney transplant, by initiating home dialysis (peritoneal dialysis or home hemodialysis), or by initiating outpatient in-center hemodialysis via arteriovenous fistula or arteriovenous graft.

Developer Rationale: Unplanned onset of ESRD has been termed "Catastrophic Onset of ESRD" and is associated with long hospitalizations and high mortality in the first year of dialysis. Unprepared and uneducated patients must be admitted to the hospital and receive urgent dialysis, usually via a central venous catheter which has been shown to cause high rates of blood stream infection and higher mortality than surgically prepared vascular access. The Optimal ESRD Starts measure quantifies the ability of a health care system to prepare patients for ESRD by identification of high-risk patients, educating them and their families about the need for dialysis or kidney transplant, helping them to make appropriate and informed choices, and then successfully transitioning to dialysis or kidney transplantation as kidney function declines to the level of ESRD.

As demonstrated in the Medical Evidence Review, Optimal ESRD Starts are better for patients, who experience fewer blood stream infections, have fewer hospital days, fewer cardiovascular events and lower mortality. Optimal ESRD Starts save resources as shown by the Canadian STARRT Trial (average estimated cost $23,965 less per patient in the first 6 months of ESRD) and by KP data (average costs $47,000 less per patient in the first 6 months of ESRD with 14.1 fewer hospital days in first 6 months of ESRD - data based previous submission). This measure has been utilized successfully to improve outcomes in Kaiser Permanente Southern California for nearly 15 years and in the national Kaiser Permanente program for 8 years.

There is a significant performance gap in Optimal ESRD Starts between the estimated U.S. performance, estimated at 30.4% in 2017 as estimated from United States Renal Data Service and CMS Fistula First data and the Kaiser Permanente national performance at 57.5% in December 2017 and 56.5% in September 2021, demonstrating the opportunity for improvement.

Numerator Statement: The number of new ESRD patients age 18 and over who initiate outpatient renal replacement therapy in the twelve month measurement period with an optimal ESRD therapy, which includes preemptive kidney transplant, home dialysis (peritoneal dialysis or home hemodialysis), or outpatient in-center hemodialysis via arteriovenous fistula or arteriovenous graft.

Denominator Statement: The number of patients age 18 and over who receive a preemptive kidney transplant or initiate long-term dialysis therapy (do not recover kidney function by 90 days) for the first time in the twelve month measurement period

Denominator Exclusions: None

Measure Type: Process

Data Source: Claims

Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: 10/02/2015

Most Recent Endorsement Date: 10/02/2015

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a ***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 maintenance process measure at the facility level of analysis that assesses the percentage of new adult ESRD patients during the measurement period who experience a planned start of renal replacement therapy by receiving a preemptive kidney transplant, by initiating home dialysis (peritoneal dialysis or home hemodialysis), or by initiating outpatient in-center hemodialysis via arteriovenous fistula or arteriovenous graft.
- The developer provides a logic model that depicts the process of Optimal End Stage Renal Disease (ESRD) Starts leads to improved patient outcomes (i.e., lower mortality and cardiovascular risk, less hospitalizations and lower costs, and higher quality of quality of life).

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 of prior review in 2015

- The developer provided evidence to support the assertion that unplanned onset of ESRD has been termed "Catastrophic Onset of ESRD" and is associated with long hospitalizations and high mortality in

the first year of dialysis. The developer provided evidence to support the assertion that unprepared and uneducated patients must be admitted to the hospital and receive urgent dialysis, usually via a central venous catheter which has been shown to cause high rates of blood stream infection and higher mortality than surgically prepared vascular access.

- The developer provides data to support that Optimal ESRD Starts save resources as shown by the:
    - Canadian Standard versus Accelerated Initiation of Renal-Replacement Therapy in Acute Kidney Injury (STARRT-AKI) Trial data - average estimated cost is $23,965 less per patient in the first 6 months of ESRD)
    - Kaiser Permanente (KP) data - average cost is $47,000 less per patient in the first 6 months of ESRD with 14.1 fewer hospital days in first 6 months of ESRD.

Changes to evidence from last review

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

☒ The developer provided updated evidence for this measure:

- The developer references Kidney Disease: Improving Global Outcomes (KDIGO) Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation. The developer cites a recommendation that patients with chronic kidney disease (CKD) who are expected to reach end-stage kidney disease (ESKD) should be informed of, educated about, and considered for kidney transplantation regardless of socioeconomic status, sex, gender identity, or race/ethnicity. (Grade D: Very low)
- The developer notes a recommendation from The European Society for Vascular Surgery which recommends an autogenous arteriovenous fistula as the primary option for vascular access.
- The Vascular Access for Haemodialysis recommends that all patients with ESKD who commence hemodialysis or are on long-term hemodialysis should dialyze with an arteriovenous fistula as first choice, an arteriovenous graft as second choice, a tunneled venous catheter as third choice and a non-tunneled temporary catheter as an option of necessity. (Grade 1a: Grade 1 as a "strong" expert recommendation and indication of the balance of benefits, risk, burden and cost and Grade A as a "high" indication of study design, directness of evidence and consistency of results)
- The Canadian Society of Nephrology recommends the following: each center should establish a dedicated team for vascular access; preserve arm veins suitable for placement of vascular access; preservation should begin in patients with progressive kidney disease and an estimated GFR of less than 30 ml/min; radio-cephalic native vessel arteriovenous fistula is the preferred type of vascular access; establish AV fistulae when the patient has an estimated GFR of 15 to 20 ml/min and progressive kidney disease.
- The developer cited a systematic review looking at a total of twenty-one studies with 29,000 participants supported the placement of Peritoneal Dialysis in ESKD patients.

Exception to evidence

- N/A

Questions for the Committee:

- *How strong is the evidence for this relationship?*
- *What is the relationship of this measure to patient outcomes?*
- *Is the evidence directly applicable to the process of care being measured?*
    - *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review.*

- *Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer includes measure performance data over six consecutive annual measurement periods that demonstrates that the KP national mean of the annual measure rate has improved from 57.1 percent in December 2015 to 58.3 percent in December 2020 but saw a slight decline at 56.5 percent in September 2021.
- During October 1, 2020 to September 30, 2021 (the most recent measurement period) the developer's data demonstrates 4,002 new ESRD patients, ranging from 103 to 1,806 patients in the eight measured Kaiser Permanente regions, with regional minimum being 42 percent and the maximum being 65 percent.
    - The developer provided measure performance rates for the U.S. population, which can be estimated from United States Renal Data System (USRDS) and CMS Fistula First data and was approximately 30.4 percent for 2017.
    - The developer states that there is a significant performance gap that exists between this U.S. performance and data analyzed at the KP regions

Disparities

- The developer used its internally sourced data to provide optimal restarts by race and ethnic categories between Q4 2017 and Q3 2021.
    - The data findings indicate that asian-pacific islanders consistently had the highest percent of optimal starts (63-69 percent) and black patients had the lowest performance (51-61 percent). White (56-59 percent) and hispanic patients (55-57 percent) performed in the middle range.

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

- Updated evidence consists of society recommendations
- Evidence includes studies showing clinical benefit of kidney transplantation compared to dialysis; UK Renal Association Vascular Access guidelines (2015), Clinical Practice Guideline of the European Society for Vascular Surgery (2018), and Canadian Society of Nephrology CPG; and systematic review and meta-analysis examining quality of life associated with PD compared to HD. Summary

of updated evidence does not include most recent KDOQI Vascular Access Guideline, however, the updated KDOQI Vascular Access Guideline is referenced in the reliability section and influenced changes in measure specifications.

- Significant number of new studies provided published since first endorsement providing support to care components considered "optimal" start in this measure.
- Evidence remains strong from initial 2015. No new information that disputes or changes the evidence.
- The developer provides narrative and a logic diagram that clearly outlines how each measure component within the measure is connected to patient health outcomes. However, it appears that the level of analysis is really at the health plan level rather than the facility level, as stated. It should be noted that the measure is only feasible in fully integrated delivery care systems or large physician groups and is not applicable to dialysis facilities because of potentially insufficient patient volumes.
- Maintenance process measure on the facility level. It is based on systematic review and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured measure derived from patient reports that the target population values the measure process or structure and finds it meaningful. Optimal ESRD starts lead to improved patient outcomes i.e. lower mortality and cardiovascular risk, less hospitalizations and lower costs, and high quality of life. Systematic review of the evidence is specific to the measure. Quality quantity and consistency of evidence was provided and evidence was graded. Developer provided evidence to support the assertion that unplanned onset of ESRD has been termed catastrophic onset of ESRD---long hospitalizations and high mortality in 1st year of dialysis.
- Patients do better when ESKD care is planned in a proactive fashion rather than reactive fashion.
- I am not aware of any new investigations that alter this measure's evidence base.
- High.
- How does the evidence relate to the specific structure, process, or outcome being measured? Not sure. Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? Good for maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? No.
- This measure was initially endorsed 10/2/2015 and is up for re-endorsement. The data for this measure is derived from claims.
- This measure improves patient outcomes and reduces cost.
- There are numerous and strong multi-dimensional data which support the superiority of an optimal start to dialysis. This is a fertile area for quality improvement. Unfortunately as structured, it is unlikely that this measure will substantially improve or optimize the initiation of dialysis. Dialysis facilities do not care for patients prior to the initiation of dialysis unless they are part of an integrated delivery system, or own, or joint venture with, the nephrology practices that are providing the upstream care for the CKD patient. Thus, it is difficult for a dialysis facility, itself, to have a sustained or appreciable impact on the pre-ESRD care dialysis. This fact is evidenced by the data provided by the measure developer (line graft demonstrating KP optimal ESRD starts). Even within the KP system, which is an integrated delivery system, the percent of optimal starts though better than the US average has minimally improved between 2015 and 2021. Additionally, a substantial number of patients are assigned to a new nephrologist when they initiate dialysis. This fact further weakens this measure, as designed, as an effective quality measure.
- Process measure, data source - claims, systematic review. Evidence graded, quality, quantity and consistency of evidence graded. KDIGO, European Society for Vascular Surgery - AV fistula,

Vascular Access for Canadian Society of Nephrology recommended. High rating for evidence. No new studies presented.

- True.
- High.
- The developer describes a maintenance process measure that attempts to analyze the percentage of patients who experience a planned start date of RRT whether in the form of a kidney transplant, or home dialysis (PD or Hemo), or by initiating outpatient in center HD rather than admitted to the hospital and receive urgent dialysis and usually via a catheter.
- The evidence matches what is being measured directly. Not aware of any new evidence.
- New studies reviewed and increase validity of the measure.
- The empirical data support that optimal starts can be improved but this measure is aimed at the wrong audience: it needs to target nephrologists/nephrology practices not dialysis providers.

1b. Performance Gap

- Yes given 57.5 % fistula rate for KP patients v 30% US. Despite measure, no consistent continued improvement at KP. No major disparities evident.
- Performance data within Kaiser from 2015 - 2021 included; the US data is from 2017. Differences in performance based on race (Kaiser data).
- Significant gap between entire US (30% with "Optimal Start") vs KP system (57-58%); also regional differences provided; disparity data given for race and ethnicity.
- More recent 2021 data was provided. Comparison to 2017 national data. Gap still exists. Population disparities.
- Developer provides evidence of a clear and persistent performance gap and considerable racial and ethnic variations.
- Measure performance data over six consecutive annual measurement periods that the KP national mean of the annual measure rate has improved from 57.1 in 12/2015 to 58.3 % in 12/2020 but saw a slight decline at 56.5% in 9/2021. Developer states that there is significant performance gap that exists between this us performance and data analyzed at the KP regions. Disparities data indicate that Asian pacific islanders consistently had the highest percent of optimal starts 63-69 percent; black patients had the lowest performance 51-61%. White 56-59 % and Hispanic patients 55-57 % performed in middle range.
- Optimal starts may differ by race; there is room for some improvement.
- Current performance of the data was provided along with disparities.
- Moderate.
- What is the patient's interpretation of their care/treatment?
- There are opportunities for improvement i.e. US population was approx. 30.4% in 2017. KP national mean has improved to 58.3 in Dec 2020 but saw decline in 2021 to 56.5. Still demonstrating higher than national average.
- Data described poor preparation for dialysis, worse among Black people.
- Moderate.
- Maintenance measure - increased emphasis on gap and disparities. KP national mean of the annual measure rate has improved from 57.1% 12/2015 to 12/2020 to 58.3%. Slight decline in 9/2021. Significant performance gap exists between US performance and data analyzed at KP regions. Asian pacific islanders 63-69% optimal starts. Followed by white patients, Hispanic and blacks rating the lowest 51 to 59%.
- True.

- Moderate.
- Current performance gap was provided and showed a gap for race.
- Based on current data, the data does pinpoint a gap in care on a regional basis. The African American community was also identified as lowest performance group.
- 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.
- Data is provided and supports that there is a gap in care.

Criteria 2: Scientific Acceptability of Measure Properties

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

Evaluators: NQF Staff

2a. Reliability: Specifications and Testing

For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

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

For maintenance measures – less emphasis if no new testing data provided.

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

Specifications:

- Measure specifications are clear and precise. The developer indicates slight changes to reporting specifications since 2015 but attests that those changes are non-consequential. Those changes include:
  - Removal of the criterion of 10 percent new hemodialysis patient limit for arteriovenous graft (AVG). Developer made this change based on recommendations from the Kidney Disease Outcomes Quality Initiative (KDOQI) 2019 Vascular Access guidelines.
    - The developer explains that removal of this criterion did not have significant impact on the reported AVG rate over time.
  - Inclusion of patients with failing kidney transplants starting or returning to dialysis. Developer made this change based on various publications that examined the high number of failed transplant patients who returned to dialysis post-transplant.
    - The developer explains that this group of patients constituted about three percent of the measure denominator population and therefore did not have significant impact on the reported measurement rate.
- Developer asserts that the data elements and collection process have not changed since the last NQF measure submission.

Reliability Testing:

- The developer conducted validity testing at the patient or encounter level; additional reliability testing is not required.

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 developer attests the specifications have not significantly changed and that additional reliability testing was not conducted. Does the Committee agree that the measure is still reliable and there is no need for repeat discussion and vote on Reliability?*

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

2b. Validity: Validity testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

For maintenance measures – less emphasis if no new testing data provided.

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 Patient/Encounter Level:
    - The developer utilized data element testing from their 2015 data element testing.
    - The developer tested the accuracy of the regional data compared to the authoritative source.
        - The renal replacement therapy information submitted by the regional care coordinator was compared to that provided by the renal replacement therapy provider on record. Among the data element elements, the developer tests it's a priori hypothesis of a 0.9 match (90 percent accuracy) between the region and source data.
        - The developer calculated the denominator data elements to be 96 percent (87-99 percent, 95CI) accurate.
        - The developer calculated the numerator data elements to be 87 percent (76-92 percent, 95CI) accurate.
        - Total element match rate was 83 percent (71-91 percent, 95CI). Based on this data, developer affirmed a positive predictive value of 0.94 to identify true optimal ESRD start, negative predictive value of 0.79 to identify a non-optimal start.
        - The developer reported that the region correctly identifies a true optimal ESRD start (positive likelihood ratio) 11.6 times more often than it incorrectly identifies a non-optimal ESRD start as optimal and that this is more than double the generally accepted rule that a test should have a ratio greater than 5. The developer also reported the ratio of reporting false negatives (negative likelihood ratio) (LR-), of 0.19 which meets the generally accepted threshold that a LR- should be less than 0.2.
    - Two changes were made to the measure specifications since the last submission:
        - One changed included, removal of criteria that 10 percent new hemodialysis patient limit for arteriovenous graft (AVG). The developer assessed the impact of removing this criterion and found that the percent AVG in 2016 (before removal of 10 percent limit) and 2021 (after removal of 10 percent limit) remained below the 10 percent AVG limit previously set in place. The developer advised that this suggests that the percent AVG have not been impacted that much by removal of this limiting criterion.
        - The second change included patients with failing kidney transplants starting to return to dialysis. The developer assessed the impact of including these patients and found

that there was a small difference of 3.3 percent in the denominator with the criterion versus the denominator without the criterion, with a range of differences from 0 percent to 5 percent in the eight markets.

- Validity testing conducted at the Accountable Entity Level:
    - The developer states that face validity was conducted but the method described by the developer does not match NQF requirements for face validity.

Exclusions

- Exclusions include:
    - Patients who are younger than 18 years old
    - Patients who are not a KP member at the start of the renal replacement therapy
    - Patients who recover GFR to the point that they are off dialysis by day 90 after the first outpatient dialysis
- In 2017, an exclusion criterion related to the 10 percent hemodialysis patient limit for AVG was removed.
    - The developer indicated that there was no change to the impact on the results because none of the regions reached the 10 percent AVG limit.
- The developer provides data on the previous patient limit exclusion but does not provide data on the other exclusions.

Risk-Adjustment

- Measure does not use risk-adjustment.

Meaningful Differences

- The developer compared the regional performance to identify statistically significant and meaningful differences in the optimal ESRD start metric. The developer used the chi-square test method to test if there are any regional differences among all regions' performance.
    - The developer employed a second set of tests to determine regional difference by comparing each regional optimal ESRD start proportion to all regions' proportion and used a 95 confidence interval to determine statistical significance.
    - The developer indicates that in 2015, for the chi-square test, the calculated statistic was 29.73 with 5 degrees of freedom and a statistically significant p-value of less than 0.0001.
    - The developer provides updated statistical analysis using 2021 data and still produces a statistically significant p-value of less than 0.05.
- Results from the developer's comparison of regional performance indicate that the optimal ESRD start proportions vary by region (.282-.567) and the regional difference to all regions group mean difference can be statistically significant (Region 1, difference = -.224, significant using 95 percent confidence level). Developer posits that identification of performance differences is possible with the optimal ESRD start metric.

Missing Data

- The developer provided information from their 2015 submission on missing data.

- The developer found missing data amongst the numerator data elements (i.e., method of renal replacement).
    - To quantify the effect of missing data, the developer calculated the performance metric under two scenarios, which represent the two extremes for the missing data.
        1. For the excluded cases, the region data elements and optimal ESRD start result are assumed not to match the source.
        2. Where the region data elements and optimal ESRD Starts result are assumed to match the source.
    - The largest change is seen in the positive likelihood ratio where the observed ratio is 12 compared to 6 in the 'As error' scenario. For all the other test metrics, the differences are much smaller.
- Using the simple comparison of confidence intervals, all the 'As error' and 'As match' confidence intervals overlap within the respective columns. The developer indicates that there is no statistical difference between the 'As error' and 'As match' value.

Comparability

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

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Are there any concerns regarding the lack of exclusion testing?*
- *Does the updated information on the AVG cap and failing kidney transplants returning to dialysis provide adequate information on the reconstituted measure's validity?*
- *The developer attests that additional validity testing was not conducted. Does the Standing Committee agree that the measure meets NQF's validity requirements and there is no need for repeat discussion and vote on Validity?*

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

Committee Pre-evaluation Comments:

2a. Reliability - Specifications

- None.
- Modification to specifications: 1) no maximum percentage of AVGs (to be more aligned with update to KDOQI Guidelines); and 2) inclusion of patients with a failing kidney transplant returning to or starting dialysis. Clarify how transition from AKI to ESKD handled in the measure. In description HHD with AVF/AVG specified, but in measure calculation there is no specification of vascular access type for HHD – clarify if HHD access type must be AVF/AVG.
- Specifications clear and no concerns about satisfactory implementation.
- No risk adjustments. No concerns.
- Specifications, logic, and calculation algorithm are clearly defined.
- Developer asserts that the data elements and collection process have not changed since the last NQF measure submission. Measure specifications are clear and precise. Slight changes to reporting specifications since 2015 but states that these changes are non-consequential. 1.

removal of 10 % new hemo patient limit for AVG. Changes made based on KDOQI 2019 Vascular Access guidelines. Did not have significant impact on the reported AVG rate over time. 2. Inclusion of patients with failing kidney transplants starting or returning to dialysis. Made changes based on various publications that examined the high number of failed transplant patients who returned to dialysis post-transplant. Developer conducted validity testing at the patient level, additional reliability testing no required.

- In organizations like KP, you have more control of patients in your organization. Outside of KP, dialysis centers may be getting patients referred from a variety of sources and the dialysis nephrologist may not be involved in care of the patient prior to ESKD care. It is difficult to have influence over patients that are not your primary patients from the start. As such, it may be difficult to consistently implement a measure like this as it could potentially penalize the ESKD provider.
- All of the data elements are clearly defined. There are no concerns with the implementation of this measure
- Moderate.
- It gives the providers interpretation; however, what is the patient's perspective?
- No changes according to developer from last NQF measure submission.
- Clear definitions and implementation.
- Low.
- Measurements specifications are clear and precise. Developer stated slight changes in reporting specifications, but changes are non-consequential. For instance, -removal of criterion of 10% new hemodialysis patients with AVG. Recommended by KDOQI 2019.
- True
- Implementation happens before dialysis starts. It has never made sense to me for this to be facility-level metric. It would be useful as a physician-level metric pre-dialysis.
- Data elements are clearly defined.
- Did not identify any issues with the specifications.
- I believe this measure has moderate reliability and that the measure can be fairly consistent in implementation.
- Data is appropriately specified but there are exclusions that are warranted: AKI, limited life expectancy.

2a2. Reliability – Testing

- ESKD incidence over time per market area was relatively flat. Data are validated.
- In feasibility section, minimum of 50 new ESRD patients is recommended due to concerns about random variation. Is this threshold based on testing?
- no concerns.
- No concerns.
- N/A; because the developer conducted validity testing at the patient or encounter level, additional reliability testing was not required.
- No concerns about reliability of measure.
- No concerns.
- There are no Reliability concerns.
- No.
- Shouldn't it include patient feedback?

- No concerns on previous reliability testing that was completed.
- No
- No
- No
- Appropriate
- No
- I have no concerns about the reliability of the measure.
- No issues.
- No
- No

2b1. Validity – Testing

- No.
- Recommend committee discussion of validity testing.
- Validity testing at patient encounter level done with satisfactory accuracy of data components
- No concerns.
- The developer tested the accuracy of the regional data compared to the "authoritative source" by comparing information submitted by the regional care coordinator to that provided by the provider on record. Denominator data elements were calculated to be 96% (87%-99% 95 CI) accurate; numerator data elements were 87% (76%-92% 95 CI) accurate, and the total data element match rate was 83% (71%-91% 95 CI). The developer calculated a positive predictive value of 0.94 to identify true optimal ESRD starts and a negative predictive value of 0.79 to identify a non-optimal start. No concerns.
- No
- No concerns.
- There are no validity concerns.
- No
- No
- Developer compared data submitted by the regional care coordinator to data in the medical record. Results are offered in the materials for both numerator and denominator.
- No
- Low
- No. Prior hypothesis was 90% accurate between region and source area. Denominator data elements was 96% accurate. Nominator data elements 87% accurate. Ratings for validity is moderate.
- Appropriate
- No
- I have no concerns about the testing results.
- The developers provided the basis for their updated information on AVG caps and failing transplant patients which I found adequate.
- No
- No

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

- Not applicable.

- No risk adjustment. Would be helpful to review denominator clarifications as some of these do appear to be exclusions.
- no concerns.
- Patients to be appropriately excluded, as in the 2015 comments from ASN and KCP, are not clear to me.
- No concerns with exclusions. The measure is a process metric and is not risk adjusted.
- Measure does not use risk adjustment. Developer provides data on the previous patient limit exclusion but does not provide data on the other exclusions.
- No concern.
- There are no risk adjustment concerns.
- No concerns.
- Unknown
- New exclusions included removal of 10 percent new hemodialysis patient limit for AVG with minimal impact. There is no risk adjustment for this measure.
- My only concern is the ease with which dialysis providers not within integrated health organizations can discount this measure, though that does not reflect a problem with the measure but with the dialysis providers.
- See above.
- Health outcome---potential social risk factor variables and the measure focus.
- True
- Here again, "start of care" for appropriate starts precedes transplant, home dialysis or vascular access type.
- Exclusions are appropriate.
- The exclusions are appropriate.
- The exclusions are appropriate. I feel the social risk factor variable are appropriate and risk adjustments are in line with common practice. The results of analysis were acceptable.
- As above, exclusions would be appropriate for AKI and limited life expectancy.

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

- None
- Recommend committee discussion of validity testing.
- No identified threats to validity; meaningful differences can be detected by measure (regional range in KP system).
- No threat, minimal missing data.
- The developer compared regional performance to identify statistically significant and meaningful differences, using the chi-square test method to test for regional differences in performance. A second set of tests was employed to determine regional difference by comparing each regional optimal ESRD start proportion to all regions' proportion and used a 95% confidence interval to determine statistical significance. Results indicate that the optimal ESRD start proportions vary by region (0.282-0.567) and the regional difference to all regions' group mean difference can be statistically significant (Region 1, difference = -0.224, significant using 95 percent confidence level).
- No
- No concerns.
- There are no concerns with threats to the validity of this measure.

- No
- Yes
- There is a concern noted that the region correctly identifies a true optimal ESRD start 11.6 times more often than it incorrectly identifies a non-optimal ESRD start. This is greater than double the expected ratio of 5.
- Identifies meaningful differences. I am not concerned about missing data.
- Unlikely
- No
- Yes
- No
- I have no concerns about threats to validity.
- They recognize that differences can occur across regions and the data support this. Missing data is addressed appropriately.
- Threats to validity: it doesn't meet the NQF requirements. Meaningful differences: the developer tested more current data and still came up with statistically
- It is not clear that the measure addresses aspects of care that can be changed.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- The developer attests that the data is generated or collected by and used by healthcare personnel during the provision of care and is abstracted from a record by someone other than person obtaining original information.
- The developer identifies administrative claims, electronic clinical data, electronic health record, and registry as the applicable data sources and states that CMS form 2728, ESRD Medical Evidence Report Medicare Entitlement and/or patient registration, can be used as the data source, as these are required for CMS payment and are a routine part of dialysis admission and preemptive kidney transplant care.
- The developer states that the process for any inconsistent or questionable data submitted has been to contact the patient's renal care team for clarification. The developer asserts that this process has been shown to be feasible when carried out by a data analyst with proper training and program support.
- The developer states that its integrated care delivery model and electronic health records (EHR) system allows consistent data collection and reporting. On the other hand, the developer also explains that lack of consistent and sufficiently powerful electronic health records system as well as existing integrated multidisciplinary CKD-care networks that are capable of consistently collecting and reporting on this measure, may be a substantial factor as to why this measure is not yet publicly reported.
- The developer explains that the measure might be difficult to roll-out in smaller practices or on the level of individual dialysis clinics because reported outcomes in measuring period might be too small.

Questions for the Committee:

- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Are the data elements in form 2728 identical to what is needed to calculate the measure?*
- *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

- None. Need at least 60 patients in the denominator.
- Reasonable to assume feasible to implement outside of Kaiser Permanente.
- data has been collected and measure implemented without significant issues/problems since initial endorsement in 2015.
- All elements are generated and attainable. No concerns with feasibility.
- No concerns with feasibility for this measure.
- Data is generated or collected by and used by healthcare personnel during provision of care and is abstracted from a record other than person obtaining original information. Administrative claims, electronic clinical data, electronic health record and registry as the applicable data sources and states that cms form 2728 esrd medical evidence report Medicare entitlement and or patient registration. Inconsistent or questionable data submitted has been to contact patient's renal care team for clarification.
- No concerns.
- Feasibility is exceptional for this measure.
- High
- Yes
- It appears the data elements needed for measure are attainable thru Admin. Claims, EHR, and Registry data.
- No concern
- Will be possible to collect these data from the dialysis facility. However the significance of the ability of the dialysis facility to impact the data or the ability to correctly attribute that impact or change to the dialysis facility is problematic.
- Developer states its integrated delivery model and electronic health records allow for consistent data collection and reporting. However, the developer states that lack of consistent data and EHR for small practices made it difficult for them to roll out their data.
- Appropriate
- Moderate rating
- Data elements are routinely generated and are found in electronic sources.
- All the data elements can be generated electronically. The major concern is to ensure the end user is entering the data correctly and in a timely manner.
- I agree that the feasibility of collecting accurate data is on moderate strength. The smaller independent providers will struggle to calculate their data (probably a manual system).
- None

Criterion 4: Use and Usability

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

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 indicates that the Optimal ESRD Starts measure is currently utilized by the Permanente Federation to track the performance of eight Kaiser Permanente Regions (Northern California, Southern California, Northwest/Oregon, Southern Washington State, Hawaii, Colorado, Georgia, and Mid-Atlantic States.)
    - Performance results from the various Kaiser Permanente Regions are compared and shared with the regional nephrologists and renal teams, leading to operational comparisons, physician and health care team education, and appropriate resource allocation.
- The developer indicates that the Optimal ESRD Starts measure is benchmarked against the average US optimal ESRD starts.
- The developer explains that while they (Kaiser Permanente) or other entities do not publicly report on Optimal ESRD Starts performance, components of the measure are publicly reported: rate of pre-emptive transplantation, rate of functioning vascular access as well as incidence and prevalence of home dialysis modalities (PD and HHD).
- The developer plans to submit measure for consideration in federal programs.
- The developer states that the Optimal ESRD Starts measure is being adopted by other large professional organizations, such as Nephology Care Alliance.

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 references feedback provided in 2015 and 2020 from partner, American Society of Nephrology Quality Committee (ASN), regarding this measure.

- ASN states that "NQF 2594 optimal ESKD starts measure was the only metric specifically addressing advanced CKD and kidney replacement planning and is a superb area for quality measure development."
- ASN emphasizes that no aspect of the practice of nephrology is in greater need of quality improvement than how patients typically initiate dialysis, and further denotes that one of the foci of this measure, pre-emptive transplantation, is one of several underutilized and highly effective options in nephrology care.
- ASN also reiterates the importance of the patient population that the measure captures, the measure's high validity performance and the committee's sentiments towards the measure stating specifically that the measure is relevant and a step in the right direction.

- The developer references feedback provided in 2015 from Kidney Care Partners (KCP) regarding this measure.
  - The feedback indicates KCP's agreement with the appropriateness of the measure and support towards implementation of the measure but also identifies the measure as only feasible in fully integrated delivery care systems or large physician groups. According the KCP's feedback, implementation of the measure is not applicable to dialysis facilities or ESRD Seamless Care Organizations (ESCOs) because of potentially insufficient patient volumes.
  - KCP also indicates that the measure does not account for the fact that approximately 40% of patients have not yet been seen the by the nephrologist they will be assigned to at the time of dialysis initiation.
  - Feedback from KCP indicates consideration towards three exclusions addressing scenarios in which a permanent access may not be appropriate:
    - Patients with a limited life expectancy, where not placing a permanent access may be more consistent with the patient's goals;
    - Patients who are uncertain whether they wish to pursue long-term treatment and desire a time-limited trial of dialysis; and
    - Patients with Acute Kidney Injury without prior CKD who ultimately don't recover renal function; these patients will likely use a catheter for their first treatment, and it is appropriate to wait up to four months to see if they will recover function before pursuing permanent access.
  - Feedback from KCP indicates three potential unintended consequences for careful consideration:
    - Potential "misuse" or misrepresentation of an "optimal start"
    - Potential for premature incentivization towards insertion of both needles for the first treatment, which may lead to damage of immature AVFs.
      - As related to the measure's exclusion of patients with a single needle in the AVF that have blood return via their catheter.
    - Likelihood of penalty to providers who are caring for a disproportionate percentage of patients with low socioeconomic status (SES).
      - Individuals with a low SES profile often have not seen a nephrologist prior to beginning treatment; safety net providers may be inappropriately penalized.
- The developer references feedback from partner, Nephrology Care Alliance, which states that "High value care for chronic kidney disease a frontline solution."

- Learnings from national renal care stakeholders have been collected since 2017, summarized, shared with all markets to support clinical leaders in the identification of successes and barriers.
- Results, including comparative reporting, are also shared with mid-level managerial leadership and front-line staff, both providers and nursing.
    - Results are reviewed quarterly at clinical population management meetings and again at provider meetings.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *The developer has provided a rationale for why the measure is not in use in an accountability program yet. Does the Standing Committee agree that the rationale is acceptable?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

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 developer indicates that the measure has been utilized successfully to improve outcomes in Kaiser Permanente Southern California for nearly 10 years and in the national Kaiser Permanente program for 3 years.
- Performance data reveals that the count of new ESKD patients increased from 3,787 in 2017 to 4,002 in 2021.
    - In 2017, the Optimal ESRD Starts measure rate was 57.5 percent, 60.7 percent in 2019 (increase), and 56.5 percent in 2021.
    - The developer attributes the episodic downward trend in 2020 through 2021 to the Covid-19 pandemic impacts.

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

- The developer did not identify unintended negative consequences of measuring Optimal ESRD Starts. Developer identified several unexpected benefits gained from implementation of this measure:
    - Moving upstream with chronic kidney disease management to improve outcomes of Optimal ESRD Starts
    - Improved integration of care with primary providers, and improved electronic tools that support this effort

- Development of alternate options such as embedded and interventional radiology placed peritoneal dialysis catheter
- Improvement in quantifying the value of care provided to chronic kidney disease/end stage kidney disease patients
- Establishing a better relationship and processes with other services as part of a multi-disciplinary care approach to care for chronic kidney disease/end stage kidney disease patients
- Enhancing earlier patient activation and participation
- Implementing electronic tools to assist with screening and risk stratification for patients with chronic kidney disease

Potential harms

- None

***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

- Internal QI purposes
- Measure has been used within 8 Kaiser Permanente Regions for QI, accountability to Regional Medical Executive Directors within Kaiser Permanente.
- Not yet publicly reported but used in Permanente Federation in 8 regions with interest in use by other national care alliances.
- This is not clear to me, if feedback on the measure is readily provided, available or explained. However, it is a metric that has been widely discussed and touted for years.
- The developer provides prior feedback (2015 and 2020) from ASN, KCP, and the Nephrology Care Alliance that is generally in support of the measure. Of note, Kidney Care Partners (KCP) has pointed out that the measure is only feasible in fully integrated delivery care systems or large physician groups and is not applicable to dialysis facilities because of potentially insufficient patient volumes. KCP has also noted that the measure does not account for the fact that approximately 40% of patients have not yet been seen the by the nephrologist they will be assigned to at the time of dialysis initiation and cautioned of the potential unintended consequence of "misuse" or misrepresentation of an "optimal start," as well as the risk of disproportionately penalizing providers caring for a higher percentage of patients with low socioeconomic status (SES). Individuals with a low SES profile often have not seen a nephrologist prior to beginning treatment, and safety net providers may be inappropriately penalized.
- 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.
- N/A
- There is an acceptable plan to use this measure in an accountability program.

- Moderate
- Unknown
- How does the developer plan to address some of the comments and feedback received? i.e., exclusions, unintended consequences.
- Use has not yet extended beyond Kaiser but will soon.
- Yes
- Feedback from KCP - 1. Pts with limited life expectancy did not have permanent access 2. Pts that wanted a trial on dialysis 3. Pts with acute kidney injury KCP on unintended consequences: 1. Misuse on optimal start 2. premature use of using 2 needles 3. Likelihood of penalty to providers for low-income patients.
- True
- Unclear
- The developer noted that the Optimal ESRD starts measure is currently utilized by Kaiser to track and is benchmarked against average US optimal start. Plans to submit this for use in quality incentive programs.
- This measure is not publicly reported in its entirety. Would support this measure being publicly reported. The developers have sought after feedback. KCP's feedback was insightful, and the developer has done a good job summarizing those unintended consequences and committed to monitoring them.
- KP is measured and uses USRDS data for comparison. Other CKD organizations support this measure. KCP has viable concerns about groups using the new fistula too early or misrepresenting themselves by manipulating the data.
- There is no public reporting. It could be incorporated into DFC website

4b. Usability

- Continued improvement was not evident casting doubt on the value of the metric.
- Has not been used in a publicly reported program within 6 years of endorsement. Per submission, is intended for use in health plans, integrated delivery systems, large nephrology group providers, and broad geographic regions; not specified for individual practitioners, small provider groups with <50 new ESRD patients in measurement period, nor intended for use by dialysis providers.
- No unintended negative consequences seen with measure's use in KP system both in Southern California over last decade and more broadly in 8 regions since measure first endorsed.
- Elements of this measure, particularly transplant and home dialysis, are used in CMS kidney contracting.
- The developer did not identify any unintended consequences with the measure.
- 1. Management with chronic kidney disease management to improve outcomes of optimal esrd starts; improved integration with primary providers and improved electronic tools to support 3. alternate options as embedded and interventional radiology placed PD catheter; 4. improvement in quantifying the value of care provided to ckd esrd patients 5. better relationship with processes with other services as part of a multi disciplinary care approach 6. enhancing earlier patient activation and participation 7. implementing electronic tools to assist with screening and risk stratification for patients with ckd No potential harms
- May lead to patients not being accepted to dialysis units if they do not have an AVF; might lead to cherry picking of patients.
- There were no harms detected in this measure.
- None

- I think the majority of patients would benefit from home dialysis; however, there are a number that are unable to perform it at home and will these numbers skew this information?
- Risk adjustment could perhaps be considered so facilities impacted by geographic disparities would not be negatively impacted.
- This is a valuable tool to improve outcomes.
- One theoretical potential harm is patient's being "coerced" into choosing one modality over the other because of financial remuneration to the facility
- Improved care with primary providers and electronic tools that support this measure; development of options as interventional radiology for placed peritoneal dialysis patients; improved in valued care for patients with CKD and ESRD; establishing a better relationship with multi approach care with other providers. early patient activation and participation.
- Appropriate
- Measure is not useful as a facility-level measure.
- The developer has indicated that the measure has been utilized successfully to improve outcomes in Kaiser for nearly 10 years and can be utilized to continue to improve quality of healthcare provided. There are no unintended or negative consequences.
- There were no actual unintended consequences identified. A developer identified a couple of hypotheticals but not on the extent of causing harm.
- KP has had success with the model and has had no significant "harms" during the implementation of the measure.
- The performance results are not usable by dialysis providers.

Criterion 5: Related and Competing Measures

Related and competing measures

- The developer did not identify any related or competing measures.

Harmonization

- Not applicable. The developer did not identify any related or competing measures, therefore the developer does not provide an explanation for harmonization.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- Yes related measures. No competing measures.
- N/A
- While not necessarily competing, elements of this measure are presently components of kidney care contracting between nephrology groups and CMMI.
- Not applicable.
- No related and competing measures.
- N/A
- To my knowledge, there are no related or competing measures.
- No
- Unknown
- Harmonization is the goal.

- None
- No
- No
- True
- None
- There are no competing measure identified.
- There are no competing measures.
- There are no competing measures.
- None

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

Member Expression of Support

- No members submitted an expression of support for this measure.

Comments

- No NQF member and public comments were received in advance of the Standing Committee evaluation.

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Have measure specifications changed since the last review? ☒ Yes ☐ No**
2. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**
3. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
    - Measure specifications are clear and precise. The developer indicates slight changes to reporting specifications since 2015 but attests that those changes are non-consequential. Those changes include:
        - Removal of the criterion of 10% new hemodialysis patient limit for arteriovenous graft (AVG). Developer made this change based on recommendations from the Kidney Disease Outcomes Quality Initiative (KDOQI) 2019 Vascular Access guidelines.
    - The developer explains that removal of this criterion did not have significant impact on the reported AVG rate over time.
    - Inclusion of patients with failing kidney transplants starting or returning to dialysis. Developer made this change based on various publications that examined the high number of failed transplant patients who returned to dialysis post-transplant.
    - The developer explains that this group of patients constituted about three percent of the measure denominator population and therefore did not have significant impact on the reported measurement rate.
    - Developer asserts that the data elements and collection process have not changed since the last NQF measure submission.

RELIABILITY: TESTING

4. **Did the developer conduct new reliability testing? ☐ Yes ☒ No**

   4a. If no, summarize the Standing Committee's previous feedback:

   - The developer conducted validity testing at the patient or encounter level; additional reliability testing is not required, but two changes were made to this measure to account for clinical guideline changes as well as to respect patient choice.

   4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:

   - Two changes were made to this measure to account for clinical guideline changes as well as to respect patient choice:
     - Removal of the criteria of 10% new hemodialysis patient limit for arteriovenous graft (AVG)
     - Inclusion of patients with failing kidney transplants starting or returning to dialysis

5. **Reliability testing level: ☐ Accountable-Entity Level ☒ Patient/Encounter Level ☐ Neither**

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

   ☒ Yes ☐ No

7. 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

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

   - The developer conducted validity testing at the patient or encounter level; additional reliability testing is not required.

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

   - The developer conducted validity testing at the patient or encounter level; additional reliability testing is not required.

10. 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.

    ☒ Yes ☐ No ☐ Not applicable

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

    ☒ Yes ☐ No ☐ Not applicable (patient/encounter level testing was not performed)

12. **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)

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

    - Measure specifications are clear and precise.
    - Developer demonstrates reliability testing using validity testing at the patient or encounter level.

VALIDITY: TESTING

14. **Did the developer conduct new validity testing? ☐ Yes ☒ No**

14a. If no, summarize the Standing Committee's previous feedback:

- The measure data elements collected for this measure are reviewed and validated at the source within each market even before it was submitted to the national reporting team for aggregated reporting. This data collection process has not changed since the last NQF measure submission in 2015. Therefore, the testing result submitted in 2015 is still relevant today.
    - Developer attests that the data elements collected for this measure in 2015 have remained unchanged for the purpose of the measurement and are still applicable.
        - Developer confirmed from a reporting perspective that since 2015, they have made changes to the reporting specifications. Those changes include:
            1. Removal of the criteria of 10% new hemodialysis patient limit for arteriovenous graft (AVG)
                - The percent AVG in 2016 (before removal of 10% limit) and 2021 (after removal of 10% limit) remained below the 10% AVG limit previously set in place. The developer advised that this suggests that % AVG have not been impacted that much by removal of this limiting criterion.
            2. Inclusion of patients with failing kidney transplants starting or returning to dialysis. These changes did not change the way the data elements were collected.
    - The developer assessed the impact of including these patients and found that there was a small difference of 3.3% in the denominator with the criterion versus the denominator without the criterion, with a range of differences from 0% to 5% in the eight markets.
- Developer explained that validity testing demonstrated that these changes had negligible impact on the individual measurement component percentage of arteriovenous graft rate level.

14b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:

- N/A

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

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

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

16. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

☒ **Yes**

☐ **No**

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

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

☐ **Face validity**

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

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

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

☒ **Yes**

☐ **No**

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

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

The developer tested the accuracy of the regional data compared to the authoritative source comparing the renal replacement therapy information submitted by the regional care coordinator to that provided by the renal replacement therapy provider on record.

- Validity testing conducted at the Patient/Encounter Level:
  - The developer utilized data element testing from their 2015 data element testing.
  - The developer tested the accuracy of the regional data compared to the authoritative source.
    - The renal replacement therapy information submitted by the regional care coordinator was compared to that provided by the renal replacement therapy provider on record. Among the data element elements, the developer tests it's a priori hypothesis of a 0.9 match (90% accuracy) between the region and source data.
- Systematic assessment of face validity of performance measure score as an indicator of quality or resource use was also conducted.
  - Developer explains that this measurement is used regularly within Kaiser Permanente and its markets as it provides insights to Kaiser Permanente leadership about the care provided for our chronic kidney disease members.

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

- The developer calculated the denominator data elements to be 96% (87%-99% 95CI) accurate.
- The developer calculated the numerator data elements to be 87% (76%-92% 95CI) accurate.
- Total element match rate was 83% (71%-91% 95CI). Based on this data, developer affirmed a positive predictive value of 0.94 to identify true optimal ESRD start, negative predictive value of 0.79 to identify a non-optimal start.
- The developer reported that the region correctly identifies a true optimal ESRD start (positive likelihood ratio) 11.6 times more often than it incorrectly identifies a non-optimal ESRD start as optimal and that this is more than double the generally accepted rule that a test should have a ratio greater than 5. The developer also reported the ratio of reporting false negatives (negative likelihood ratio)(LR-), of 0.19 which meets the generally accepted threshold that a LR- should be less than 0.2.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

- No concerns

22. **Risk Adjustment**

22a. **Risk-adjustment method**

☒ None (only answer Question 20b and 20e) ☐ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

22b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

22c. **Social risk adjustment:**

22c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

22c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

22c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

22d.**Risk adjustment summary:**

22d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

22d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

22d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

22d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

22d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

22e**. Assess the risk-adjustment approach**

- None

23. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**
    - No concerns. Developer attests to and provides data to support its assertion that the measure is capable of identifying meaning differences in performance.
24. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    - Not applicable
25. **Please describe any concerns you have regarding missing data.**
    - No concerns. Developer demonstrates that the extent of missing data does not impact the validity of the measure.

ADDITIONAL RECOMMENDATIONS

26. **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.**
    - None

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

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]

Yes

[Yes Please Explain]

Summary of new or updated evidence:

- KDIGO Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation recommends that all patients with chronic kidney disease (CKD) G4-G5 (glomerular filtration rate [GFR] < 30ml/min/1.73 m2) who are expected to reach end-stage kidney disease (ESKD) (excluding those listed in Rec 1.1.3) be informed of, educated about, and considered for kidney transplantation
- The European Society for Vascular Surgery recommends an autogenous arteriovenous fistula as the primary option for vascular access.
- The Vascular Access for Haemodialysis recommends that all patients with end stage kidney disease who commence haemodialysis or are on long-term haemodialysis should dialyse with an arteriovenous fistula as first choice, an arteriovenous graft as second choice, a tunnelled venous catheter as third choice and a nontunnelled temporary catheter as an option of necessity.
- The Canadian Society of Nephrology recommends the following: each center should establish a dedicated team for vascular access; preserve arm veins suitable for placement of vascular access; preservation should begin in patients with progressive kidney disease and an estimated GFR of less than 30 ml/min; radio-cephalic native vessel arteriovenous fistula is the preferred type of vascular access; establish AV fistulae when the patient has an estimated GFR of 15 to 20 ml/min and progressive kidney disease.
- A systematic review looking at a total of twenty-one studies with 29,000 participants by authors Chuasuwan et al. in "Health and Quality of Life Outcomes (2020) 18:191" supported the placement of Peritoneal Dialysis in ESKD patients.

[Response Ends]

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]

The process model below illustrates how the health care system is being measured and how the individual measure components are closely related to the desired health outcomes which include lowering patient mortality and cardiovascular risk, improving patient quality of life, and reducing hospitalization and cost of care.

Diagram 1. A process diagram that describes how the health care system is being measured and how each measure component within the measure is connected to patient health outcomes.

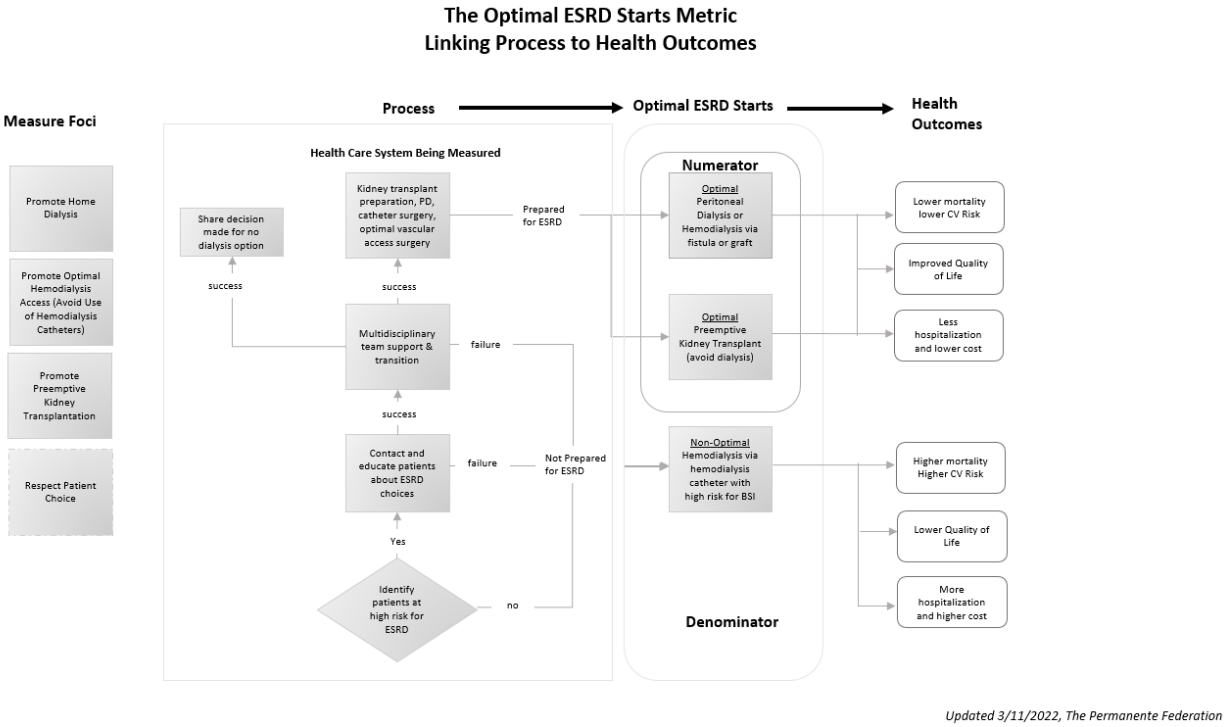


Diagram. A process diagram that describes how the health care system is being measured and how each measure component within the measure is connected to patient health outcomes.

[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]

Clinical Practice Guideline recommendation (with evidence review)

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

[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]

Chadban SJ, Ahn C, Axelrod DA, Foster BJ, Kasiske BL, Kher V, Kumar D, Oberbauer R, Pascual J, Pilmore HL, Rodrigue JR, Segev DL, Sheerin NS, Tinckam KJ, Wong G, Knoll GA. KDIGO Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation. Transplantation. 2020 Apr;104(4S1 Suppl 1):S11-S103. doi: 10.1097/TP.0000000000003136. PMID: 32301874.

https://kdigo.org/wp-content/uploads/2018/08/KDIGO-Txp-Candidate-GL-FINAL.pdf

[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]

Recommendation 1.1: We recommend that all patients with chronic kidney disease (CKD) G4-G5 (glomerular filtration rate [GFR] < 30ml/min/1.73 m2) who are expected to reach end-stage kidney disease (ESKD) (excluding those listed in Rec 1.1.3) be informed of, educated about, and considered for kidney transplantation regardless of socioeconomic status, sex, gender identity, or race/ethnicity (1D).

[Response Ends]

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

[Response Begins]

Grade D

Grade D: quality of evidence "very low", meaning the estimate of effect is very uncertain, and often will be far from the truth.

[Response Ends]

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

[Response Begins]

Grade	Quality of Evidence	Meaning
A	High	We are confident that the true effect lies close to that of the estimate of the effect.

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

Table. This table contains the Evidence Grading System and lists the grade, the associated quality of evidence and the associated meaning of the grade.

[Response Ends]

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

[Response Begins]

Grade 1

Grade Level 1: "We recommend".

Implications for patients: Most people in your situation would want the recommended course of action and only a small portion would not.

Implications for clinicians: Most patients should receive the recommended course of action.

Implications for policy: The recommendation can be evaluated as a candidate for developing a policy or a performance measure.

Grade Level 2: "We suggest"

Implications for patients: The majority of people in your situation would want the recommended course of action, but many would not.

Implications for clinicians: Different choices will be appropriate for different patients. Each patient needs help to arrive at a management decision consistent with her or his values and preferences.

Implications for policy: The recommendation is likely to require substantial debate and involvement of stakeholders before policy can be determined.

Not Graded: Used, typically, to provide guidance based on common sense or where the topic does not allow adequate application of evidence. The most common examples include recommendations regarding monitoring intervals, counseling, and referral to other clinical specialists. Ungraded recommendations are generally written as declarative statements, but not meant to be interpreted as being strong recommendations than Level 1 or Level 2 recommendations.

[Response Ends]

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

[Response Begins]

Grade Level 1: "We recommend".

Implications for patients: Most people in your situation would want the recommended course of action and only a small portion would not.

Implications for clinicians: Most patients should receive the recommended course of action.

Implications for policy: The recommendation can be evaluated as a candidate for developing a policy or a performance measure.

Grade Level 2: “We suggest”

Implications for patients: The majority of people in your situation would want the recommended course of action, but many would not.

Implications for clinicians: Different choices will be appropriate for different patients. Each patient needs help to arrive at a management decision consistent with her or his values and preferences.

Implications for policy: The recommendation is likely to require substantial debate and involvement of stakeholders before policy can be determined.

Not Graded: Used, typically, to provide guidance based on common sense or where the topic does not allow adequate application of evidence. The most common examples include recommendations regarding monitoring intervals, counseling, and referral to other clinical specialists. Ungraded recommendations are generally written as declarative statements, but not meant to be interpreted as being strong recommendations than Level 1 or Level 2 recommendations.

[Response Ends]

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

[Response Begins]

https://kdigo.org/wp-content/uploads/2018/08/KDIGO-Txp-Candidate-GL-FINAL.pdf

page S28

Table. A table that shows the topics searched, citations screened, number of studies included, and summary tables or evidence profiles pertaining to Access to Transplantation.

Topics	Topics Searched	Citations Screened	Included Studies, n	Summary Tables / Evidence Profiles
1. Access to Transplantation	Txp vs. Wtl.	1832	8	+

Table 3. Work products for the guideline

https://kdigo.org/wp-content/uploads/2018/08/KDIGO-Txp-Candidate-STs-EPs-FINAL.pdf

page 6

Table. A table that shows the evidence profile for Kidney Transplantation Vs Waitlisting including predictor, outcome, number of studies, number of patients, quality of studies, consistency across studies, directness of the evidence, other considerations and summary of findings.

Predictor	Outcome	# of Studies	Total N of Patients	Methodological Quality of Studies	Consistency Across Studies	Directness of the Evidence	Other Considerations	Summary of Findings - Quality of Evidence	Summary of Findings - Description of Findings and Outcome Importance
Age	Death	6	14149 (174-5961)	No limitations (0)	No important inconsistencies (0)	Direct (0)	None (0)	High	Txp superior to waitlist in almost all age groups*^; Critical
Transplantation in HIV+	Death	1	317	Serious limitations (-1)	N/A	Direct (0)	Sparse (-2)	Very Low	Txp was comparable to waitlist in patients who were HIV+; Critical
Obesity	Death	1	7443	No limitations (0)	N/A	Direct (0)	Sparse (-2)	Low	Txp similarly superior to waitlist among obese and nonobese; Critical

Predictor	Outcome	# of Studies	Total N of Patients	Methodological Quality of Studies	Consistency Across Studies	Directness of the Evidence	Other Considerations	Summary of Findings - Quality of Evidence	Summary of Findings - Description of Findings and Outcome Importance
Obesity	Graft Loss	1	7443	No limitations (0)	N/A	Direct (0)	Sparse (-2)	Low	Txp similarly superior to waitlist among obese and nonobese; Critical
Overall Summary:	*	*	*	*	*	*	*	Quality of Overall Evidence:	*
Transplant generally found to be superior to continued waitlist status regardless of age or obesity	*	*	*	*	*	*	*	Variable	*

Evidence Profile: Kidney Transplantation vs. Waitlisting

GL = Guideline, N/A = Not Applicable, NS = Nonsignificant predictor, Txp = transplantation.

*1 study found similar RR (0.12-0.34) of transplant vs. waitlist across age groups 18-34 years through >65 years (lowest age cohort was non-significant, likely due to lack of statistical power). 4 studies restricted to elderly (>60-70 years) all found significantly lower death with transplant (RR/HR=0.36-0.67), including in a subgroup restricted to >= 75 years old. 1 study of children found large differences in death, favoring transplant over waitlist across 3 age strata (0-5, 6-12 13-18 years; HR=0.09-0.52); however, in the small subset of 0/5 year olds, the RR was not statistically significant.

*cells intentionally left blank

[Response Ends]

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

[Response Begins]

Overall summary was that transplant generally were found to be superior to continued waitlist status regardless of age or obesity.

https://kdigo.org/wp-content/uploads/2018/08/KDIGO-Txp-Candidate-STs-EPs-FINAL.pdf

page 6

Table. A table that shows the evidence profile for Kidney Transplantation Vs Waitlisting including predictor, outcome, number of studies, number of patients, quality of studies, consistency across studies, directness of the evidence, other considerations and summary of findings.

Predictor	Outcome	# of Studies	Total N of Patients	Methodological Quality of Studies	Consistency Across Studies	Directness of the Evidence	Other Considerations	Summary of Findings - Quality of Evidence	Summary of Findings - Description of Findings and Outcome Importance
Age	Death	6	14149 (174-5961)	No limitations (0)	No important inconsistencies (0)	Direct (0)	None (0)	High	Txp superior to waitlist in almost all age groups*^; Critical
Transplantation in HIV+	Death	1	317	Serious limitations (-1)	N/A	Direct (0)	Sparse (-2)	Very Low	Txp was comparable to waitlist in patients who were HIV+; Critical

Predictor	Outcome	# of Studies	Total N of Patients	Methodological Quality of Studies	Consistency Across Studies	Directness of the Evidence	Other Considerations	Summary of Findings - Quality of Evidence	Summary of Findings - Description of Findings and Outcome Importance
Obesity	Death	1	7443	No limitations (0)	N/A	Direct (0)	Sparse (-2)	Low	Txp similarly superior to waitlist among obese and nonobese; Critical
Obesity	Graft Loss	1	7443	No limitations (0)	N/A	Direct (0)	Sparse (-2)	Low	Txp similarly superior to waitlist among obese and nonobese; Critical
Overall Summary:	*	*	*	*	*	*	*	Quality of Overall Evidence:	*
Transplant generally found to be superior to continued waitlist status regardless of age or obesity	*	*	*	*	*	*	*	Variable	*

Evidence Profile: Kidney Transplantation vs. Waitlisting

GL = Guideline, N/A = Not Applicable, NS = Nonsignificant predictor, Txp = transplantation.

*1 study found similar RR (0.12-0.34) of transplant vs. waitlist across age groups 18-34 years through >65 years (lowest age cohort was non-significant, likely due to lack of statistical power). 4 studies restricted to elderly (>60-70 years) all found significantly lower death with transplant (RR/HR=0.36-0.67), including in a subgroup restricted to >= 75 years old. 1 study of children found large differences in death, favoring transplant over waitlist across 3 age strata (0-5, 6-12 13-18 years; HR=0.09-0.52); however, in the small subset of 0/5 year olds, the RR was not statistically significant.
*cells intentionally left blank

[Response Ends]

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

[Response Begins]

NA

[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]

Schold JD, Buccini LD, Goldfarb DA, Flechner SM, Poggio ED, Sehgal AR. Association between Kidney Transplant Center Performance and the Survival Benefit of Transplantation Versus Dialysis. Clin J Am Soc Nephrol. 2014 Oct 7;9(10):1773-80. doi: 10.2215/CJN.02380314. Epub 2014 Sep 18. PubMed PMID:

25237071; PubMed Central PMCID: PMC4186511.

Description: A retrospective cohort study of adults wait-listed for kidney transplantation in the United States from 2003 to 2010 using the Scientific Registry of Transplant Recipients was conducted. The primary aim was to investigate whether measured center performance modifies the survival benefit of transplantation versus dialysis.

Summary: Among 223,808 waitlisted patients, 59,199 and 32,764 patients received a deceased or living donor transplant, respectively. Median follow-up from listing was 43 months (25th percentile=25 months, 75th percentile=67 months), and there were 43,951 total patient deaths. Deceased donor transplantation was independently associated with lower mortality at each center performance level compared with remaining on the waiting list; adjusted hazard ratio was 0.24 (95% confidence interval, 0.21 to 0.27) among 11,972 patients listed at high-performing centers, adjusted hazard ratio was 0.32 (95% confidence interval, 0.31 to 0.33) among 203,797 patients listed at centers performing as expected, and adjusted hazard ratio was 0.40 (95% confidence interval, 0.35 to 0.45) among 8039 patients listed at low-performing centers.

Findings indicate that measured center performance modifies the survival benefit of kidney transplantation, but the benefit of transplantation remains highly significant even at centers with low measured quality. Policies that concurrently emphasize improved center performance with access to transplantation should be prioritized to improve ESRD population outcomes.

Impact on conclusion: Findings are consistent with earlier meta-analyses and reviews.

Bouaoun L, Villar E, Ecochard R, Couchoud C. Excess risk of death increases with time from first dialysis for patients on the waiting list: implications for renal allograft allocation policy. Nephron Clin Pract. 2013;124(1-2):99-105. doi: 10.1159/000355549. Epub 2013 Oct 26. PubMed PMID: 24192719.

Description: The study used data from the French Renal Epidemiology and Information Network Registry to quantify the risk of death among patients on the waiting list.

Summary: During 45,013 person-years of follow-up, 7,224 patients died, 5,956 (82%) more than expected relative to the general population. The excess risk of death increased by 45% per additional year on the waiting list (18%-79%), p =

0.0005). The excess death rate of wait-listed patients was 1.7 times (1.1-2.7) higher than that of patients with kidney transplantation during the study period.

Impact on conclusion: Findings are consistent with earlier meta-analyses and reviews.

Bisigniano L, López-Rivera A, Tagliafichi V, Fernández VJ, Soratti CA. Analysis of mortality while on waiting list for kidney transplant in adults in Argentina 2005-2009. Transplant Proc. 2012 Sep;44(7):2239-41. doi: 10.1016/j.transproceed.2012.07.128. PubMed PMID: 22974963.

Description: The study compares the survival of deceased donor transplant patients to patients on the waiting list in Argentina.

Summary:

We analyzed 1682 patients transplanted average age 48.14 + 13.48 years and 3647 patients on waiting lists average age 47.88 + 14.32 years. For patients transplanted 30-day survival was 99.8% at 1 year 96.2% and 5 years of 79.9%. For patients on the waiting list survival at 30 days was 99.7% at 1 year and 5 years 94.6% 66.6%. Chi-square was 42.77, P =<.0001. HR 0.64 (95% CI 0.56 to 0.73). Cox regression for patients on waiting lists HR 1.40 (95% CI 1.20 –1.63) P<=.0001. The time dependent Cox regression showed for patients transplanted at 30 days, <1 year >1 year showed HR 4.18 (95% CI 2.88–6.06) P<=.0001, HR 0.40 (95% CI 0.27 to 0.61) P<=.0001 and HR 0.19 (95% CI 1.12– 0.29) P <=.0001, respectively.

Survival, both at baseline and in the long term, is better in transplant patients as compared to patients on waiting list. In Cox time–dependant regression the risk of death during the first 30 days is 4 times higher in transplant patients. This reverses and at 1 year, transplant patients are 60% less likely to die, and after one year this probability is 81% lower (P =<.0001).

Impact on conclusion: Findings are consistent with earlier meta-analyses and reviews.

Kontodimopoulos N, Niakas D. An estimate of lifelong costs and QALYs in renal replacement therapy based on patients' life expectancy. Health Policy. 2008 Apr;86(1):85-96. Epub 2007 Nov 9. PubMed PMID: 17996975.

Description: To estimate lifelong costs and quality adjusted life years (QALYs) of hemodialysis (HD), peritoneal dialysis (PD) and renal transplantation (Tx) in Greece, based on individual patient life expectancy.

Summary: Estimated lifelong QALYs were 4.37 (HD), 3.94 (PD) and 16.11 (Tx) (P < 0.001). Annual HD and PD costs per patient were estimated at € 36,247 and € 30,719 respectively. For Tx, average 1st year, 3-year and lifelong (undiscounted) costs were € 31,714, € 43,275 and € 151,274 respectively.

Impact on conclusion: While the data does not translate directly to the U.S., this provides an estimate on the comparative cost benefit of transplant versus dialysis.

New evidence identified for this update. All preceding studies in this section were included in the last submission:

Ryu JH, Koo TY, Ro H, Cho JH, Kim MG, Huh KH, Park JB, Lee S, Han S, Kim J, Oh KH, Yang J; KNOW-KT Study group. Better health-related quality of life in kidney transplant patients compared to chronic kidney disease patients with similar renal function. PLoS One. 2021 Oct 4;16(10):e0257981. doi: 10.1371/journal.pone.0257981. PMID: 34606505; PMCID: PMC8489710.

Description: To investigate the longitudinal changes in health-related quality of life (HRQOL) with respect to both general health status and CKD-specific health status in kidney transplant (KT) patients, and to compare these HRQOL patterns in KT and CKD patients at the same CKD stages using data from patients enrolled in the KoreaN cohort study for Outcome in patients With Kidney Transplantation (KNOW-KT) and the KoreaN cohort study for Outcome in patients with Chronic Kidney Disease (KNOW-CKD) studies.

Summary: KT patients had higher SF-36 scores with respect to both physical and mental QOL than CKD patients at the same CKD stage; however they had similar CKD-targeted scores compared to CKD patients with similar renal function. KT was a significant prognostic factor associated with better QOL, independent of renal function.

Impact on conclusion: The data does not translate directly to the U.S., but this is consistent with other studies and reviews showing that KT yields improvements to QOL.

de Brito DCS, Machado EL, Reis IA, Moreira DP, Nébias THM, Cherchiglia ML. Modality transition on renal replacement therapy and quality of life of patients: a 10-year follow-up cohort study. Qual Life Res. 2019 Jun;28(6):1485-1495. doi: 10.1007/s11136-019-02113-z. Epub 2019 Jan 21. PMID: 30666548.

Description: Assesses changes in a cohort of end-stage renal disease (ESRD) dialysis patients over 10 years (for survivors).

Summary: Transplantation was shown to be treatment with more benefits to patients with ESRD. Transplant patients sustained greater social participation, job retention, and improvement in SF-36 scores compared to those who did not have KT as part of RRT. More time on dialysis was associated with a reduction on QOL.

Impact on conclusion: This is consistent with other studies and reviews showing that KT yields improvements to QOL.

[Response Ends]

Group 2 - Evidence - Systematic Reviews Table

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

[Response Begins]

Vascular Access for Haemodialysis

UK Renal Association

6th Edition, January 2015

https://ukkidney.org/sites/renal.org/files/vascular-access.pdf

[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]

1. Preferred type of vascular access (Guideline 1.1)

Guideline 1.1 – Incident and prevalent patients

We recommend that all patients with end stage kidney disease who commence haemodialysis or are on long-term haemodialysis should dialyse with an arteriovenous fistula as first choice, an arteriovenous graft as second choice, a tunnelled venous catheter as third choice and a nontunnelled temporary catheter as an option of necessity (1A).

7. Prevention of catheter related infections (Guidelines 7.1-7.4)

Guideline 7.1 – Minimise the use of venous catheters

We recommend that central venous catheters should be employed as a method of last resort for longer term vascular access to reduce the overall risk of infectious complications and the burden of central venous stenosis in haemodialysis patients (1B).

[Response Ends]

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

[Response Begins]

Grade 1

The modified GRADE system has been adopted by the Renal Association Clinical Practice Guidelines Committee and has been used to grade the recommendations in all of the modules in the 5th edition of the Renal Association guidelines. It explicitly describes both the strength of the recommendations and the quality of the underlying evidence, with the aim of maximising applicability to standard clinical practice (1-4). The modified GRADE system grades level of expert recommendation as “strong” (Grade 1) or “weak” (Grade 2) according to balance of benefits, risk, burden and cost. The quality or level of evidence is assessed as “high” (Grade A), “moderate” (Grade B), “low” (Grade C) or “very low” (D) depending on factors such as study design, directness of evidence and consistency of results (1-4).

1. Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328:1490

2. Jaeschke R, Guyatt GH, Dellinger P et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008; 337:327-330

3. Uhlig K, MacLeod A, Craig J et al. Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int 2006; 70:2058-2065.

4. Kidney Disease: Improving Global Outcomes. KDIGO clinical practice guidelines for the prevention, diagnosis, evaluation and treatment of Hepatitis C in chronic kidney disease. Kidney Int 2008; 73 (S109): S1-S99.

[Response Ends]

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

[Response Begins]

The modified GRADE system has been adopted by the Renal Association Clinical Practice Guidelines Committee and has been used to grade the recommendations in all of the modules in the 5th edition of the Renal Association guidelines. It explicitly describes both the strength of the recommendations and the quality of the underlying evidence, with the aim of maximising applicability to standard clinical practice (1-4). The modified GRADE system grades level of expert recommendation as “strong” (Grade 1) or “weak” (Grade 2) according to balance of benefits, risk, burden and cost. The quality or level of evidence is assessed as “high” (Grade A), “moderate” (Grade B), “low” (Grade C) or “very low” (D) depending on factors such as study design, directness of evidence and consistency of results (1-4).

1. Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328:1490

2. Jaeschke R, Guyatt GH, Dellinger P et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008; 337:327-330

3. Uhlig K, MacLeod A, Craig J et al. Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int 2006; 70:2058-2065.

4. Kidney Disease: Improving Global Outcomes. KDIGO clinical practice guidelines for the prevention, diagnosis, evaluation and treatment of Hepatitis C in chronic kidney disease. Kidney Int 2008; 73 (S109): S1-S99.

[Response Ends]

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

[Response Begins]

Grades A, B

The modified GRADE system has been adopted by the Renal Association Clinical Practice Guidelines Committee and has been used to grade the recommendations in all of the modules in the 5th edition of the Renal Association guidelines. It explicitly describes both the strength of the recommendations and the quality of the underlying evidence, with the aim of

maximising applicability to standard clinical practice (1-4). The modified GRADE system grades level of expert recommendation as “strong” (Grade 1) or “weak” (Grade 2) according to balance of benefits, risk, burden and cost. The quality or level of evidence is assessed as “high” (Grade A), “moderate” (Grade B), “low” (Grade C) or “very low” (D) depending on factors such as study design, directness of evidence and consistency of results (1-4).

1. Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328:1490

2. Jaeschke R, Guyatt GH, Dellinger P et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008; 337:327-330

3. Uhlig K, MacLeod A, Craig J et al. Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int 2006; 70:2058-2065.

4. Kidney Disease: Improving Global Outcomes. KDIGO clinical practice guidelines for the prevention, diagnosis, evaluation and treatment of Hepatitis C in chronic kidney disease. Kidney Int 2008; 73 (S109): S1-S99.

[Response Ends]

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

[Response Begins]

The modified GRADE system has been adopted by the Renal Association Clinical Practice Guidelines Committee and has been used to grade the recommendations in all of the modules in the 5th edition of the Renal Association guidelines. It explicitly describes both the strength of the recommendations and the quality of the underlying evidence, with the aim of maximising applicability to standard clinical practice (1-4). The modified GRADE system grades level of expert recommendation as “strong” (Grade 1) or “weak” (Grade 2) according to balance of benefits, risk, burden and cost. The quality or level of evidence is assessed as “high” (Grade A), “moderate” (Grade B), “low” (Grade C) or “very low” (D) depending on factors such as study design, directness of evidence and consistency of results (1-4).

1. Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328:1490

2. Jaeschke R, Guyatt GH, Dellinger P et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008; 337:327-330

3. Uhlig K, MacLeod A, Craig J et al. Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int 2006; 70:2058-2065.

4. Kidney Disease: Improving Global Outcomes. KDIGO clinical practice guidelines for the prevention, diagnosis, evaluation and treatment of Hepatitis C in chronic kidney disease. Kidney Int 2008; 73 (S109): S1-S99.

[Response Ends]

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

[Response Begins]

The review included 62 cohort studies comprising of 586,337 participants.

[Response Ends]

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

[Response Begins]

Ravani et al. included 62 cohort studies comprising of 586,337 patients.

Catheter versus Fistula: Persons using catheters had increased risk of all-cause mortality (risk ratio [RR]=1.53, 95% confidence interval [95% CI]=1.41–1.67), fatal (RR=2.12, 95% CI=1.79–2.52) and nonfatal (RR=4.66, 95% CI=2.63–8.26) infection, major cardiovascular event (RR=1.38, 95% CI=1.24–1.54), and hospitalization (RR=1.68, 95% CI=1.33–2.12) compared with persons using fistulas.

Catheter versus Graft: Persons using catheters had increased risk of all-cause mortality (RR=1.38, 95% CI=1.25–1.52), fatal (RR=1.49, 95% CI=1.15– 1.93) and nonfatal (RR=2.78, 95% CI=1.80–4.29) infection, cardiovascular event (RR=1.26, 95% CI=1.11–1.43), and hospitalization (RR=1.51, 95% CI=1.30–1.75) compared with those individuals using grafts.

In absolute terms, catheter use is associated with 80–134 additional deaths per 1000 person-years compared with fistula use and 60–125 additional deaths per 1000 person-years compared with graft use. Graft use is associated with 18–54 additional deaths for every 1000 persons each year compared with fistula use.

The direction of the risk for mortality (lower in fistula and graft users than catheter users) was highly consistent across existing studies. Despite marked study heterogeneity, the direction of association with mortality was consistent in 49 of 51 (96%) comparisons.

[Response Ends]

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

[Response Begins]

While individual studies might address adverse events or harms, the meta-analysis does not provide summary information.

[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]

An updated literature search did not reveal any large studies comparing outcomes for catheter versus other access types.

[Response Ends]

Group 3 - Evidence - Systematic Reviews Table

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

[Response Begins]

Schmidli J, Widmer MK, Basile C, et al. Editor's Choice – Vascular Access: 2018 Clinical Practice Guidelines of the European Society for Vascular Surgery (ESVS). European Journal of Vascular and Endovascular Surgery. 2018;55(6):757-818.

https://www.ejves.com/article/S1078-5884(18)30080-7/fulltext#secsectitle0090

[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]

Recommendation 3: An autogenous arteriovenous fistula is recommended as the primary option for vascular access.

[Response Ends]

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

[Response Begins]

Class I, level A

Class I: Evidence and/or general agreement that a given treatment or procedure is beneficial, useful, effective.

Level A: Data derived from multiple randomized clinical trials or meta-analyses.

[Response Ends]

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

[Response Begins]

Class I: Evidence and/or general agreement that a given treatment or procedure is beneficial, useful, effective.

Class II: Conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of the given treatment or procedure.

Class IIa: Weight of evidence/opinion is in favor of usefulness/efficacy.

Class IIb: Usefulness/efficacy is less well established by evidence/opinion.

Class III: Evidence or general agreement that the given treatment or procedure is not useful/effective, and in some cases may be harmful.

Level of evidence A: Data derived from multiple randomized clinical trials or meta-analyses.

Level of evidence B: Data derives from a single randomized clinical trial or large non-randomized studies.

Level of evidence C: Consensus of opinion of the experts and/or small studies, retrospective studies, registries.

[Response Ends]

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

[Response Begins]

Class I: Evidence and/or general agreement that a given treatment or procedure is beneficial, useful, effective.

Class II: Conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of the given treatment or procedure.

Class IIa: Weight of evidence/opinion is in favor of usefulness/efficacy.

Class IIb: Usefulness/efficacy is less well established by evidence/opinion.

Class III: Evidence or general agreement that the given treatment or procedure is not useful/effective, and in some cases may be harmful.

Level of evidence A: Data derived from multiple randomized clinical trials or meta-analyses.

Level of evidence B: Data derives from a single randomized clinical trial or large non-randomized studies.

Level of evidence C: Consensus of opinion of the experts and/or small studies, retrospective studies, registries.

[Response Ends]

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

[Response Begins]

Class I: Evidence and/or general agreement that a given treatment or procedure is beneficial, useful, effective.

Class II: Conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of the given treatment or procedure.

Class IIa: Weight of evidence/opinion is in favor of usefulness/efficacy.

Class IIb: Usefulness/efficacy is less well established by evidence/opinion.

Class III: Evidence or general agreement that the given treatment or procedure is not useful/effective, and in some cases may be harmful.

Level of evidence A: Data derived from multiple randomized clinical trials or meta-analyses.

Level of evidence B: Data derives from a single randomized clinical trial or large non-randomized studies.

Level of evidence C: Consensus of opinion of the experts and/or small studies, retrospective studies, registries.

[Response Ends]

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

[Response Begins]

1 systematic review and meta-analysis (k=83 studies included and pooled):

Murad MH, Elamin MB, Sidawy AN, Malaga G, Rizvi AZ, Flynn DN, Casey ET, McCausland FR, McGrath MM, Vo DH, El-Zoghby Z, Duncan AA, Tracz MJ, Erwin PJ, Montori VM. Autogenous versus prosthetic vascular access for hemodialysis: a systematic review and meta-analysis. J Vasc Surg. 2008 Nov;48(5 Suppl):34S-47S. doi: 10.1016/j.jvs.2008.08.044. PMID: 19000592.

Ng LJ, Chen F, Pisoni RL, Krishnan M, Mapes D, Keen M, Bradbury BD. Hospitalization risks related to vascular access type among incident US hemodialysis patients. Nephrol Dial Transplant. 2011 Nov;26(11):3659-66. doi: 10.1093/ndt/gfr063. Epub 2011 Mar 3. PMID: 21372255.

[Response Ends]

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

[Response Begins]

NA

[Response Ends]

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

[Response Begins]

NA

[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]

NA

[Response Ends]

Group 4 - Evidence - Systematic Reviews Table

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

[Response Begins]

1) J Am Soc Nephrol 17: S1–S27, 2006

2) Evidence for these guidelines was based on this systematic review: Mustafa RA, Zimmerman D, Rioux JP, Suri RS, Gangji A, Steele A, MacRae J, Pauly RP, Perkins DN, Chan CT, Copland M, Komenda P, McFarlane PA, Lindsay R, Pierratos A, Nesrallah GE. Vascular access for intensive maintenance hemodialysis: a systematic review for a Canadian Society of Nephrology clinical practice guideline. Am J Kidney Dis. 2013 Jul;62(1):112-31

http://www.vascularaccesssociety.com/guidelines.html

[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]

I. Planning for Vascular Access *(Supports proactive planning for ESRD and the use of fistula over catheter)*

1. Each center should establish a dedicated team for vascular access. (Grade D, opinion)

2. Preserve arm veins suitable for placement of vascular access. Preservation should begin in patients with progressive kidney disease and an estimated GFR of less than 30 ml/min. (Grade D, opinion)

3. The preferred type of vascular access is a radio-cephalic native vessel arteriovenous fistula. (Grade C)

II. Access Timing, Placement, and Maturation *(Supports proactive planning for ESRD)*

1. Establish AV fistulae when the patient has an estimated GFR of 15 to 20 ml/min and progressive kidney disease. (Grade D, opinion)

[Response Ends]

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

[Response Begins]

Grade D, opinion

Grade C

The grading of the evidence supporting each recommendation is based upon the scheme developed by the Canadian Hypertension Education Program, see citation

Zarnke KB, Campbell NR, McAlister FA, Levine M: A novel process for updating recommendations for managing hypertension: Rationale and methods. *Can J Cardiol* 16: 1094–1102, 2000

[Response Ends]

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

[Response Begins]

The grading of the evidence supporting each recommendation is based upon the scheme developed by the Canadian Hypertension Education Program, see citation

Zarnke KB, Campbell NR, McAlister FA, Levine M: A novel process for updating recommendations for managing hypertension: Rationale and methods. *Can J Cardiol* 16: 1094–1102, 2000

[Response Ends]

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

[Response Begins]

The grading of the evidence supporting each recommendation is based upon the scheme developed by the Canadian Hypertension Education Program, see citation

Zarnke KB, Campbell NR, McAlister FA, Levine M: A novel process for updating recommendations for managing hypertension: Rationale and methods. *Can J Cardiol* 16: 1094–1102, 2000

[Response Ends]

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

[Response Begins]

The grading of the evidence supporting each recommendation is based upon the scheme developed by the Canadian Hypertension Education Program, see citation

Zarnke KB, Campbell NR, McAlister FA, Levine M: A novel process for updating recommendations for managing hypertension: Rationale and methods. *Can J Cardiol* 16: 1094–1102, 2000

[Response Ends]

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

[Response Begins]

NA

[Response Ends]

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

[Response Begins]

Section A. Hemodialysis via a hemodialysis catheter has worse outcomes than hemodialysis by arteriovenous fistula or arteriovenous graft

Evidence indicates increased mortality and morbidity when starting hemodialysis with a catheter. A recent systematic review indicates persons using catheters for hemodialysis tend to have the highest risks for death, infections, cardiovascular events, and hospitalization compared with other vascular access types.

Ravani P, Palmer SC, Oliver MJ, Quinn RR, MacRae JM, Tai DJ, Pannu NI, Thomas C, Hemmelgarn BR, Craig JC, Manns B, Tonelli M, Strippoli GF, James MT. Associations between hemodialysis access type and clinical outcomes: a systematic review. J Am Soc Nephrol. 2013 Feb;24(3):465-73. doi: 10.1681/ASN.2012070643. Epub 2013 Feb 21. Review. PubMed PMID: 23431075; PubMed Central PMCID: PMC3582202.

Note: There is an earlier systematic review (Mustafa) which included twenty studies compared to Ravani with 62. There is considerable overlap in the studies utilized by both.

[Response Ends]

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

[Response Begins]

NA

[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]

NA

[Response Ends]

Group 5 - Evidence - Systematic Reviews Table

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

[Response Begins]

Comparisons of quality of life between patients underwent peritoneal dialysis and hemodialysis: a systematic review and meta-analysis

Anan Chuasuwan, Siriporn Pooripussarakul, Ammarin Thakkinstian, Atiporn Ingsathit & Oraluck Pattanaprateep

Health Qual Life Outcomes 18, 191 (2020).

https://doi.org/10.1186/s12955-020-01449-2

[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]

Since this is not a guideline, the conclusion from the systematic review is provided: "Patients with chronic kidney disease (CKD) stage 5 or ESRD treated with PD had better generic HRQoL measured by SF-36 and EQ-5D than HD patients. In addition, PD had higher specific HRQoL by KDQOL than HD patients in subdomain of physical functioning, role limitations due to emotional problems, effects and burden of kidney disease."

[Response Ends]

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

[Response Begins]

This paper is is a systematic review and not a guideline, therefore, it does not have any recommendation, only conclusions. The authors evaluated the quality of the individual studies using the Newcastle-Ottawa Scale, which is designed for evaluation of non-randomized studies, and the evaluation is shown below.

Table 1. Risk of bias assessment for cohort by Newcastle Ottawa Scale (NOS)

Author (Year)	Study Design	Selection (Max 5) - Representativeness of the sample	Selection (Max 5) - Sample size	Selection (Max 5) - Non-respondents	Selection (Max 5) - Ascertainment of the exposure	Comparability (Max 2)	Outcome (Max 3) - Assessment of the outcome	Outcome (Max 3) - Statistical test	
Merkus (1997) [15]	CS	*	*	*	*	*	*	*	
Wight (1998) [40]	CS	*	*	*	*	*	*	*	
Blake (2000) [41]	CS	*	*	*	**	*	*	*	
Diaz-Buxo (2000) [42]	CS	*	*	*	**	*	*	*	
Wasserfallen (2004) [44]	CS	*	*	*	*	**	*	*	
Lee (2005) [26]	CS	*	*	*	*	*	*	*	
Kutner (2005) [45]	CS	*	*	*	*	**	*	*	
Kalender (2007) [4]	CS	*	*	*	*	**	*	*	
Zhang (2007) [46]	CS	*	*	*	*	*	*	*	
Sayin (2007) [47]	CS	*	*	*	**	*	*	*	
Borowiak (2009) [48]	CS	*	*	*	**	*	**	*	
Kontodimopoulos (2009) [18]	CS	*	*	*	**	*	*	*	
Ibrahim (2011) [49]	CS	*	*	*	**	*	*	*	
Turkmen (2012) [50]	CS	*	*	*	**	**	*	*	
Okpechi (2013) [51]	CS	*	*	*	**	**	*	*	
Czyzewski (2014) [52]	CS	*	*	*	**	*	*	*	
Yang (2015) [14]	CS	*	*	*	**	*	*	*	
Kostro (2016) [2]	CS	*	*	*	*	**	*	*	
Chang (2016) [53]	CS	*	*	*	*	*	*	*	

Author (Year)	Study Design	Selection (Max 4) - Representativeness of the exposed cohort	Selection (Max 4) - Selection of the non exposed cohort	Selection (Max 4) - Ascertainment of the exposure	Selection (Max 4) - Demonstration that outcome of interest was not present at start of study	Comparability (Max 2)	Outcome (Max 3) - Assessment of outcome	Outcome (Max 3) - Was follow-up long enough for outcomes to occur	Outcome (Max 3) - Adequacy of follow up of cohorts
Merkus (1999) [10]	Pro	*	*	*	*	*	*	*	*
Harris (2002) [43]	Pro	*	*	*	**	**	*	*	*

*Indicates cell intentionally left blank

CS, cross sectional; Pro, prospective.

Table 1. Risk of bias assessment for cohort by Newcastle Ottawa Scale (NOS)

[Response Ends]

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

[Response Begins]

not applicable.

[Response Ends]

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

[Response Begins]

not applicable

[Response Ends]

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

[Response Begins]

not applicable

[Response Ends]

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

[Response Begins]

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement guideline and registered at PROSPERO (number CRD42016048574). A total of 21 out of 7995 studies published between 1997 and 2016 were eligible, see Figure 1, and included around 29,000 participants (6035 PD and 22,967 HD).

Figure 1. Flow chart of study selection

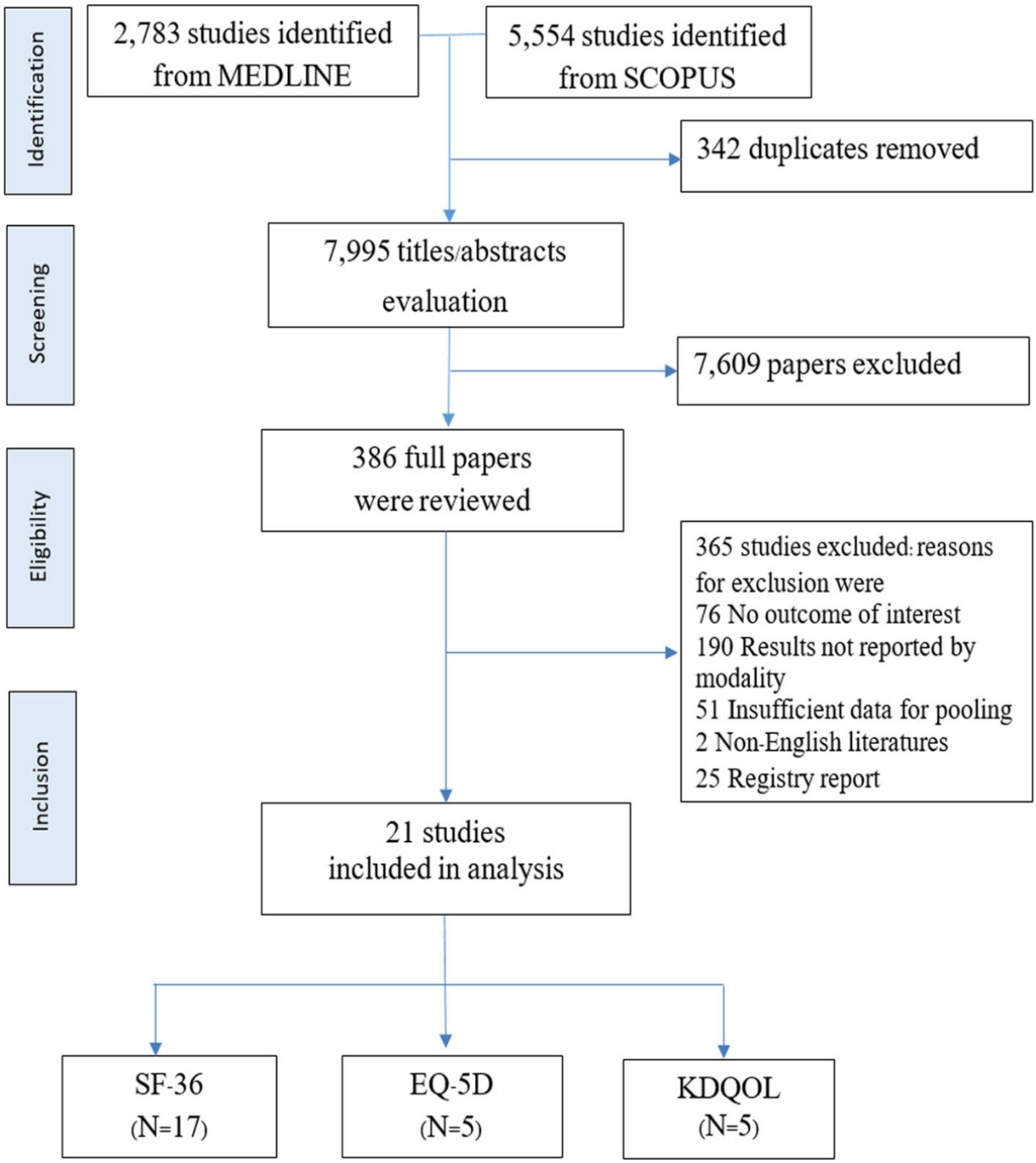


Figure 1. Flow chart of study selection

[Response Ends]

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

[Response Begins]

The systematic review and meta-analysis performed to compare HRQoL between PD and HD used both generic (SF-36) and kidney specific QOL (KDQOL) instrument, and another common generic HRQoL, EQ-5D. According to the authors, this approach can be included in additional cost-effectiveness analysis or cost-utility analysis. Not only that, the authors stated that by analyzing each subdomain of the instrument, this could help to understanding the real individual affecting

subdomain that contribute to patient HRQoL and may provide useful information and an opportunity to find the solution for specific management plan.

[Response Ends]

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

[Response Begins]

Not identified.

[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]

None conducted

[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]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[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]

Unplanned onset of ESRD has been termed "Catastrophic Onset of ESRD" and is associated with long hospitalizations and high mortality in the first year of dialysis. Unprepared and uneducated patients must be admitted to the hospital and

receive urgent dialysis, usually via a central venous catheter which has been shown to cause high rates of blood stream infection and higher mortality than surgically prepared vascular access. The Optimal ESRD Starts measure quantifies the ability of a health care system to prepare patients for ESRD by identification of high risk patients, educating them and their families about the need for dialysis or kidney transplant, helping them to make appropriate and informed choices, and then successfully transitioning to dialysis or kidney transplantation as kidney function declines to the level of ESRD.

As demonstrated in the Medical Evidence Review, Optimal ESRD Starts are better for patients, who experience fewer blood stream infections, have fewer hospital days, fewer cardiovascular events and lower mortality. Optimal ESRD Starts save resources as shown by the Canadian STARRT Trial (average estimated cost $23,965 less per patient in the first 6 months of ESRD) and by KP data (average costs $47,000 less per patient in the first 6 months of ESRD with 14.1 fewer hospital days in first 6 months of ESRD - data based previous submission). This measure has been utilized successfully to improve outcomes in Kaiser Permanente Southern California for nearly 15 years and in the national Kaiser Permanente program for 8 years.

There is a significant performance gap in Optimal ESRD Starts between the estimated U.S. performance, estimated at 30.4% in 2017 as estimated from United States Renal Data Service and CMS Fistula First data and the Kaiser Permanente national performance at 57.5% in December 2017 and 56.5% in September 2021, demonstrating the opportunity for improvement.

[Response Ends]

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

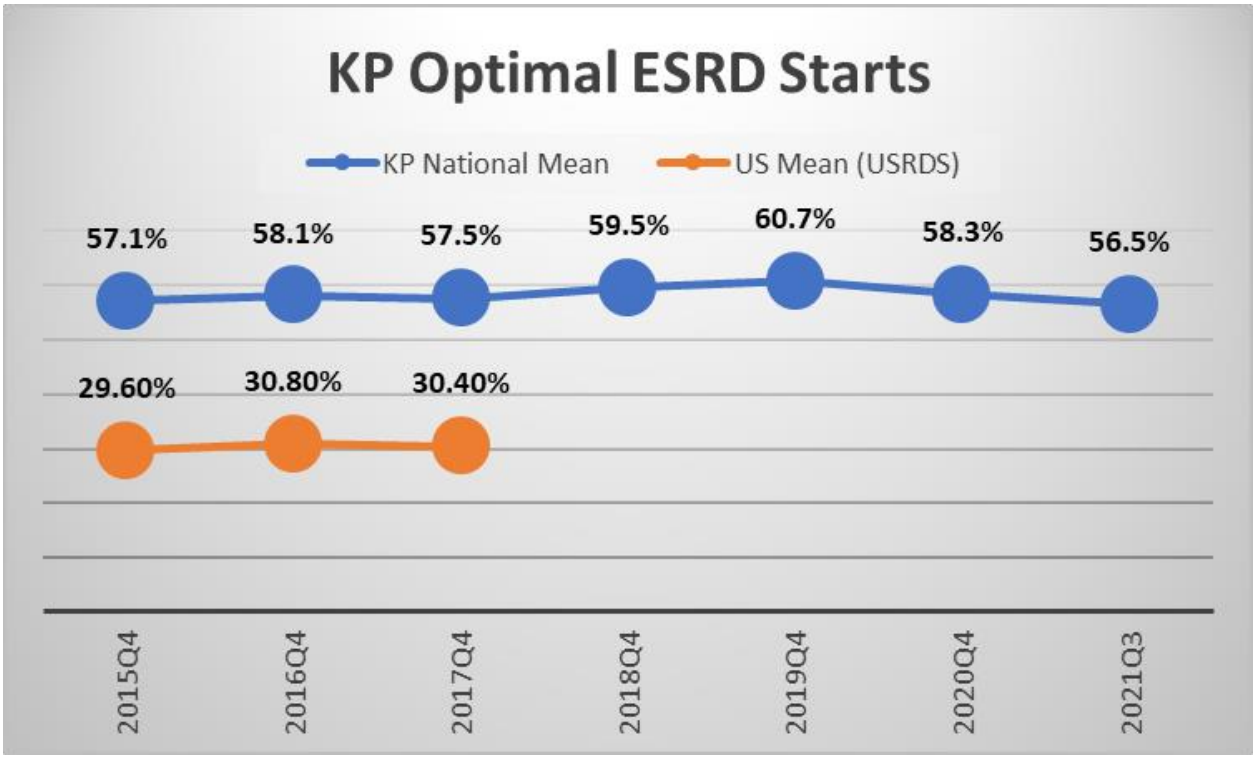
Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement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

[Response Begins]

Performance scores across the Kaiser Permanente (KP) Regions for the last seven years are shown graphically below. Over six consecutive annual measurement periods the KP national mean has improved from 57.1% in December 2015 to 58.3% in December 2020 but saw a slight decline at 56.5% in September 2021. For the most recent measurement period (October 1, 2020 to September 30, 2021) the total number of new ESRD patients was 4,002, ranging from 103 to 1,806 patients in the eight measured Kaiser Permanente regions, with regional minimum being 42% and the maximum being 65%.

Performance for the U.S. population can be estimated from USRDS and CMS Fistula First data and was approximately 30.4% for 2017. Thus a significant performance gap exists between this U.S. performance and what can be achieved.

Line Graph. A line graph that depicts the annual measure rate since 2015 up to 2021 for Kaiser Permanente and compares the trend line to the US annual rate.



Line Graph. A line graph that depicts the annual measure rate since 2015 up to 2021 for Kaiser Permanente and compares the trend line to the US annual rate.

Since all new ESRD patients are included (there is no sampling), and the result for each patient is either yes/Optimal or no/non-Optimal, we do not believe a descriptive standard deviation calculation applied to the total KP population is very meaningful. Variation or inaccuracy could be introduced via measurement error (missing data, incorrect reporting of initial modality, date or vascular access) which is addressed in the Reliability and Validity Testing section.

[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]

Not applicable

[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]

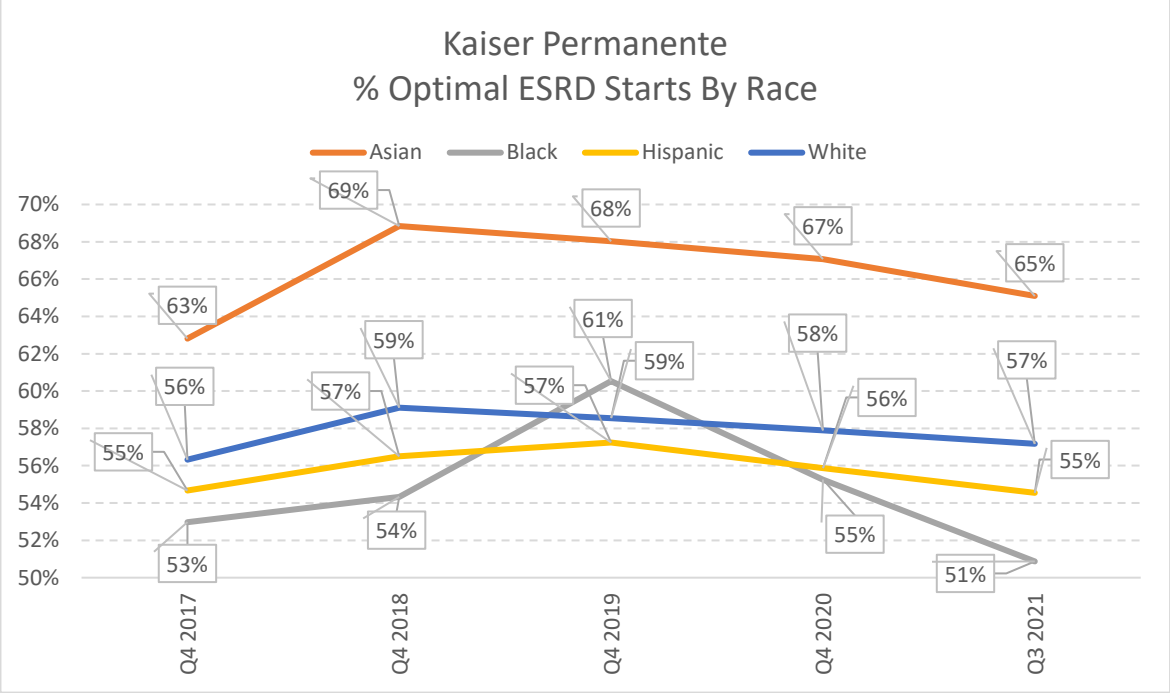
For the most recent measurement period (October 1, 2020 to September 30, 2021), the total number of new ESKD patients was 4,002, ranging from 103 to 1,806 patients in the respective eight Kaiser Permanente markets. Table 1 below shows the characteristics of the patient population at Kaiser Permanente as of measurement period Q3 2021. The line graph below shows the trend line for Optimal ESRD Starts rate by race at Kaiser Permanente from 2017 to 2021. The trend line doesn't show differences between white, black, and Hispanic for Optimal ESRD Starts.

Table 1. Characteristics of ESKD Kaiser Permanente Patient Population In Optimal ESRD Starts Denominator As of Q3 2021

Characteristics	**New KP ESKD Patients during measurement period**	**KP Optimal ESRD Starts during measurement period**
n	4,002	2,268
Age, mean (SD)	64 (14)	64 (14)
Female	39.4%	40.4%
Race	*	*
White	27.8%	28.0%
Black	18.4%	16.4%
Hispanic	27.9%	26.6%
Asian Pacific Islander	21.0%	23.9%
Alaskan Native American	0.3%	0.1%
Multi racial	1.8%	1.9%
Missing race	2.9%	3.1%

Table 1. Characteristics of Kaiser Permanente Patient Population In Optimal ESRD Starts Denominator As of Q3 2021

*cells intentionally left empty



Line Graph 1. Line graph showing Kaiser Permanente % Optimal ESRD Starts Rate By Race

[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]

Information From 2022

We recognized that there is a gap in care that warrants a national performance measure such as the Optimal ESRD Starts due to poor pre-ESRD care coordination and lack of established multi-disciplinary care teams.

Information From 2015 Submission

The primary focus of this measure are 1) preemptive kidney transplantation, 2) initial dialysis therapy at home by PD or HD, and 3) initial hemodialysis via AVF or AVG. An important contributing factor is 4) access to nephrology care prior to ESRD.

1) While there a number of articles addressing disparities in kidney transplantation, only a single article was located addressing disparities in preemptive kidney transplantation, concluding that preemptive deceased donor kidney transplant occurs most often among Caucasians (versus Blacks) with private insurance.
2) There were no recent articles located that address disparities impacting dialysis modality choice (PD vs. HD) in the US. An article from 1999 suggests there may be disparities, but is not specific. Also, there is great disparity in the utilization of peritoneal dialysis between different countries, with the US being one of the lower.
3) One article was located suggesting that lower initial AVF creation in African American patients may be due to differences in arm vein diameters. Another article addresses geographic variability, gender and racial disparities in hemodialysis access.
4) A single article was located suggesting that zip codes with higher black populations are associated with lower access to nephrology care, before requiring dialysis.

Focus 1) Preemptive kidney transplantation: This article addresses racial and insurance type disparity in preemptive kidney transplant, deceased donor.

Clin J Am Soc Nephrol. 2013 Apr;8(4):575-82. doi: 10.2215/CJN.05310512. Epub 2013 Jan 31. Preemptive deceased donor kidney transplantation: considerations of equity and utility. Grams ME(1), Chen BP, Coresh J, Segev DL. (conclusion above)

Articles identifying disparities in kidney transplant rates by race, age, socioeconomic status and type of insurance (not necessarily preemptive)

Clin J Am Soc Nephrol. 2012 Sep;7(9):1490-7. doi: 10.2215/CJN.13151211. Epub 2012 Jul 26. Association of race and insurance type with delayed assessment for kidney transplantation among patients initiating dialysis in the United States. Johansen KL(1), Zhang R, Huang Y, Patzer RE, Kutner NG.

CONCLUSIONS: Racial and insurance-related disparities in transplant assessment potentially delay transplantation, particularly among younger patients.

J Natl Med Assoc. 2007 Aug;99(8):923-32. Trends in kidney transplantation rates and disparities. Stolzmann KL(1), Bautista LE, Gangnon RE, McElroy JA, Becker BN, Remington PL.

CONCLUSION: These results demonstrate a growing disparity in transplantation rates by demographic characteristics and a consistent disparity in transplantation by socioeconomic characteristics.

Clin J Am Soc Nephrol. 2013 Dec;8(12):2149-57. Differences in access to kidney transplantation between Hispanic and non-Hispanic whites by geographic location in the United States. Arce CM(1), Goldstein BA, Mitani AA, Lenihan CR, Winkelmayer WC.

CONCLUSIONS: After accounting for geographic location and controlling for competing risks (e.g., Hispanic survival advantage), the disparity in access to deceased donor transplantation was markedly attenuated among Hispanics compared with non-Hispanic whites. To overcome the geographic disparities that Hispanics encounter in the path to transplantation, organ allocation policy revisions are needed to improve donor organ equity.

Am J Transplant. 2013 Jun;13(6):1557-65. Racial differences in determinants of live donor kidney transplantation in the United States. Purnell TS(1), Xu P, Leca N, Hall YN.

CONCLUSIONS: In the United States, significant disparities in rates of LDKT persist, but determinants of these disparities vary by race-ethnicity. Efforts to expand preESKD insurance coverage, to improve access to high-quality predialysis care and to overcome socioeconomic barriers are important targets for addressing disparities in LDKT.

Am J Kidney Dis. 2012 Jun;59(6):849-57. Center-level factors and racial disparities in living donor kidney transplantation. Hall EC(1), James NT, Garonzik Wang JM, Berger JC, Montgomery RA, Dagher NN, Desai NM, Segev DL.

CONCLUSIONS: Racial disparity in attainment of LDKT exists at every transplant center in the country. Centers with higher rates of LDKT attainment for all races had less disparity; these high-performing centers might provide insights into policies that might help address this disparity.

J Natl Med Assoc. 2007 Aug;99(8):923-32.Trends in kidney transplantation rates and disparities. Stolzmann KL(1), Bautista LE, Gangnon RE, McElroy JA, Becker BN, Remington PL.

CONCLUSION: These results demonstrate a growing disparity in transplantation rates by demographic characteristics and a consistent disparity in transplantation by socioeconomic characteristics. Future studies should focus on identifying specific barriers to transplantation among different subpopulations in order to target effective interventions.

Focus 2) Initial therapy is a Home Dialysis modality:

Perit Dial Int. 1999;19 Suppl 2:S419-22. Socioeconomic aspects of peritoneal dialysis in North America: role of non medical factors in the choice of dialysis. Venkataraman V(1), Nolph KD.

SUMMARY: Patients initiating dialysis therapy must make a choice between hemodialysis (HD) and peritoneal dialysis (PD). Controversy persists over the relative merits of each modality in the treatment of end-stage renal disease (ESRD). Issues relating to survival, morbidity, economics, and patient characteristics will all determine the final choice of therapy. Non-medical factors are the most important determinant of dialysis modality selection. In the United States, HD has been the more commonly used modality, while PD is underrepresented. This disparity arises from multiple factors including reactions (sometimes incorrect) to the healthcare financing structure, physician biases, and changing demographic

patterns in the ESRD population. We discuss these issues and present collected evidence showing that increased use of PD may have substantial overall benefit.

J Am Soc Nephrol. 2012 Mar; 23(3): 533–544. Global Trends in Rates of Peritoneal Dialysis. Arsh K. Jain, Peter Blake, Peter Cordy and Amit X. Garg

SUMMARY: PD prevalence varies from 79.4% in Hong Kong to 0.7% in Luxembourg with 7.0% in the US. "Several economic influences, including health care financing and delivery, physician reimbursement, and resource availability, have been suggested to affect trends in use. For example, countries with private dialysis providers generally use PD for a smaller proportion of dialysis patients than countries in which public providers dominate. The proliferation of HD units in some countries has increased the availability of HD, creating an incentive to use that capacity rather than home dialysis modalities. Some have raised concerns that nephrology training programs are deficient in PD and do not adequately prepare young nephrologists to provide care for PD patients."

Focus 3) Initial hemodialysis vascular access is an arteriovenous fistula or an arteriovenous graft:

J Vasc Surg. 2012 Aug;56(2):424-31. Ethnic differences in arm vein diameter and arteriovenous fistula creation rates in men undergoing hemodialysis access. Ishaque B(1), Zayed MA, Miller J, Nguyen D, Kaji AH, Lee JT, O´Connell J, de Virgilio C.

CONCLUSIONS: African American patients are less likely than non-African American patients to undergo AVF during first-time hemodialysis access surgery. This ethnic discrepancy appears to be due to smaller arm vein diameters in African American patients. In African American patients with appropriate vein diameters who do undergo AVF, primary and functional patencies are equivalent to non-African American patients.

J Am Soc Nephrol. 2002 Aug;13(8):2117-24. National profile of practice patterns for hemodialysis vascular access in the United States. Reddan D(1), Klassen P, Frankenfield DL, Szczech L, Schwab S, Coladonato J, Rocco M, Lowrie EG, Owen WF Jr; National ESRD CPM Work Group.

CONCLUSION: Despite translation of practice guidelines for hemodialysis vascular access into national CPMs, there is substantial disparity in angioaccess (for hemodialysis) allocation in the United States. Quality improvement strategies to improve the prevalence of fistulae should focus on selected regions and include physician education about their practice patterns and potential biases.

Focus 4) access to nephrology care prior to ESRD

J Am Soc Nephrol. 2010 Jul;21(7):1192-9.Racial composition of residential areas associates with access to pre-ESRD nephrology care. Prakash S(1), Rodriguez RA, Austin PC, Saskin R, Fernandez A, Moist LM, O´Hare AM.

SUMMARY: Referral to a nephrologist before initiation of chronic dialysis occurs less frequently for blacks than whites, but the reasons for this disparity are incompletely understood. Here, we examined the contribution of racial composition by zip code on access and quality of nephrology care before initiation of renal replacement therapy (RRT). We retrospectively studied a cohort study of 92,000 white and black adults who initiated RRT in the United States between June 1, 2005, and October 5, 2006. The percentage of patients without pre-ESRD nephrology care ranged from 30% among those who lived in zip codes with <5% black residents to 41% among those who lived in areas with >50% black residents. In adjusted analyses, as the percentage of blacks in residential areas increased, the likelihood of not receiving pre-ESRD nephrology care increased. Among patients who received nephrology care, the quality of care (timing of care and proportion of patients who received a pre-emptive renal transplant, who initiated therapy with peritoneal dialysis, or who had a permanent hemodialysis access) did not differ by the racial composition of their residential area. In conclusion, racial composition of residential areas associates with access to nephrology care but not with quality of the nephrology care received.

[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.

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins] Yes

Three changes were made to this measure to account for clinical guideline changes as well as to respect patient choice:

1. to remove the criteria of 10% new hemodialysis patient limit for arteriovenous graft (AVG).
2. to include patients with failing kidney transplants starting or returning to dialysis

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]

- Remove the criterion of 10% new hemodialysis patient limit for arteriovenous graft (AVG). This change is based on recommendations from the KDOQI 2019 Vascular Access guidelines (website: https://www.ajkd.org/article/S0272-6386(19)31137-0/fulltext). Removal of this criterion did not have significant impact on the reported AVG rate over time.
- Include patients with failing kidney transplants starting or returning to dialysis. This is based on various publications that looked at the high number of failed transplant patients who then went on dialysis (1,2). This group of patient population made up about 3% of the measure denominator population and therefore did not have significant impact on the reported measurement rate.

1. Chan M. Initial Vascular Access Type in Patients with a Failed Renal Transplant. CJASN. 2014; 1225–1231.

2. Perl J. Reduced survival and quality of life following return to dialysis after transplant failure: the Dialysis Outcomes and Practice Patterns Study. Nephrol Dial Transplant. 2012; 4464–4472.

[Response Ends]

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]

Optimal End Stage Renal Disease (ESRD) Starts

[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]

Optimal End Stage Renal Disease (ESRD) Starts is the percentage of new adult ESRD patients during the measurement period who experience a planned start of renal replacement therapy by receiving a preemptive kidney transplant, by initiating home dialysis (peritoneal dialysis or home hemodialysis), or by initiating outpatient in-center hemodialysis via arteriovenous fistula or arteriovenous graft.

[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]

Other (specify)

[Other (specify) Please Explain]

Pre-emptive Kidney Transplant

Palliative Care and End-of-Life Care

Renal

Renal: Chronic Kidney Disease (CKD)

Renal: End Stage Renal Disease (ESRD)

Surgery: General Surgery

Surgery: Vascular Surgery

[Response Ends]

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

[Response Begins]

Care Coordination

Care Coordination: Transitions of Care

Disparities Sensitive

Health and Functional Status: Quality of Life

Other (specify)

[Other (specify) Please Explain]

Patient activation and engagement

Person-and Family-Centered Care: Person-and Family-Centered Care

Safety: Healthcare Associated Infections

Screening

[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]

Adults (Age >= 18)

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]

Clinician: Group/Practice

Facility

Health Plan

Integrated Delivery System

Population: Regional and State

[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]

Ambulatory Care

Inpatient/Hospital

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]

none available

[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]

No data dictionary/code table – all information provided in the submission form

[Response Ends]

Attachment: 2594_NQF_Renal_Measure_2594_Data_Elements.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 number of new ESRD patients age 18 and over who initiate outpatient renal replacement therapy in the twelve month measurement period with an optimal ESRD therapy, which includes preemptive kidney transplant, home dialysis (peritoneal dialysis or home hemodialysis), or outpatient in-center hemodialysis via arteriovenous fistula or arteriovenous graft.

[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]

New Information

The Optimal ESRD Starts numerator is the total number of new patients age 18 and over who initiate renal replacement therapy for the first time and do not come off dialysis by 90 days, with one of the following:

- A preemptive kidney transplant or simultaneous pancreas-kidney transplant (SPK). Preemptive means that the patient has never experienced out-patient dialysis, OR

- Initial home or self-dialysis modality, including planned and "successful urgent start" peritoneal dialysis (PD) and home hemodialysis (HHD) via an arteriovenous fistula or arteriovenous graft. ”Successful urgent start” peritoneal dialysis means that the patient never experienced outpatient hemodialysis via a hemodialysis catheter before starting outpatient peritoneal dialysis, OR
- Initial outpatient hemodialysis (HD), including self-hemodialysis (SHD), via arteriovenous fistula (AVF) prepared surgically without use of artificial materials. The patient may have a hemodialysis catheter in place if it is not used. Do not count patient with a single needle in AVF with blood return via catheter, OR
- Initial outpatient hemodialysis (HD), including self-hemodialysis (SHD), via arteriovenous graft (AVG). The patient may have a hemodialysis catheter if it is not used. Do not count patient with a single needle in AVG with blood return via catheter.

From Old Submission

The item underlined in this sentence was removed from the new submission:

- Initial outpatient hemodialysis (HD), including self-hemodialysis (SHD), via arteriovenous graft (AVG), **limited to no more than 10% of all patients starting in-center hemodialysis#**.

 # Arteriovenous fistula (AVF) is highly preferred for hemodialysis over an arteriovenous graft (AVG). AVF are associated with many fewer follow-up encounters with vascular surgery and interventional radiology to remove clots, dilate and replace. CMS has recognized AVF superiority in its Fistula First Quality Initiative, which continues to collect data and promote practice improvement methods.

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

The number of patients age 18 and over who receive a preemptive kidney transplant or initiate long-term dialysis therapy (do not recover kidney function by 90 days) for the first time in the twelve month measurement period

[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]

New Information

The population being measured are patients age 18 and over who 1) receive a preemptive kidney transplant (having never received outpatient dialysis), including simultaneous pancreas and kidney transplant, plus 2) patients age 18 and over initiating long-term maintenance dialysis who do not recover kidney function by 90 days.

The population includes patients who start renal replacement therapy and then are lost to follow up (lose insurance, move away) and/or die.

The denominator is the number of the above patients within the measured entity during the 12-month measurement period.

Clarifications based on the above definition (not exclusions):

1. The denominator does not include patients who initiate outpatient dialysis but then recover GFR to the point where they can stop dialysis treatments by 90 days after the first outpatient dialysis.
3. The denominator does not include patients who previously reached ESRD, such as

- Patients who previously were on dialysis 90 days or more who then recovered kidney function for a while, but then restarted dialysis
- Patients who switch from one dialysis modality to another, for example switching from in-center hemodialysis to home dialysis.

3. The denominator does not include patients who died without experiencing outpatient dialysis or a kidney transplant.

From Old Submission

The item underlined in this sentence was removed from the new submission:

2. The denominator does not include patients who previously reached ESRD, such as

- Patients who previously were on dialysis 90 days or more who then recovered kidney function for a while, but then restarted dialysis
- Patients who switch from one dialysis modality to another, for example switching from in-center hemodialysis to home dialysis.
- **Patients with failing kidney transplants starting or returning to dialysis.**

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

None

[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]

None

[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]

As there is no patient sampling (all patients who reach ESRD are included), there is no stratified sampling.

For comparative purposes and tracking within Kaiser Permanente, the metric has been calculated (stratified) by geographic medical regions or areas. Results by geographic regions/areas are shown here.

Table 1. Table shows % optimal ESRD starts by geographic markets as of 2021 Q3

Geographic Markets	# of New ESRD Patients	# of KP New Patients With Optimal ESRD Starts	% Optimal ESRD Starts
Market1	103	64	62.1%
Market2	162	85	52.5%
Market3	201	117	58.2%
Market4	128	82	64.1%
Market5	132	48	36.4%
Market6	139	79	56.8%
Market7	1,806	970	53.7%
Market8	1,331	815	61.2%
Kaiser Permanente	4,002	2,260	56.5%

Table 1. Table shows % optimal ESRD starts by geographic markets as of 2021 Q3

[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]

No risk adjustment or risk stratification

[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]

New Information

1. The target population is all new ESRD patients as described in sp.15. Data is validated by each market via chart review, compiled and submitted on standardized spreadsheets for quality reporting.
2. Determine denominator:

- Eliminate patients who do not meet the denominator definition sp.15.
    a. Eliminate patients who recovered kidney function by day 90
    b. Eliminate patients who previously were on dialysis 90 days or more who then recovered kidney function then later restarted dialysis
    c. Eliminate patients changing dialysis modality
    d. Eliminate patients who died without experiencing outpatient dialysis or a kidney transplant
- Eliminate patients with incomplete data if unavailable

3. Count patients in each category. Each denominator patient must be assigned to one and only one of the groups below per numerator criteria in sp.13.

    Group A: Preemptive kidney transplant

    Group B: Peritoneal Dialysis (Home)

    Group C: Home Hemodialysis

    Group D: In-center HD with AVF

    Group E: In-center HD with AVG

    Group F: In-center HD with Catheter

4. Note: Denominator = A + B + C + D + E + F
5. Calculate Optimal ESRD Starts = ((A + B + C + D + E'))/Denominator) x 100%
6. Calculate Modality Sub-metrics

- Preemptive Kidney Transplant Starts + (A/Denominator) x 100%
- Home Dialysis Starts = ((B + C))/Denominator) x 100%
- Optimal AVF & AVG Starts = ((D + E'))/Denominator) x 100%
- Non-Optimal ESRD Starts = 100% - Optimal ESRD Starts

From Old Submission

Remove underlined item from sentence below since this item is no longer excluded in the measure specification.

2. Determine denominator:

c. Eliminate patients starting dialysis after failed transplant

[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]

There is no sampling. All patients who reach ESRD and start dialysis or have a preemptive kidney transplant are included.

[Response Ends]

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

[Response Begins]

Claims

Electronic Health Records

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]

The data collection instrument is in Question 1 of the Additional section. It can be completed from records maintained by the renal care team as patients reach ESRD, and submitted to the measure analyst every 3 months.

CMS 2728 Form: Within KP we do not have access to this data, but all the essential data elements are available on the CMS 2728 Form which is submitted for every new ESRD patient in the US (whether they have Medicare coverage or not). The only missing data is the date of stopping dialysis if recover from acute renal failure by 90 days, and in most cases, a 2728 Form is not submitted for these patients. Patients who recover kidney function and stop dialysis by 90 days are not included in the denominator or numerator. We anticipate that this will be the source of data for organizations outside of KP in the future.

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]

Available in attached appendix in Question 1 of the Additional Section

[Response Ends]

2a. Reliability

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.02. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]

No

[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]

No additional risk adjustment analysis included

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

[Response Begins]

Claims

Electronic Health Data

Electronic Health Records

Other (specify)

[Other (specify) Please Explain]

Data request forms faxed back by dialysis facilities and kidney transplant centers

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]

The Data Submission Form (appendix) is completed by each of the participating Kaiser Permanente (KP) regions. The information is abstracted from care coordination databases of ESRD patients/members. Since initiation of dialysis and transplantation outside of KP require authorization, it is believed that no patients are missed. The Data Submission Forms are combined into a single database for calculation of Optimal ESRD Starts and the results are reported as total Optimal ESRD Starts and its components (home dialysis, preemptive kidney transplant, in-center hemodialysis via fistula, graft and catheter). This data collection process has not changed since the last submission of this measure.

Since there was no change in the data collection process since the last submission, we felt the validity testing conducted in the prior submission still holds true.

From Prior 2015 Submission

For validity testing of the data elements, a randomized sample set from the above dataset was selected. For each randomized patient, a data questionnaire was mailed to the dialysis facility where the patient initiated dialysis or to the kidney transplant program where the patient was transplanted. The questionnaire requested verification that the patient had indeed started dialysis or received a kidney transplant and the date of initiation of dialysis or transplantation (verifying denominator inclusion), if dialysis, the dialysis modality and if hemodialysis, the vascular access used for the first treatment (catheter, arteriovenous fistula or arteriovenous graft). Forms that were not returned by deadline were followed up by phone calls and faxes to the facility. All but two were completed and returned; those facilities were no longer under KP contract and claimed they no longer had the information. This information was tested against the information on the Data Submission Form (above).

[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]

Information From 2015 Submission

01-01-2012 - 12-31-2012

[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]

Clinician: Group/Practice

Facility

Health Plan

Integrated Delivery System

[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]

Information From 2015 Submission

All six (6) KP regions participating in the Optimal ESRD Program are included in the validity testing and analysis. Regional membership distribution, in thousands, is 226, 236, 482, 538, 3,385, and 3,566. Regions are located in northern California, southern California, Hawaii, northern Oregon and southern Washington, Colorado, and Georgia.

[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]

Information From 2015 Submission

The Optimal ESRD Starts numerator is the total number of new patients age 18 and over who initiate renal replacement therapy for the first time and do not come off dialysis by 90 days, with one of the following:

- A preemptive kidney transplant or simultaneous pancreas-kidney transplant (SPK). Preemptive means that the patient has never experienced out-patient dialysis, OR
- Initial home or self-dialysis modality, including planned and "successful urgent start" peritoneal dialysis (PD) and home hemodialysis (HHD) via an arteriovenous fistula or arteriovenous graft. ”Successful urgent start” peritoneal dialysis means that the patient never experienced outpatient hemodialysis via a hemodialysis catheter before starting outpatient peritoneal dialysis, OR
- Initial outpatient hemodialysis (HD), including self-hemodialysis (SHD), via arteriovenous fistula (AVF) prepared surgically without use of artificial materials. The patient may have a hemodialysis catheter in place if it is not used. Do not count patient with a single needle in AVF with blood return via catheter, OR
- Initial outpatient hemodialysis (HD), including self-hemodialysis (SHD), via arteriovenous graft (AVG), limited to no more than 10% of all patients starting in-center hemodialysis#. The patient may have a hemodialysis catheter if it is not used. Do not count patient with a single needle in AVG with blood return via catheter.

An arteriovenous fistula (AVF) is highly preferred for hemodialysis over an arteriovenous graft (AVG). AVF are associated with many fewer follow-up encounters with vascular surgery and interventional radiology to remove clots, dilate and replace. CMS has recognized AVF superiority in its Fistula First Quality Initiative, which continues to collect data and promote practice improvement methods.

Nevertheless, not every patient is suitable for an AVF, and these patients require an AVG for hemodialysis which is still much better than hemodialysis by catheter. In our 3 year experience measuring Optimal ESRD Starts in Kaiser Permanente less than 5% of new hemodialysis patients start with an AVG as their initial access.

[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]

Information From 2015 Submission

There are no differences in the data used.

[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]

No social risk factors were collected for this measure. However, we have patient reported data in our data system.

[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]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

[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]

The data elements used in the OES measure reporting are collected from our 8 Kaiser Permanente regions every quarter on new ESKD members, and the data elements collected include patient unique identifier, dialysis start date or kidney transplant date, initial ESKD modality, initial vascular access, kidney function recovery date, dialysis facility or transplant program, and groupers. As part of their data quality review process, each Kaiser Permanente regional ESKD coordinator goes through a series of checking and testing of each patient or case with their dialysis or kidney transplant coordinators to ensure accuracy of patient identifier, start date, modality and vascular access. Additional chart or systematic review is also conducted to ensure accuracy of the data. After data is submitted to the national reporting team, another set of review is carried out by the reporting team to ensure patient identifier submitted has a match in the electronic health record reporting data table. Once the patient data is aggregated at the regional level, the denominator and numerator is sent back to each Kaiser Permanente region for final review before the data is published.

[Response Ends]

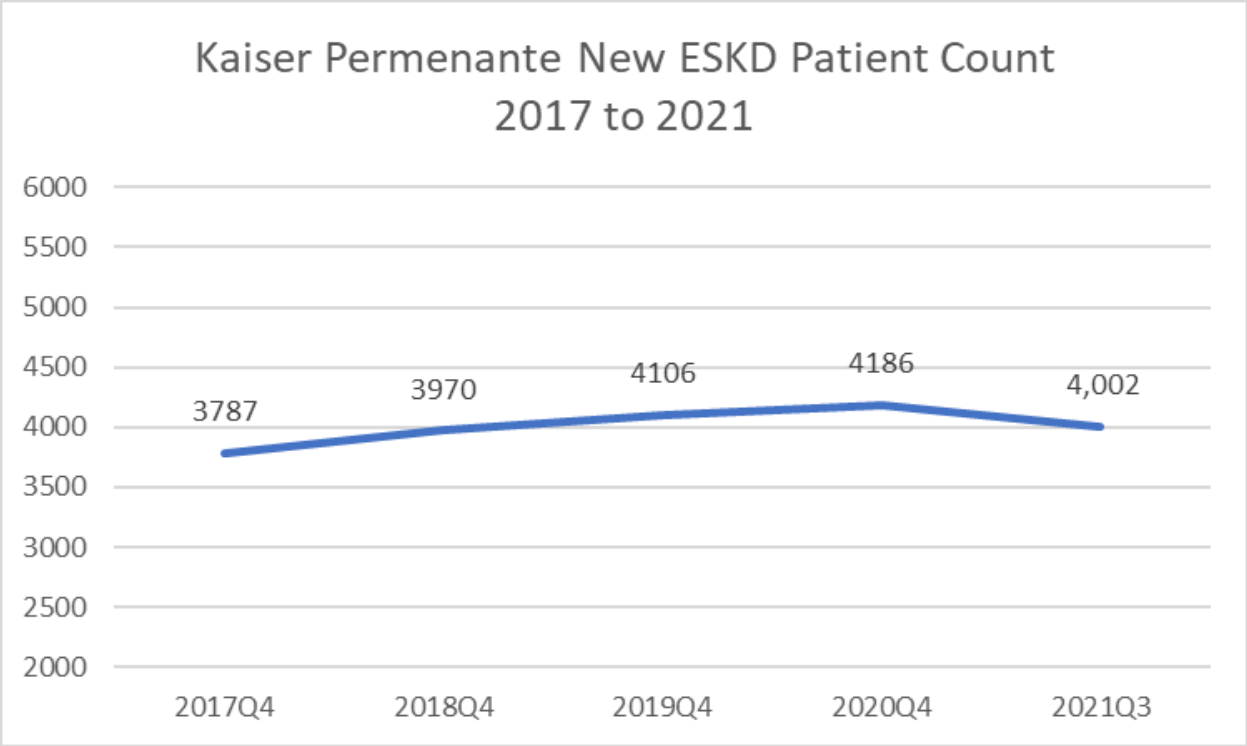
2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

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]

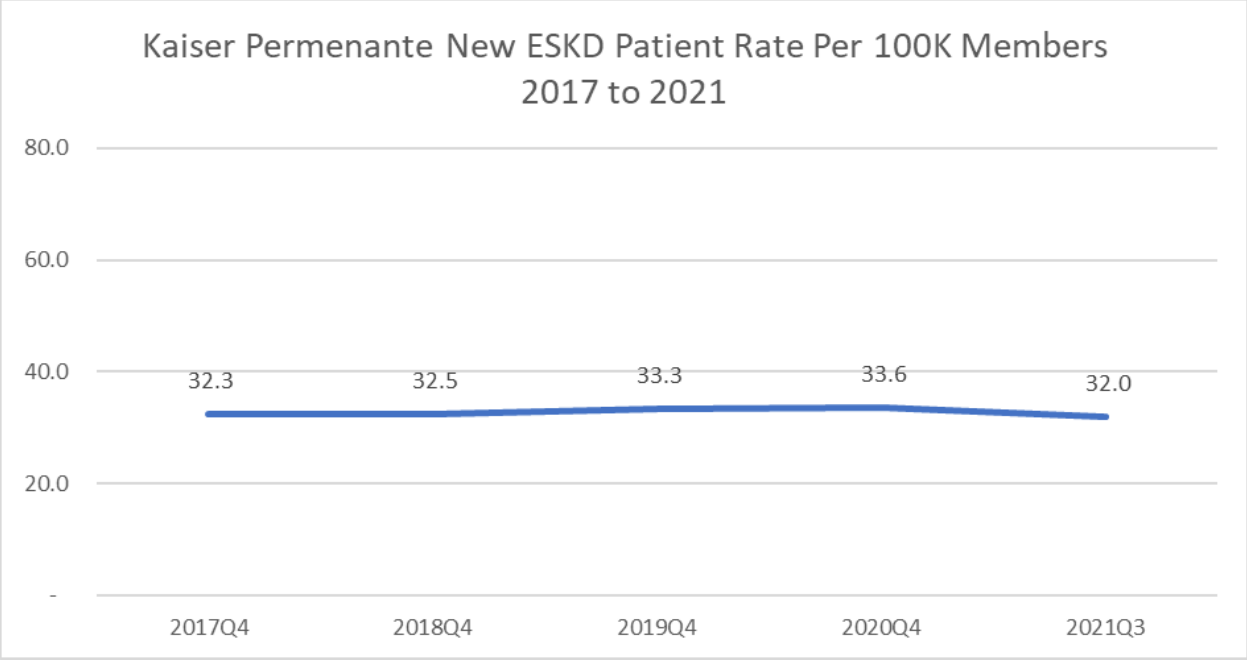
Given that the data collection was performed on all new ESKD members at Kaiser Permenante, reported results were not stratified by sample size. The line graph below shows the number of new ESKD patients reported from 2017 to 2021. Adjusting the patient volume to our annual membership, one can see that the ESKD rate per 100,000 members has a relatively flat line suggesting a stable patient volume over time.

Line Graph 1. Line graph depicts number of new ESKD patients at Kaiser Permanente reported from 2017 to 2021

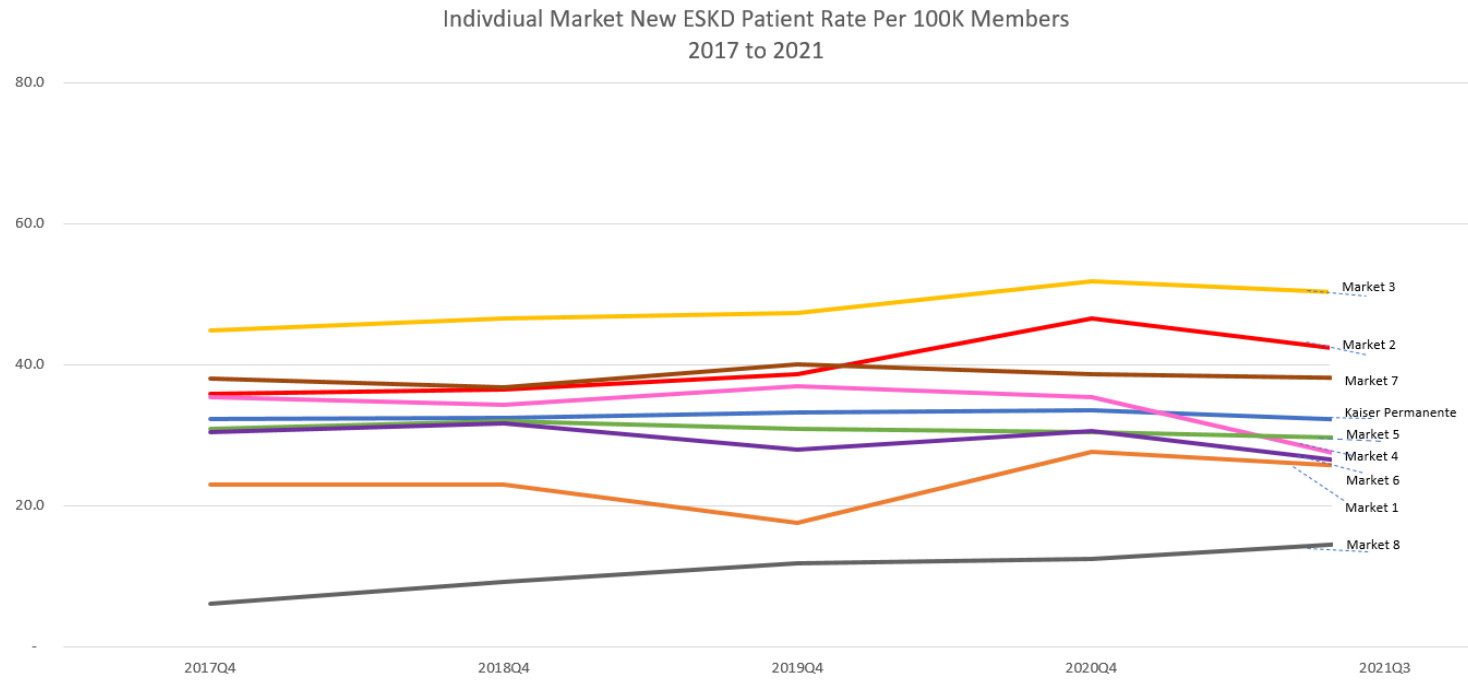


Line Graph 1. Line graph depicts number of new ESKD patients at Kaiser Permanente reported from 2017 to 2021

Line Graph 2. Line graph shows the rate ofLine Graph 2. Line graph shows the rate of new ESKD patients per 100,000 members reported from 2017 to 2021 new ESKD patients per 100,000 members reported from 2017 to 2021



Line Graph 2. Line graph shows the rate of new ESKD patients per 100,000 members reported from 2017 to 2021



Line Graph 3. Line graph shows the individual market's rate of new ESKD patients per 100,000 members reported from 2017 to 2021

Line Graph 3. Line graph shows the individual market's rate of new ESKD patients per 100,000 members reported from 2017 to 2021

[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 annual ESKD patient volume remained steady with a gradual increase from 2017 to 2020 before a small drop in 2021, most likely related to the lockdown from the Covid19 pandemic. Given the annual fluctuations in membership, the reported number of new ESKD patients was adjusted to annual membership and the annual ESKD rate per 100,000 members suggested a stable ESKD patient population over time, as shown in Line Graph 2. Drilling down to the individual market level, it is noted that most of the lines are relatively flat with some markets with smaller population of less than 200 seeing small fluctuations, as shown in Line Graph 3. The analysis showed that the ESKD 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 that the measure rate is precise.

[Response Ends]

2b. Validity

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

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

Empirical validity testing

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[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]

Information From 2022

Patient-Level Testing. As mentioned in section 2a.10, the measure data elements collected for this measure are reviewed and validated at the source ie dialysis and kidney transplant coordinators within each market even before it was submitted to the national reporting team for aggregated reporting. This data collection process has not changed since the last NQF measure submission in 2015. Therefore, the testing result submitted in 2015 is still relevant today.

Empirical Validity Testing. The data elements collected for this measure remained unchanged for the purpose of the measurement. Therefore, the statistical results from the validity testing conducted in 2015 are still applicable. Nevertheless, from a reporting perspective, since 2015, we have made some changes to the reporting specifications, as outlined in the Maintenance Update section: 1) Remove the criteria of 10% new hemodialysis patient limit for arteriovenous graft (AVG), 2) Include patients with failing kidney transplants starting or returning to dialysis. These changes did not change the way the data elements were collected. A validity testing was done showcasing these changes and their impact on the measurement.

Systematic assessment of face validity of performance measure. This measurement is used regularly within Kaiser Permanente and its markets as it provides insights to Kaiser Permanente leadership about the care provided for our chronic kidney disease members. A recent 2021 DaVita white paper entitled "High-Value Care for Chronic Kidney Disease: A Front-Line Solution" authored by the Nephrology Care Alliance Quality Working Group suggested that Optimal ESRD Start is linked to delivering high-value kidney care. Not only that, a new CMS CKCC payment model implemented in 2022 will measure outcomes very similar to the Optimal ESRD Start measurement (Gaurav).

References

Nephrology Care Alliance Quality Working Group, High-Value Care for Chronic Kidney Disease: A Front-Line Solution, Nephrology Care Alliance, https://nephrologycarealliance.com/assets/docs/top-banner/NCA%20Quality%20Metrics%20White%20Paper%202021.pdf?utm_source=&utm_medium=social&utm_term=display_organic&utm_content=&utm_campaign=

Gaurav Jain and Daniel Weiner, Value based care in nephrology: the Kidney Care Choices Model and other Reforms,

Kidney360, Publish Ahead of Print, 10.34067/KID.0004552021

Information From 2015 Submission

We test the accuracy of the regional data compared to the authoritative source comparing the renal replacement therapy information submitted by the regional care coordinator (region) to that provided by the renal replacement therapy provider on record (source), i.e. dialysis unit where the initial dialysis occurred (hemodialysis) or managed (peritoneal dialysis) or the transplant center on record as performing the kidney transplant.

Among the data element elements, we test our a priori hypothesis of a .9 match (90% accuracy) between the region and source data. This is performed at three levels, denominator, numerator, and total.

1. At the denominator level, we test the accuracy of the qualification elements, for example, a first time dialysis. We measure the proportion of records where the region response and source matched and use a 95% confidence interval (CI) to determine if the estimated proportion is different from our hypothesized value of .9.

2. At the numerator level, we compare the method of renal replacement submitted by the region to the authoritative source. This is to assess the accuracy of the regional data element identifying the mode of initial renal replacement therapy. A 95% CI is used to determine if the estimated proportion differs from our hypothesized value of .9.
3. Total element accuracy, denominator and numerator accuracy combined, is tested comparing the match proportion to our hypothesized value of .9 and a 95% CI.
4. We evaluate the performance metric, optimal ESRD start, using sensitivity, specificity, and positive and negative predictive values with 95% confidence intervals.
5. There are two sample cases where the source, dialysis center, is no longer under contract. For these two cases the source results are not available. To judge the potential effect of the missing data, we consider two scenarios, (1) the region result does not match the source (worst case), and (2) the region result does match the source (best case).

Data element match proportion confidence intervals are calculated such that the lower confidence limit is never less than zero and the upper confidence limit is never greater than one (Fleiss).

Fleiss JL (1981). Statistical Methods for Rates and Proportions, 2nd edition. John Wiley & Sons, New York.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

Information From 2022

Given that there was no change in the data collection process, the prior statistical results from the validity testing conducted in 2015 are still relevant. However, as mentioned in 2b.02, we have made some changes to the reporting specifications. Below is the result showing the impact of the changes at the individual measurement component rate level.

1) Remove the criteria of 10% new hemodialysis patient limit for arteriovenous graft (AVG) in 2016. Comparison of the impact of removal of this criterion on the individual measurement component result for measurement periods 2016 and 2021:

Table 1. Table shows percent by AVG, flag for AVG limit in place, flag for AVG limit reached between 2021 and 2016 AVG rate

Markets	Year	2016 %AVG	2016 - AVG limit of <10% in Optimal Start Measure in place	2016 - AVG 10% Reached?	Year	2021 %AVG	2021 - AVG limit of <10% in Optimal Start Measure in place	2021 - AVG 10% Reached?
Market 1	2016	0.0%	Y	N	2021	1.0%	N	N
Market 2	2016	4.8%	Y	N	2021	4.3%	N	N
Market 3	2016	0.0%	Y	N	2021	6.5%	N	N
Market 4	2016	7.5%	Y	N	2021	5.5%	N	N

Markets	Year	2016 %AVG	2016 - AVG limit of <10% in Optimal Start Measure in place	2016 - AVG 10% Reached?	Year	2021 %AVG	2021 - AVG limit of <10% in Optimal Start Measure in place	2021 - AVG 10% Reached?
Market 5	2016	5.3%	Y	N	2021	7.6%	N	N
Market 6	2016	3.2%	Y	N	2021	0.7%	N	N
Market 7	2016	2.1%	Y	N	2021	3.2%	N	N
Market 8	2016	6.8%	Y	N	2021	3.2%	N	N
Kaiser Permanente	2016	4.1%	Y	N	2021	3.5%	N	N

Table 1. Table shows percent by AVG, flag for AVG limit in place, flag for AVG limit reached between 2021 and 2016 AVG rate

2) Include patients with failing kidney transplants starting or returning to dialysis.

Table 2. Table shows number of new ESKD patients before and after the inclusion of the criteria, and the difference between the two denominators, using 2021 data.

Entity	Year	denominator 1 (include pat failing kidney tx and went on dialysis)	denominator 2 (exclude pat failing kidney tx and went on dialysis)	Difference between denominator 1 and denominator 2	% Difference
Market 1	2021	103	103	0	0.0%
Market 2	2021	162	159	3	1.9%
Market 3	2021	201	196	5	2.5%
Market 4	2021	128	125	3	2.3%
Market 5	2021	132	132	0	0.0%
Market 6	2021	139	132	7	5.0%
Market 7	2021	1806	1728	78	4.3%
Market 8	2021	1331	1293	38	2.9%
Kaiser Permanente	2021	4002	3868	134	3.3%

Table 2. Table shows number of new ESKD patients before and after the inclusion of the criteria, and the difference between the two denominators, using 2021 data.

Information From 2015 Submission

Among the 73 sample records, there are 12 errors comprised of 9 errors at the numerator level, where the region recorded mode of renal replacement does not match the source record (AV Fistula and Catheter), and 3 errors at the denominator level, where the patient does not meet the qualifying criteria (does not start renal replacement or has

previously started treatment). Two records have missing source results due to the dialysis center not currently under contract, Table 1.

For the denominator, numerator, and total element match proportions, Table 2, we exclude the two records with no source results (scenario 1, exclude). To consider the potential effect of the excluded records, we include the records under two additional scenarios, that the region result does not match the source (scenario 2, as error) and where the region result matches the source (scenario 3, as match).

1. For the denominator match, the match proportion is .96 (68/71). This is the proportion of sample records where the initial regional submission matches the result from re-examination. The 95% lower and upper confidence limits are .87 and .99, respectively. Under the alternative scenarios, the denominator accuracy of the region is .93 (68/73) with CI .84-.98, if the missing validation data are considered errors and .96 (70/73) with CI .88-.99 if the two missing results are considered matches.
2. For the numerator match, the match proportion is .87 (59/68). The three denominator record errors are excluded. The 95% lower and upper confidence limits are .76 and .93. Assuming that the two cases without source data do not match the source, the region accuracy is .84 (59/70) with CI .73-.92. Assuming that the regional data match the source, then the accuracy is .87 (61/70) with CI .76-.94.
3. For the total element match, the match proportion is .83 (59/71). The 95% lower and upper confidence limits are .71 and .91. Assuming that the region results does not match the missing source results, the match proportion is .81 (59/73) and CI .70-.89. Assuming that the regional data matches the source, then the accuracy is .84 (61/73) with CI .73-.91.
4. Collapsing Table 1 using the optima ESRD start definition yields the distribution for the performance metric, Table 3, a 2 x 2 table. The region results are displayed in the rows and source in the columns. For example, among the total of 68 cases, the regions list 35 of the cases as having an optimal ESRD start. Of these 35, 33 are true optimal ESRD start cases as reported by the source. The region result matches the source in 59 cases, 33 where region and source are both 'Yes' and 26 where region and source are both 'No'. Table 3 is used to calculate the optimal ESRD starts test results, Table 4.
5. For each scenario, Table 4 lists the regional accuracy (match proportion), prevalence of a true (source) optimal ESRD start in the sample, sensitivity and specificity of the regional data to correctly identify an optimal ESRD start among true optimal ESRD start cases and to correctly identify cases that are not an optimal ESRD start among such cases, positive and negative likelihood ratios, and the positive and negative predictive values. For the observed data, excluding the missing data, the match proportion is .87 (59/68) with CI 0.76-0.93, prevalence of a true (source) optimal ESRD start is .59 (40/68). Sensitivity and specificity are .82 with CI (0.67, 0.92) and .93 with CI (0.75-0.99) respectively. Positive and negative predictive values are .94 with CI (0.80, 0.99) and .79 with CI (0.61, 0.90).
6. Results for the missing data scenarios are discussed in section 2b7 Missing Data Analysis and Minimizing Bias.

Table 1. Distribution of treatment counts for region and source

Source (Dialysis Unit/Transplant Center)	*	*	*	*	*	*	*	*
Region	**AV Fistula**	**Catheter**	**Peritoneal**	**Transplant**	**Graft**	**Not first dialysis**	**No source result**	**Total**
AV Fistula	19	2	0	0	0	1	0	22
Catheter	7	26	0	0	0	2	2	37
Peritoneal	0	0	11	0	0	0	0	11
Transplant	0	0	0	3	0	0	0	3

Source (Dialysis Unit/Transplant Center)	*	*	*	*	*	*	*	*
Graft	0	0	0	0	0	0	0	0
Total	26	28	11	3	0	3	2	73
Notes	Not first dialysis: Error for Optimal ESRD Start qualifications, exclude from denominator. No source result: Source information not available, exclude from analysis.	*	*	*	*	*	*	*

Table 1. Distribution of treatment counts for region and source.

* Cell intentionally left empty

Table 2. Match proportion results

*	*	**Denominator**	**Numerator**	**Proportion**	**95% Confidence Interval**	*
Match	**Scenario**	**n**	**x**	**p = x/n**	**LCL**	**UCL**
Total	Exclude	71	59	0.831	0.719	0.906
*	As error	73	59	0.808	0.696	0.888
*	As match	73	61	0.836	0.727	0.909
Numerator	Exclude	68	59	0.868	0.759	0.934
*	As error	70	59	0.843	0.732	0.915
*	As match	70	61	0.871	0.765	0.936
Denominator	Exclude	71	68	0.958	0.873	0.989
*	As error	73	68	0.932	0.841	0.975
*	As match	73	70	0.959	0.877	0.989

Table 2. Match proportion results

* Cells intentionally left blank.

Table 3. Distribution of Optimal ESRD Start counts for Region and Source

Source	*	*	*
Region	**Yes**	**No**	**Total**
Yes	33	2	35
No	7	26	33
Total	40	28	68

Table 3. Distribution of Optimal ESRD Start counts for Region and Source.

Table 4. Optimal ESRD Start metric test results: match proportion, prevalence, sensitivity, specificity, predictive value (95% confidence interval)

*	*	*	*	*	*	*	**Predictive Value**	*
Scenario	**Proportion**	**Prevalence**	**Sensitivity**	**Specificity**	**LR+**	**LR-**	**Positive**	**Negative**
Exclude	0.868 (59/68)	0.588	0.825	0.929	11.5	0.2	0.943	0.788
*	(0.759, 0.934)	*	(0.666, 0.921)	(0.75, 0.988)	*	*	(0.795, 0.99)	(0.606, 0.904)
As error	0.843 (59/70)	0.571	0.786	0.867	5.9	0.3	0.892	0.743
*	(0.732, 0.915)	*	(0.628, 0.892)	(0.684, 0.956)	*	*	(0.736, 0.965)	(0.564, 0.869)
As match	0.871 (61/70)	0.600	0.833	0.933	12.5	0.2	0.946	0.800
*	(0.765, 0.936)	*	(0.68, 0.925)	(0.765, 0.988)	*	*	(0.805, 0.991)	(0.625, 0.909)

Table 4. Optimal ESRD Start metric test results: match proportion, prevalence, sensitivity, specificity, predictive value (95% confidence interval)

* Cells intentionally left blank.

[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]

Information From 2022

The analysis assessing the impact of removing the criteria of 10% new hemodialysis patient limit for arteriovenous graft (AVG) showed that % AVG in 2016 (before removal of 10% limit) and 2021 (after removal of 10% limit) remained below the 10% AVG limit previously set in place. This suggests that % AVG have not been impacted that much by removal of this limiting criterion.

The analysis assessing the impact of including patients with failing kidney transplants starting or returning to dialysis in the measurement found that there was a small difference of 3.3% in the denominator with the criterion versus the denominator without the criterion, with a range of differences from 0% to 5% in the eight markets.

Information From 2015 Submission

At the element level, the accuracy of region data compared to the source is good to excellent with a match in the denominator of .96, in the numerator of .87, and total match of .83. At the 95% confidence level these proportions are not significantly different from our hypothesized proportion of .9. The region data is a valid representation of the source at the element level.

At the performance metric level, accuracy is very good where the region assessment matches the source with a proportion of .87 (59/68) and the 95% CI encompasses values from good to excellent (.76, .93). With a sensitivity proportion of .82, the region assessment is very good at identifying cases that are true optimal ESRD starts with a 95%

confidence that the accuracy is within .67 to .92. Specificity, the probability that the region assessment correctly identifies cases that are not optimal ESRD starts, is excellent, .93.

The probability of a true optimal ESRD start among cases identified by the region as optimal ESRD start, the positive predictive value (PPV), is excellent at .94. Conversely, identification of a true non-optimal ESRD start among region reported cases as non-optimal ESRD starts, negative predictive value (NPV), is good at .79. Both PPV and NPV are dependent upon prevalence of true optimal ESRD starts. As the use of optimal ESRD starts increase (prevalence), the PPV increases and NPV decreases.

A third pair of test assessments, based upon the sensitivity and specificity values, are the likelihood ratios. This pair is independent of the prevalence of the optimal ESRD start. The positive likelihood ratio (LR+), sensitivity/(1 – specificity), is the odds of a true optimal ESRD start among region cases identified as an optimal ESRD start (sensitivity) compared to region cases identified as a non-optimal ESRD start ((1-specificity) is 11.6. That is, the region correctly identifies a true optimal ESRD start 11.6 times more often than it incorrectly identifies a non-optimal ESRD start as optimal. This is more than double the generally accepted rule that a test should have a ratio greater than 5. The negative likelihood ratio (LR-), calculated as (1-sensitivity)/specificity, is the odds of a true optimal ESRD start among cases identified by the region as non-optimal compared to region correctly identified non-optimal ESRD starts among non-optimal. The resulting odds, .19, meets the generally accepted threshold that a LR- should be less than 0.2.

[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]

Information From 2015 Submission

To identify statistically significant and meaningful differences in the optimal ESRD start metric, we compare the regions to the all regions (regional aggregate) performance using CY2012 study performance results. The validity sample is drawn from the same CY2012 population. The regional optimal ESRD start data, number of cases and proportion receiving an optimal ESRD start, is used to test for significant differences among and between the regional performance and the all regions performance. A chi-square test is performed to test if there are any regional differences from the all regions performance. A second set of tests to determine which region is different compares each regional optimal ESRD start proportion to the all regions proportion and a 95% confidence interval is used to determine statistical significance.

[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]

Information From 2022

We updated the statistical results using data from measurement period 2021 Q3. Results comparing the markets to all markets' optimal ESRD start proportion are in Table 1. The table includes the number of cases, the number of optimal

ESRD starts, optimal ESRD start proportion, differences between market and program level results, standard deviation, t-value, p-value and the 95% confidence interval upper and lower bound.

Table 1. Optimal ESRD Start results by region, measurement period 2021 Q3

Entity	N	Optimal ESRD Starts Count	Optimal ESRD Starts Proportion	Differences	Std Dev	t Value	Pr > |t|	Lower 95% CL for Mean	Upper 95% CL for Mean
Market 1	103	64	0.621	0.05	0.487	12.94	<.0001	0.526	0.717
Market 2	162	85	0.531	-0.04	0.501	13.50	<.0001	0.453	0.609
Market 3	201	117	0.587	0.02	0.494	16.86	<.0001	0.518	0.656
Market 4	128	82	0.641	0.07	0.482	15.05	<.0001	0.556	0.725
Market 5	132	48	0.364	-0.20	0.483	8.65	<.0001	0.280	0.447
Market 6	139	79	0.568	0.00	0.497	13.48	<.0001	0.485	0.652
Market 7	1806	970	0.537	-0.03	0.499	45.76	<.0001	0.514	0.560
Market 8	1331	815	0.617	0.05	0.486	46.27	<.0001	0.591	0.643
Kaiser Permanente	4002	2260	0.567	0.00	0.496	72.34	<.0001	0.551	0.582

Table 1. Optimal ESRD Start results by region, measurement period 2021 Q3

Information From 2015 Submission

For the chi-square test, the calculated statistic is 29.73 with 5 degrees of freedom. This is statistically significant (p-value is < .0001).

Results comparing the region to all region optimal ESRD start proportion are in Table 6. The table includes the number of cases, the number of optimal ESRD starts, optimal ESRD start proportion, difference from the all regions performance and the 95% confidence interval. One region, Region 1, is identified as significantly different from the all regions performance. For region 1, among the 110 cases newly diagnosed as ESRD in CY2012, 31 received an optimal ESRD start as identified by the region for a start proportion of .28, difference from the all regions is -.22 with a 95% CI (-0.32, -0.13).

Table 6. Optimal ESRD Start results by region, measurement period CY 2012.

Region	N	Optimal ESRD Start Count	Opt ESRD Start Proportion	Difference	LCL	UCL
1	110	31	0.282	-0.224	-0.315	-0.133
2	87	42	0.483	-0.023	-0.136	0.090

Region	N	Optimal ESRD Start Count	Opt ESRD Start Proportion	Difference	LCL	UCL
3	76	32	0.421	-0.085	-0.204	0.035
4	1031	516	0.500	-0.005	-0.042	0.031
5	134	76	0.567	0.061	-0.029	0.151
6	1162	618	0.532	0.026	-0.009	0.061
All Regions	2600	1315	0.506	*	*	*

Table 6. Optimal ESRD Start results by region, measurement period CY 2012

*This cell was intentionally left blank.

Table 7. Optimal ESRD Start, region to national (.355*) performance comparison, measurement CY2012 using 95% confidence intervals.

Region	N	Optimal ESRD Start Count	Opt ESRD Start Proportion	LCL	UCL	Significance+
1	110	31	0.282	0.202	0.377	NS
2	87	42	0.483	0.375	0.592	Sig
3	76	32	0.421	0.310	0.540	NS
4	1031	516	0.500	0.470	0.531	Sig
5	134	76	0.567	0.479	0.652	Sig
6	1162	618	0.532	0.503	0.561	Sig
All Regions	2600	1315	0.506	0.486	0.525	Sig

Table 7. Optimal ESRD Start, region to national (.355*) performance comparison, measurement CY2012 using 95% confidence intervals

* Estimate based upon USRDS and CMS Fistula First data, see Appendix for calculation

+ NS, Not significantly different, 95% confidence interval contains the national value, LCL < national value < UCL; Sig, Significantly different, 95% confidence interval does not include national value.

[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]

Information From 2022

In the updated statistical analysis using 2021 data, the p-value was less than 0.05 (which means there is significant difference), supporting the same findings back in 2015, see below.

Information From 2015 Submission

The results in Tables 6 and 7 indicate that there is sufficient variation in regional results to identify significant differences in performance among regions both within Kaiser Permanente and to the national optimal ESRD start rate.

The Table 6 results indicate that the optimal ESRD start proportions vary by region (.282-.567) and the regional difference to the all regions group mean difference can be statistically significant (Region 1, difference = -.224, significant using 95% confidence level). Thus indicating that identification of performance differences is possible with the optimal ESRD start metric.

The individual region and the aggregate, all region, results are compared to the 2012 national rate (0.355, estimated from the USRDS and CMS Fistula First data, see Appendix for the calculation), Table 7, using the 95% confidence intervals for the region and all region optimal ESRD start rates. Four of the regional (2, 4, 5, and 6) and the all region aggregate confidence intervals do not include the national optimal ESRD start performance value, .355 indicating statistically significant differences from the national rate.

[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]

Information From 2015 Submission

To quantify the effect of missing data, the performance metric is calculated under the two scenarios (1) for the excluded cases, the region data elements and optimal ESRD start result are assumed not to match the source; and (2) where the region data elements and optimal ESRD start result are assumed to match the source. These two scenarios represent the two extremes for the missing data.

Missing data are only amongst the numerator data elements, the method of renal replacement.

[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]

Information From 2015 Submission

There are two cases with missing source data. Both cases are the result of the dialysis center not being under contract during the validation period with one case from each of the two largest regions. One missing case is from a national dialysis chain center and the other from an independent dialysis center. No dialysis center has more than one record in the study sample. Sample selection uses a random selection of cases across the entire CY2012 population of ESRD patients newly diagnosed in the period. The proportion of missing cases (2/73) is less than 3%.

Sensitivity analysis results for the data elements are displayed in Table 2. For the total element match, the observed match proportion can range from a low of .81 when the region data elements do not match the source for the missing cases (Scenario = As error) to a high of .84 when the region data elements match the source (Scenario = As match). When the region does not match the source, the match proportion is significantly different from .9 at the 95% level (CI, .70-.89). For the numerator match, the match proportion has a low of .84 and a high of .87. For the denominator, the match proportion is .93 when the region does not match the source and .96 when it does. The numerator and denominator extremes are not statistically different from the a priori match proportion, .09.

Performance metric results are in Table 4. The optimal ESRD start metric is .84 (95% CI .73-.92) under the 'As error', both sensitivity and specificity drop to .78 and .87 respectively, the positive likelihood ratio is now 6 with a negative LR increasing to .3. For the 'As match' scenario, the performance metric is .87 (95% CI .77-.94), sensitivity is .83 (95% CI .68-.93), specificity is .93 (.77-.99), positive LR increases slightly to 12.5 and negative LR remains at .2.

[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]

Information From 2015 Submission

Given the relatively small number of cases with missing data, less than 3% (2/73), we expect a small effect to be seen in the sensitivity analysis and because the match proportion is greater than .5 that any differences are greater when missing data are treated as errors than when treated as matches. This is seen the Table 2 and 4 results. The largest change is seen in the positive LR where the observed ratio is 12 compared to 6 in the 'As error' scenario. For all of the other test metrics, the differences are much smaller. Using the simple comparison of confidence intervals, all of the 'As error' and 'As match' confidence intervals overlap within the respective columns indicating that there is no statistical difference between the 'As error' and 'As match' value. The missing data has no statistically significant effect upon the observed 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]

These are the exclusion criteria. The first 3 criteria were described in the 2015 submission:

1. Patients who recover GFR to the point that they are off dialysis by day 90 after the first outpatient dialysis.
2. Patients who are not a KP member at the start of the renal replacement therapy.
3. Patients who are younger than 18 years old.

In 2017, an exclusion criterion related to the 10% hemodialysis patient limit for arteriovenous graft (AVG) that was in place up to 2016 was removed.

[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]

Table 1. Table shows percent by AVG, flag for AVG limit in place, flag for AVG limit reached between 2021 and 2016 AVG rate

Markets	Year	2016 %AVG	2016 - AVG limit of <10% in Optimal Start Measure in place	2016 - AVG 10% Reached?	Year	2021 %AVG	2021 - AVG limit of <10% in Optimal Start Measure in place	2021 - AVG 10% Reached?
Market 1	2016	0.0%	Y	N	2021	1.0%	N	N
Market 2	2016	4.8%	Y	N	2021	4.3%	N	N
Market 3	2016	0.0%	Y	N	2021	6.5%	N	N
Market 4	2016	7.5%	Y	N	2021	5.5%	N	N
Market 5	2016	5.3%	Y	N	2021	7.6%	N	N
Market 6	2016	3.2%	Y	N	2021	0.7%	N	N
Market 7	2016	2.1%	Y	N	2021	3.2%	N	N
Market 8	2016	6.8%	Y	N	2021	3.2%	N	N
KP	2016	4.1%	Y	N	2021	3.5%	N	N

Table 1. Table shows percent by AVG, flag for AVG limit in place, flag for AVG limit reached between 2021 and 2016 AVG rate

[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]

The removal of the criterion of 10% new hemodialysis patient limit for arteriovenous graft (AVG) was made based on recommendations from the KDOQI 2019 Vascular Access guidelines (website: https://www.ajkd.org/article/S0272-6386(19)31137-0/fulltext), which supported use of arteriovenous graft as equally beneficial as arteriovenous fistula for ESKD patients and moved away from the Fistula First policy.

[Response Ends]

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

[Response Begins]

No risk adjustment or stratification

[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]

not applicable

[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]

While this measure of Optimal Start of ESRD does not directly measure the very important work of supportive care with planned no renal replacement therapy, it does strongly support and encourage this work in a compassionate way. It encourages early, careful, and continual life care planning and shared decision-making with the patient on the pathway of supportive care with planned no renal replacement therapy to allow adequate time to prevent a non-optimal start should the patient change course.

[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]

[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]

not applicable

[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]

not applicable

[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]

not applicable

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

not applicable

[Response Ends]

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

[Response Begins]

not applicable

[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]

not applicable

[Response Ends]

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

[Response Begins]

not applicable

[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]

not applicable

[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]

not applicable

[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)

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

[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 electronic health records (EHRs)

[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]

Not applicable

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

no effort at the moment

[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]

Collecting and submitting data: Despite the need to transcribe electronically stored information onto the submission form, determining Optimal ESRD Starts has been feasible within Kaiser Permanente. Since the metric outcomes are viewed by medical group leadership as well as by nephrologists and their teams, there has been motivation to maintain accurate data lists and to verify data before submitting for measure calculation.

When deciding to utilize this measure, it is important to consider the size of the denominator. A minimum of 50 new ESRD patients is required to decrease the impact of random variation in patients needing to start dialysis. At the level of analysis specified (health plans, integrated delivery systems, large nephrologist groups) this should usually not be a problem. In some circumstances. It would

be reasonable to extend the time period for data collection to insure there are a 50 or more denominator patients. Even if a period of 18-24 months were needed, overlapping time periods can allow for more frequent reporting, such as every 6 months.

Operational use: The success of measuring and improving Optimal ESRD Starts has been enhanced by the sponsorship and follow through of leadership. Use of this measure has been demonstrated to improve health outcomes for patient members and to reduce the use of health care resources, goals shared by all health care systems.

Useful Sub-metrics: In addition to providing the percentage of patients with Optimal ESRD Starts, the metric yields the numbers and percentages of patients in each category: preemptive kidney transplant, home peritoneal dialysis, home hemodialysis and successful AVF/AVG placement for in-center hemodialysis. Teams can determine how they are doing in each of these categories in comparison to peer groups, and understand where they have the best opportunities for improvement.

Missing/incomplete data: Missing denominator patients has not been an issue as authorization for dialysis or kidney transplant is required in order to receive payment. The same will be true when 2728 Forms are utilized as the data source, since they are required for Medicare Payment and are a routine part of dialysis admission and preemptive kidney transplant care. The process for any inconsistent or questionable data submitted has been to contact the patient's renal care team for clarification. This process has been shown to be feasible when carried out by a data analyst with proper training and program support. The additional cost of determining Optimal ESRD starts may not be great if it is carried out by a previously established quality program already employing data analysts.

[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]

No outside fees or licensing are required.

[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]

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

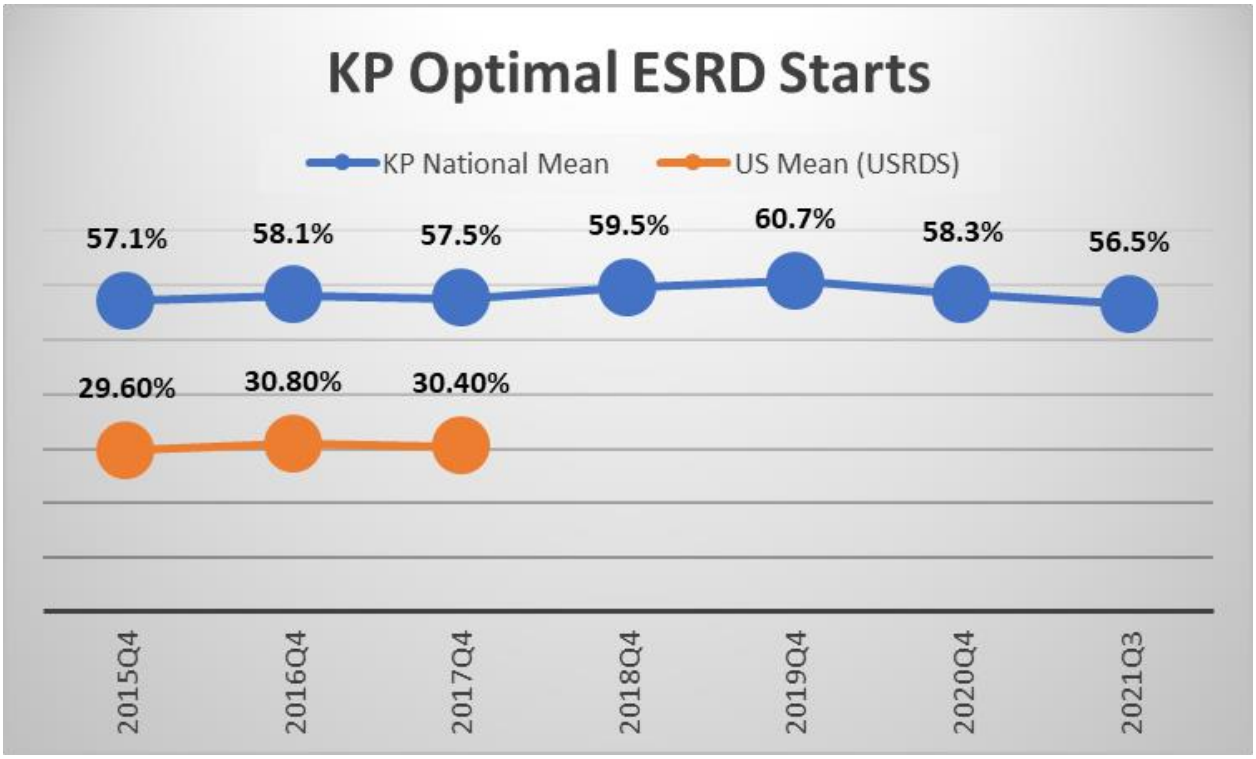
[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]

Name of program and sponsor:

Optimal ESRD Starts is currently utilized by the Permanente Federation (sponsor for this NQF submission) to track the performance of 8 Kaiser Permanente Regions.

This measure is benchmarked against the average US optimal ESRD starts rate calculated using data from the USRDS, see graph below.

Line Graph. A line graph that depicts the annual measure rate since 2015 up to 2021 for Kaiser Permanente and compares the trend line to the US annual rate.



Line Graph. A line graph that depicts the annual measure rate since 2015 up to 2021 for Kaiser Permanente and compares the trend line to the US annual rate.

Geographic area and number and percentage of accountable entities and patients as of 2021 included:

KP Regions participating (with patient counts from last measure) are Northern (1331) and Southern (1806) California,

Northwest/Oregon (162) and Southern Washington State (103), Hawaii (128), Colorado (139), Georgia (132) and Mid-Atlantic States (201).

NOTE: ESRD Optimal Starts is intended to be used to measure and improve performance in health plans, integrated delivery systems, large nephrology provider groups, and broad geographic Regions with large numbers of ESRD patients. It is not specified for individual practitioners, or small provider groups with fewer than 50 new ESRD patients during the measurement period. It is not intended for use by dialysis providers.

Quality Improvement (Internal to the specific organization)

[Quality Improvement (Internal to the specific organization) Please Explain]

Name of program and sponsor:

Optimal ESRD Starts is currently utilized by the Permanente Federation (sponsor for this NQF submission) to track the performance of 8 Kaiser Permanente Regions.

Purpose:

1) Quality improvement by comparing results from the various Kaiser Permanente Regions and sharing these results with the regional nephrologists and renal teams, leading to operational comparisons, physician and health care team education, and appropriate resource allocation.

2) Accountability to the Regional Executive Medical Directors. The measure for all 8 participating KP regions is reported every 3 months to each Executive Medical Director. Executive Medical Directors control medical group resources and hold the physicians and administrative teams accountable for improving Optimal ESRD Starts.

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

KP Regions participating (with patient counts from last measure) are Northern (1300) and Southern (1800) California,

Northwest/Oregon (170) and Southern Washington State (120), Hawaii (130), Colorado (140), Georgia (130) and Mid-Atlantic States (270).

NOTE: ESRD Optimal Starts is intended to be used to measure and improve performance in health plans, integrated delivery systems, large nephrology provider groups, and broad geographic Regions with large numbers of ESRD patients. It is not specified for individual practitioners, or small provider groups with fewer than 50 new ESRD patients during the measurement period. It is not intended for use by dialysis providers.

Not in use

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Quality Improvement (internal to the specific organization)

[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]

Optimal ESKD Starts (OES) metric was created in early 2000's as a composite outcome measure to evaluate success of pre-ESKRD care. For patients with progressive chronic kidney disease (CKD), approaching transition to End Stage Kidney Disease (ESKD), few outcomes are generally considered to be associated with better outcomes or aligned with patient and caregivers' preferences:

- Pre-emptive transplantation, when patient received kidney transplant prior to initiation of dialysis, thus avoiding need for dialysis all together.
- Starting maintenance dialysis with one of the forms of home dialysis, peritoneal dialysis (PD) or home hemodialysis (HHD)
- Starting outpatient in-center HD with mature, ready to use arterio-venous fistula (AVF) or graft (AVG)
- Receiving maximal medical therapy to relieve symptoms related to ESKD/uremia, without dialysis

In its essence, OES composite measure serves as a marker of comprehensive CKD care, which includes identifying patients at risk of progressive CKD early, risk stratifying them to prioritize therapeutic interventions, providing education about CKD nature and ESKD therapy options and optimizing their transition to ESKD or medical care without dialysis.

OES is one of the major outcome goals monitored in Kaiser Permanente in assessing healthcare system performance in care for patients with CKD. After becoming an NQF measure, OES received support from ASN and KCP (1,2) is being looked at or adopted by other integrated healthcare systems (5) and is frequently referenced as a desired outcome by commercial companies working in CKD space as well as part of the proposed Kidney Care Choices (KCC) CMS/CMMI model. (4)

While KP or other entities do not publicly report on OES performance, components of the measure are publicly reported: rate of pre-emptive transplantation, rate of functioning vascular access as well as incidence and prevalence of home dialysis modalities (PD and HHD). In addition, goal of increasing prevalence of home dialysis and kidney transplantation

are central parts of Advancing American Kidney Health (3) presidential executive order, while increasing use of permanent hemodialysis vascular access is a part of standard publicly reportable dialysis units CMS QIP.

Kaiser's integrated care delivery model and sophisticated electronic health records (EHR) system allows consistent data collection and reporting. Lack of consistent and powerful enough EHR as well as existing integrated multidisciplinary CKD-care network in US healthcare community, capable of consistently collecting and reporting on this composite measure, might be the main factor why this measure is not publicly reported at this point. However, this might be a reason to promote it, to ensure other integrated health care systems, ACO's and ESCO's adopt this measure and report on success of CKD and transitional care management.

We understand that this measure might be difficult to roll-out in smaller practices or on the level of individual dialysis clinics. Reported outcomes in measuring period might be too small, and for dialysis clinics, there is no opportunity for intervention in pre-dialysis care.

In summary, we are advocating to measure to be supported by NQF, with expectation that it will be eventually adopted for public reporting by additional health care systems, growing commercial entities and, CMS/CMMI promoted care models such as KCC.

1. Himmelfarb, J. NQF renal project, https://www.asn-online.org/policy/webdocs/15.7.13asncommentsonnqfrenalmeasures.pdf,
2. McConigal, L. NQF Renal Project. https://kidneycarepartners.org/wp-content/uploads/2015/04/ltNQFRenalProjectCommentsLeadRepsSIGNED04-13-15.pdf),
3. https://www.federalregister.gov/documents/2019/07/15/2019-15159/advancing-american-kidney-health
4. https://jasn.asnjournals.org/content/31/3/602
5. https://pubmed.ncbi.nlm.nih.gov/21282303/

NOTE: ESRD Optimal Starts is intended to be used to measure and improve performance in health plans, integrated delivery systems, large nephrology provider groups, and broad geographic Regions with large numbers of ESRD patients. It is not specified for individual practitioners, or small provider groups with fewer than 50 new ESRD patients during the measurement period. It is not intended for use by dialysis providers.

[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]

With support of ASN, we are currently in the process of submitting Optimal ESRD Starts as an optional quality measure to the MIPS. If successful, it will have a path for public reporting for interested stakeholders. In addition, Optimal ESRD Starts is getting adopted by other large professional organizations, such as Nephology Care Alliance as target. ("High value care for chronic kidney disease a frontline solution")

reference

"High value care for chronic kidney disease a frontline solution" Nephrology Care Alliance.

[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]

Communication of performance results is shared widely within the organization and actual performance results are shared quarterly with markets.

[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]

Measure results are shared quarterly with market leaders over email communications and at regular meetings. Results are summarized at the market level and markets can view their rates, denominator, and numerator counts. Trend data is also provided for each market to understand changes over time.

[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]

Since 2017, renal care national leaders held regular Optimal Starts Check In Calls with each market clinical leaders to learn and understand their successes and barriers. Learnings were collected and summarized and then shared with all markets. Results, including comparative reporting, are also shared with mid-level managerial leadership and front line staff, both providers and nursing. Results are reviewed quarterly at clinical population management meetings and again at provider meetings.

[Response Ends]

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

[Response Begins]

A summary of feedback obtained from markets recently included the following:

- Early CKD: successes and barriers
- Transitions and Optimal ESRD Starts: successes and barriers
- Regional activities to increase incidence and support patients to remain on home modalities
- Transplant activities to increase transplant rates
- Dignified Journeys (palliative care and life care planning)
- Program-wide, inter-regional coordination of efforts (contracting, commercial to Medicare benefits, claims management, clinical care and quality)
- Measured performance and implementation are shared with local medical center teams and leadership as well as regional leadership (nephrology and senior administration). Feedback is obtained at local and regional meetings that meet (monthly, quarterly, annually).
- This metric accurately reflects successful CKD to dialysis (or transplant) transitions.

[Response Ends]

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

[Response Begins]

Feedback from American Society of Nephrology Quality Committee in 2020 regarding this measure:

The NQF 2594 optimal ESKD starts measure was the only metric specifically addressing advanced CKD and kidney replacement planning. The measure captures the percentage of patients with a preemptive kidney transplant, home dialysis initiation, or outpatient in-center hemodialysis via AVF or arteriovenous graft. The measure was rated as having high validity. Committee discussion centered on the importance of having an evidence-based target given potential patient

nonadherence, accelerated progression to kidney failure, and lack of timely nephrology referral.Most committeemembers felt that this metric was a step in the right direction, particularly considering the Advancing American Kidney Health Initiative by Health and Human Services and the Centers for Medicare and Medicaid Innovation kidney care models.17 These mandatory and optional models emphasize home dialysis, transplantation, and coordinated outpatient

dialysis initiation, making the optimal ESKD starts metric particularly relevant.

Feedback from ASN on July 13, 2015 regarding this measure:

Ensuring an optimal, individualized start to dialysis care is a crucially important goal for every patient. ASN strongly supports efforts to encourage focus on this goal; perhaps no aspect of the practice of nephrology is in greater need of quality improvement than how patients typically initiate dialysis. In contrast to virtually all other measures in the ESRD Quality Incentive Program (QIP), which operate on the margins, the optimal start measure stands to significantly improve outcomes and quality of life for patients. The nephrology community must do a better job holding itself accountable to start improving patients' dialysis starts, and the society commends Kaiser for developing measure #2594 with this objective in mind. Even among patients followed by a nephrologist for 6 months or more prior to initiation, the rate of starting dialysis with a catheter still exceeds 70%—a dismal trend that shows little sign of change overall. Although dialysis patient mortality in the US is improving, it remains significantly higher than other developed countries; only 50% of the patients on dialysis survive more than 3

years—which is double the average mortality rate for cancer patients. Notably, improvements in dialysis patient mortality can be largely attributed to an increase in AVFs and a decline in catheter rates, and while AVFs are not the appropriate choice for every patient many more people would benefit from receiving this access prior to starting dialysis than do today. Furthermore, experiences in certain areas of the country suggest that change is achievable.

For example, one care facility in the Pacific Northwest serving a safety net population of predominantly uninsured/underinsured/undocumented individuals has reduced the proportion of patients that start dialysis with a catheter to less than 30%, compared to the national average of 70%. Another facility in that region serving socioeconomically disadvantaged patients has reduced this proportion to less than 50%. Selection of access type is just one of numerous aspects involved in the initiation of dialysis that would be improved through greater emphasis on ensuring an optimal start for each patient.

Pre-emptive transplantation is another underutilized and highly effective option that this measure would promote; nationwide, less than three percent of patients receive a pre-emptive transplant. The society fully endorses the principles laid out in measure #2594 and concurs, in concept, that this is a superb area for quality measure development. As the NQF draft report notes, the prior coordination between PCPs and nephrologists necessary to achieve optimal—and

individualized—starts (as defined by this measure) would best be facilitated in an integrated delivery care system. However, feasibility of implementation remains the issue. Indeed, Kaiser clarifies that "the measure is currently utilized in

an integrated care delivery system," and reiterated "difficulty in using the measure in any type of unit with less than 50 patients." This measure is not a dialysis facility level metric but rather a health system level metric—albeit a

metric that, lamentably, very few health systems are equipped to achieve to date. ASN appreciates the concerns raised regarding whether this measure is feasible in the majority of care settings at this time. Care must be taken to ensure that this measure is only used in settings where providers have an opportunity to influence patient preparation for dialysis. Looking beyond this specific measure, ASN suggests that as a starting point for pushing the

community to do better, nephrologists should begin to create greater accountability by identifying some strata of patients under the care of the nephrology care team for some defined length of time to be the denominator in an assessment of optimal starts. Notably, this information is captured by Form 2728 in question 18 b, which stratifies the amount of time patients have been under the care of nephrologists for six months or less, 6-12 months, and >12 months). ASN would consider it reasonable to start with patients under the care of nephrologists for at least six months and exclude patients who begin dialysis during an acute

kidney injury episode. The society would also encourage the Centers for Medicare and Medicaid Services (CMS) to begin rigorously tracking all available data related to optimal starts and collecting information about feasibility in integrated and non-integrated health systems. Collecting and sharing such data would help facilitate adoption of this and similar optimal start measures.

Feedback from Kidney Care Partners on April 13, 2015 regarding this measure:

The percentage of new ESRD patients during the measurement period who experience a planned start of renal replacement therapy by receiving a preemptive kidney transplant, initiating home dialysis, or initiating outpatient in-center hemodialysis via arteriovenous fistula (AVF) or AV graft. Comment: Generally support with noted qualifications, but oppose for implementation in dialysis facilities. KCP acknowledges that this team/population/integrated-system level measure is an important and conceptually appropriate measure, but believes it is only feasible in fully integrated delivery care systems or large physician groups. As constructed, the measure cannot be applied to dialysis facilities, or even ESCOs, because neither includes CKD patients. While the "clinician/team" level specified in the measure (i.e., the nephrology practice) could be applied, sufficient patient volumes will be an issue. Specifically, the number of patients starting ESRD care within a practice within the defined 12-month period of time might not be sufficient to provide statistically valid data. Likewise, the measure doesn't account for the fact that approximately 40% of patients have not yet been seen the by the nephrologist they will be assigned to at the time of dialysis initiation.

Additionally, we recommend that the developer consider three exclusions addressing scenarios in which a permanent access may not be appropriate. These

three groups represent a small number of patients cumulatively, but they should nevertheless be appropriately be excluded:

- Patients with a limited life expectancy, where not placing a permanent access may be more consistent with the patient's goals;
- Patients who are uncertain whether they wish to pursue long-term treatment and desire a time-limited trial of dialysis; and
- Patients with Acute Kidney Injury without prior CKD who ultimately don't recover renal function; these patients will likely use a catheter for their first treatment, and it is appropriate to wait up to four months to see if they will recover function before pursuing permanent access.

Finally, we have identified three potential unintended consequences that should be closely monitored:

• There could be some "misuse" or misrepresentation of an "optimal start". For instance, patients who have not received appropriate pre-ESRD care

might be persuaded to acutely start peritoneal dialysis in the hospital, and then be kept on this course so as to improve the measured entity's

performance.

• The measure excludes patients with a single needle in the AVF that have blood return via their catheter. This could incentivize premature insertion of

both needles for the first treatment, leading to damage of immature AVFs. 6 Catheters could then be used in subsequent treatments, without impacting

measure performance.

• The measure will likely penalize providers caring for a disproportionate percentage of patients with low socioeconomic status, since these individuals

most often have not seen a nephrologist prior to beginning treatment; safety net providers may be inappropriately penalized.

[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]

Feedback from American Society of Nephrology, ASN and Kidney Care Partners supported continued implementation of this measure at Kaiser Permanente since it significantly improves outcomes and quality of life for patients. Being a composite measure that includes outpatient in-center hemodialysis via AVF or arteriovenous graft, the renal care committee at Kaiser Permanente also reviewed clinical guidelines from KDOQI to guide development of this measure. In 2018, the KDOQI clinical guideline for vascular access recommended moving away from a Fistula First approach. As a result of this, we removed from the measure specification the limitation of 10% placed on patients with arteriovenous graft. Additionally, we also recognized that many patients with failed kidney transplants would return back to dialysis, therefore, we modified the measure specification to include this group of patients.

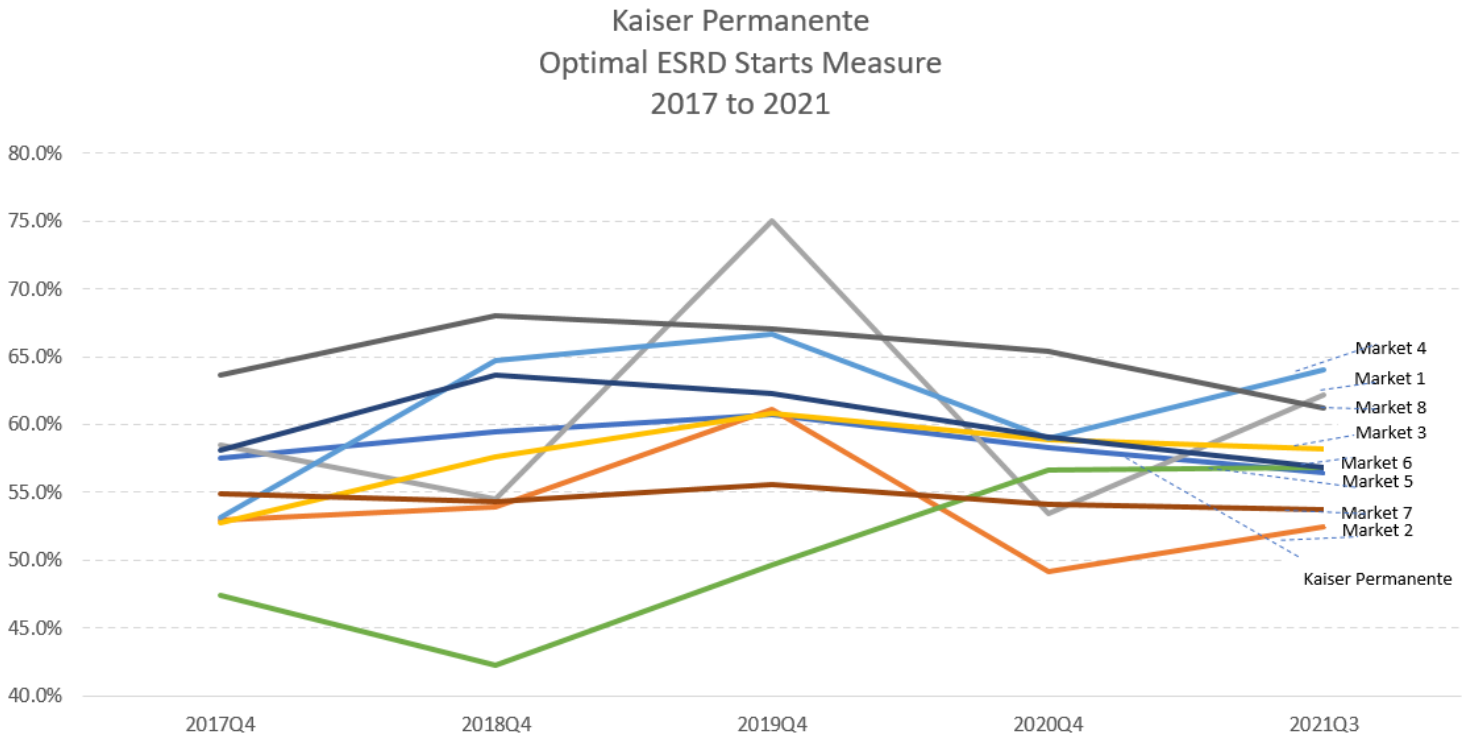
[Response Ends]

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]

Below is a trend graph of the Optimal ESRD Starts measure for all 8 KP markets from 2017 to 2021. The number of new ESKD patients increased from 3,787 in 2017 to 4,002 in 2021. In 2017, the Optimal ESRD Starts measure rate was at 57.5%, and continued to climb to 60.7% in 2019 before trending down to 56.5% in 2021. This downward trend in 2020 through 2021 is expected due to the impact of the global COVID19 pandemic.



Line Graph. A line graph that depicts the annual age sex race adjusted measure rate since 2017 up to 2021 for Kaiser Permanente and individual market rates.

[Response Ends]

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

[Response Begins]

Unintended negative consequences of measuring Optimal ESRD Starts were not identified during implementation.

There can be complications from creating an arteriovenous fistula or graft, such as surgical complications or heart failure from circulatory overload. Such complications would not be the result of using the measure.

A hypothetical negative consequence could be an increase in the number of arteriovenous fistulas created for hemodialysis that are never used due to patient death, recovery or stabilization of kidney function, or a later decision to forego dialysis treatment. To the extent to which that occurs, this should be outweighed by an increase in the number of patients prepared for starting hemodialysis who will not require a dialysis catheter with its high risks for infection and early mortality.

Another possible negative consequence could be this - by increasing home dialysis rates, the number of patients starting hemodialysis with CVC could go up since some patients who were planning on starting on home dialysis start on hemodialysis for multiple reasons such as acute kidney injury. Our data shows that despite increasing the overall use of home dialysis at Kaiser Permanente, the number of patients starting hemodialysis with CVC did not go up.

On the positive side, Kaiser Permanente focuses on doing what is best for the patient, which also aligns with KDOQI's Patient Life-Plan first. Therefore, we recognize that the best choice of treatment for certain ESKD patients is hemodialysis catheter because it is the treatment that will have the least adverse effects and complications. At the same time, we also recognize that some ESKD patients may need to have a "no dialysis" option.

[Response Ends]

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

[Response Begins]

There are several unexpected benefits gained from implementation of this measure and they are listed as follows:

1. moving upstream with CKD Management to improve outcomes of Optimal ESRD Starts
2. better integration of care w/ primary providers, and improved electronic tools that support this effort
3. development of alternate options such as embedded and interventional radiology placed PD catheter
4. improvement in quantifying the value of care provided to CKD/ESKD patients
5. Establishing a better relationship and processes with other services as part of a multi-disciplinary care approach to care for CKD/ESKD patients
6. Enhancing earlier patient activation and participation
7. Implementing electronic tools to assist with screening and risk stratification for patients with CKD

[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]

[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]

There are no competing measures to the Optimal ESRD Starts measure.

[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]

There are two related measures, 0256 and 0257, but no competing measures. These measures and Optimal ESRD Starts are complementary with different rationale and different data collection methods. Optimal ESRD Starts focuses on patients who need to start renal replacement therapy, including hemodialysis, whereas measures 0256 and 0257 both focus on improving vascular access for patients already on hemodialysis. The Measure 0256 Hemodialysis Vascular Access – Minimizing use of catheters as Chronic Dialysis Access metric is a percentage of patients currently on maintenance hemodialysis with a chronic catheter in place continuously for 90 days or more. As opposed to Optimal ESRD Starts, which is an incidence rate for new ESRD patients, measure 0256 is a prevalence measure of the existing hemodialysis population. Another difference is that even a single first treatment with a catheter is a negative Optimal ESRD Start outcome, whereas measure 0256 requires a catheter to be present for 90 days or longer. While the denominator populations are not harmonized, Optimal ESRD Starts is complimentary as more Optimal ESRD Start without a hemodialysis catheter will lower chronic catheter prevalence. The Measure 0257 Hemodialysis Vascular Access – Maximizing Placement of Arterial Venous Fistula metric is a percentage of patients on maintenance hemodialysis using an autogenous arteriovenous fistula (AVF). Like optimal ESRD Starts, it focuses on increasing the use of arteriovenous fistulas as the best type of vascular access for hemodialysis. As opposed to Optimal ESRD Starts, which is an incidence rate for new ESRD patients, measure 0257 is a prevalence measure of the existing hemodialysis population. While the denominator populations are not harmonized, Optimal ESRD Starts is complimentary. An Optimal ESRD Start with an AVF will result in higher AVF prevalence. In summary, Optimal ESRD starts is quite different in focus (Pre-ESRD patient planning versus managing patients already on hemodialysis), covers home dialysis and transplant as well as inpatient hemodialysis, and is the only metric to impact patients before and as they transition to ESRD. It is an incidence rate at the point of reaching ESRD as opposed to a prevalence rate in patients already on hemodialysis. Optimal ESRD Starts tells how a health care entity is performing in the build up to ESRD to optimize each patient's modality choice, and the other two measures address how an organization is doing after patients reach ESRD, limited only to hemodialysis.

[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]

N/A

[Response Ends]