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 #: 3689

Measure Title: First Year Standardized Waitlist Ratio (FYSWR)

Measure Steward: Centers for Medicare & Medicaid Services

Brief Description of Measure: The FYSWR measure tracks the number of incident patients in a practitioner (inclusive of physicians and advanced practice providers) group who are under the age of 75 and were listed on the kidney or kidney-pancreas transplant waitlist or received a living donor transplant within the first year of initiating dialysis. For each practitioner group, the First Year Standardized Waitlist Ratio (FYSWR) is calculated to compare the observed number of waitlist events in a practitioner group to its expected number of waitlist events. The FYSWR uses the expected waitlist events calculated from a Cox model, adjusted for age and patient comorbidities at incidence of dialysis. For this measure, patients are assigned to the practitioner group based on the National Provider Identifier (NPI)/Unique Physician Identifier Number (UPIN) information entered on the CMS Medical Evidence 2728 form.

Developer Rationale: A measure focusing on the outcome of waitlisting is appropriate for several reasons. First, in preparing patients for suitability for waitlisting, dialysis practitioners optimize their health and functional status, improving their overall health state. Second, waitlisting is a necessary step prior to potential receipt of a deceased donor kidney transplant (receipt of a living donor kidney is also accounted for in the measure), which is known to be beneficial for survival and quality of life [1]. Third, dialysis practitioners exert substantial control over the processes that result in waitlisting. This includes proper education of dialysis patients on the option for transplant, referral of appropriate patients to a transplant center for evaluation, and assisting patients with completion of the transplant evaluation process, in order to increase their candidacy for transplant waitlisting. These types of activities are included as part of the conditions for coverage for Medicare certification of ESRD dialysis facilities. Finally, wide regional and facility variations in waitlisting rates highlight substantial room for improvement for this measure [2-5].

Additionally, this measure focuses specifically on the population of patients incident to dialysis, examining for waitlist or living donor transplant events occurring within a year of dialysis initiation. This will evaluate and encourage rapid attention from dialysis practitioner groups to the optimization of health of patients to ensure early access to the waitlist, which has been demonstrated to be particularly beneficial [6-9]. This measure contrasts with the other proposed waitlisting measures, which focus on a prevalent population of dialysis patients and encourage maintenance of patients on the waitlist (Percent of Prevalent Patients Waitlisted and Percent of Prevalent Patients Waitlisted in Active Status).

1. *Tonelli M, Wiebe N, Knoll G, et al. Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. American Journal of Transplantation 2011;11:2093-2109.*

Abstract: Individual studies indicate that kidney transplantation is associated with lower mortality and improved quality of life compared with chronic dialysis treatment. We did a systematic review to summarize

the benefits of transplantation, aiming to identify characteristics associated with especially large or small relative benefit. Results were not pooled because of expected diversity inherent to observational studies. Risk of bias was assessed using the Downs and Black checklist and items related to time-to-event analysis techniques. MEDLINE and EMBASE were searched up to February 2010. Cohort studies comparing adult chronic dialysis patients with kidney transplantation recipients for clinical outcomes were selected. We identified 110 eligible studies with a total of 1 922 300 participants. Most studies found significantly lower mortality associated with transplantation, and the relative magnitude of the benefit seemed to increase over time ($p < 0.001$). Most studies also found that the risk of cardiovascular events was significantly reduced among transplant recipients. Quality of life was significantly and substantially better among transplant recipients. Despite increases in the age and comorbidity of contemporary transplant recipients, the relative benefits of transplantation seem to be increasing over time. These findings validate current attempts to increase the number of people worldwide that benefit from kidney transplantation.

2. *Ashby VB, Kalbfleisch JD, Wolfe RA, et al. Geographic variability in access to primary kidney transplantation in the United States, 1996-2005. American Journal of Transplantation 2007; 7 (5 Part 2):1412-1423.*

Abstract: This article focuses on geographic variability in patient access to kidney transplantation in the United States. It examines geographic differences and trends in access rates to kidney transplantation, in the component rates of wait-listing, and of living and deceased donor transplantation. Using data from Centers for Medicare and Medicaid Services and the Organ Procurement and Transplantation Network/Scientific Registry of Transplant Recipients, we studied 700,000+ patients under 75, who began chronic dialysis treatment, received their first living donor kidney transplant, or were placed on the waiting list pre-emptively. Relative rates of wait-listing and transplantation by State were calculated using Cox regression models, adjusted for patient demographics. There were geographic differences in access to the kidney waiting list and to a kidney transplant. Adjusted wait-list rates ranged from 37% lower to 64% higher than the national average. The living donor rate ranged from 57% lower to 166% higher, while the deceased donor transplant rate ranged from 60% lower to 150% higher than the national average. In general, States with higher wait-listing rates tended to have lower transplantation rates and States with lower wait-listing rates had higher transplant rates. Six States demonstrated both high wait-listing and deceased donor transplantation rates while six others, plus D.C. and Puerto Rico, were below the national average for both parameters.

3. *Satayathum S, Pisoni RL, McCullough KP, et al. Kidney transplantation and wait-listing rates from the international Dialysis Outcomes and Practice Patterns Study (DOPPS). Kidney Intl 2005 Jul; 68 (1):330-337.*

Abstract: BACKGROUND: The international Dialysis Outcomes and Practice Patterns Study (DOPPS I and II) allows description of variations in kidney transplantation and wait-listing from nationally representative samples of 18- to 65-year-old hemodialysis patients. The present study examines the health status and socioeconomic characteristics of United States patients, the role of for-profit versus not-for-profit status of dialysis facilities, and the likelihood of transplant wait-listing and transplantation rates.

METHODS: Analyses of transplantation rates were based on 5267 randomly selected DOPPS I patients in dialysis units in the United States, Europe, and Japan who received chronic hemodialysis therapy for at least 90 days in 2000. Left-truncated Cox regression was used to assess time to kidney transplantation. Logistic regression determined the odds of being transplant wait-listed for a cross-section of 1323 hemodialysis patients in the United States in 2000. Furthermore, kidney transplant wait-listing was determined in 12 countries from cross-sectional samples of DOPPS II hemodialysis patients in 2002 to 2003 (N= 4274).

RESULTS: Transplantation rates varied widely, from very low in Japan to 25-fold higher in the United States and 75-fold higher in Spain (both P values <0.0001). Factors associated with higher rates of transplantation included younger age, nonblack race, less comorbidity, fewer years on dialysis, higher income, and higher education levels. The likelihood of being wait-listed showed wide variation internationally and by United States region but not by for-profit dialysis unit status within the United States.

CONCLUSION: DOPPS I and II confirmed large variations in kidney transplantation rates by country, even after adjusting for differences in case mix. Facility size and, in the United States, profit status, were not associated

with varying transplantation rates. International results consistently showed higher transplantation rates for younger, healthier, better-educated, and higher income patients.

4. *Patzer RE, Plantinga L, Krisher J, Pastan SO. Dialysis facility and network factors associated with low kidney transplantation rates among United States dialysis facilities. Am J Transplant. 2014 Jul; 14(7):1562-72.*

Abstract: Variability in transplant rates between different dialysis units has been noted, yet little is known about facility-level factors associated with low standardized transplant ratios (STRs) across the United States End-stage Renal Disease (ESRD) Network regions. We analyzed Centers for Medicare & Medicaid Services Dialysis Facility Report data from 2007 to 2010 to examine facility-level factors associated with low STRs using multivariable mixed models. Among 4098 dialysis facilities treating 305 698 patients, there was wide variability in facility-level STRs across the 18 ESRD Networks. Four-year average STRs ranged from 0.69 (95% confidence interval [CI]: 0.64-0.73) in Network 6 (Southeastern Kidney Council) to 1.61 (95% CI: 1.47-1.76) in Network 1 (New England). Factors significantly associated with a lower STR (p<0.0001) included for-profit status, facilities with higher percentage black patients, patients with no health insurance and patients with diabetes. A greater number of facility staff, more transplant centers per 10,000 ESRD patients and a higher percentage of patients who were employed or utilized peritoneal dialysis were associated with higher STRs. The lowest performing dialysis facilities were in the Southeastern United States. Understanding the modifiable facility-level factors associated with low transplant rates may inform interventions to improve access to transplantation.

5. *Melanson TA, Gander JC, Rossi A, et al. Variation in Waitlisting Rates at the Dialysis Facility Level in the Context of Goals for Improving Kidney Health in the United States. Kidney International Reports 2021;6:1965-1968.*

No abstract.

6. *Meier-Kriesche, Herwig-Ulf, and Bruce Kaplan. "Waiting time on dialysis as the strongest modifiable risk factor for renal transplant outcomes: A Paired Donor Kidney Analysis." Transplantation 74.10 (2002): 1377-1381.*

Abstract: BACKGROUND: Waiting time on dialysis has been shown to be associated with worse outcomes after living and cadaveric transplantation. To validate and quantify end-stage renal disease (ESRD) time as an independent risk factor for kidney transplantation, we compared the outcome of paired donor kidneys, destined to patients who had ESRD more than 2 years compared to patients who had ESRD less than 6 months.

METHODS: We analyzed data available from the U.S. Renal Data System database between 1988 and 1998 by Kaplan-Meier estimates and Cox proportional hazards models to quantify the effect of ESRD time on paired cadaveric kidneys and on all cadaveric kidneys compared to living-donated kidneys.

RESULTS: Five- and 10-year unadjusted graft survival rates were significantly worse in paired kidney recipients who had undergone more than 24 months of dialysis (58% and 29%, respectively) compared to paired kidney recipients who had undergone less than 6 months of dialysis (78% and 63%, respectively; P<0.001 each). Ten-year overall adjusted graft survival for cadaveric transplants was 69% for preemptive transplants versus 39% for transplants after 24 months on dialysis. For living transplants, 10-year overall adjusted graft survival was 75% for preemptive transplants versus 49% for transplants after 24 month on dialysis.

CONCLUSIONS: ESRD time is arguably the strongest independent modifiable risk factor for renal transplant outcomes. Part of the advantage of living-donor versus cadaveric-donor transplantation may be explained by waiting time. This effect is dominant enough that a cadaveric renal transplant recipient with an ESRD time less than 6 months has the equivalent graft survival of living donor transplant recipients who wait on dialysis for more than 2 years.

7. *Meier-Kriesche, H. U., Port, F. K., Ojo, A. O., Rudich, S. M., Hanson, J. A., Cibrik, D. M., ... & Kaplan, B. (2000). Effect of waiting time on renal transplant outcome. Kidney international, 58(3), 1311-1317.*

Abstract: BACKGROUND: Numerous factors are known to impact on patient survival after renal transplantation. Recent studies have confirmed a survival advantage for renal transplant patients over those

waiting on dialysis. We aimed to investigate the hypothesis that longer waiting times are more deleterious than shorter waiting times, that is, to detect a "dose effect" for waiting time.

METHODS: We analyzed 73,103 primary adult renal transplants registered at the United States Renal Data System Registry from 1988 to 1997 for the primary endpoints of death with functioning graft and death-censored graft failure by Cox proportional hazard models. All models were corrected for donor and recipient demographics and other factors known to affect outcome after kidney transplantation.

RESULTS: A longer waiting time on dialysis is a significant risk factor for death-censored graft survival and patient death with functioning graft after renal transplantation ($P < 0.001$ each). Relative to preemptive transplants, waiting times of 6 to 12 months, 12 to 24 months, 24 to 36, 36 to 48, and over 48 months confer a 21, 28, 41, 53, and 72% increase in mortality risk after transplantation, respectively. Relative to preemptive transplants, waiting times of 0 to 6 months, 6 to 12 months, 12 to 24 months, and over 24 months confer a 17, 37, 55, and 68% increase in risk for death-censored graft loss after transplantation, respectively.

CONCLUSIONS: Longer waiting times on dialysis negatively impact on post-transplant graft and patient survival. These data strongly support the hypothesis that patients who reach end-stage renal disease should receive a renal transplant as early as possible in order to enhance their chances of long-term survival.

8. *Schold JD, Huml AM, Poggio ED et al. Patients with High Priority for Kidney Transplant Who Are Not Given Expedited Placement on the Transplant Waiting List Represent Lost Opportunities. J Am Soc Nephrol 2021;32:1733-1746.*

Abstract: BACKGROUND: Kidney transplantation is associated with the best outcomes for most patients with ESKD. The national Kidney Allocation System prioritizes patients with Estimated Post-Transplant Survival (EPTS) scores in the top 20% for expedited access to optimal deceased donor kidneys.

METHODS: We studied adults aged 18 years in the United States Renal Data System with top 20% EPTS scores who had been preemptively waitlisted or initiated dialysis in 2015–2017. We evaluated time to waitlist placement, transplantation, and mortality with unadjusted and multivariable survival models.

RESULTS: Of 42,445 patients with top 20% EPTS scores (mean age, 38.0 years; 57% male; 59% White patients, and 31% Black patients), 7922 were preemptively waitlisted. Among 34,523 patients initiating dialysis, the 3-year cumulative waitlist placement incidence was 37%. Numerous factors independently associated with waitlisting included race, income, and having noncommercial insurance. For example, waitlisting was less likely for Black versus White patients, and for patients in the lowest-income neighborhoods versus those in the highest-income neighborhoods. Among patients initiating dialysis, 61% lost their top 20% EPTS status within 30 months versus 18% of patients who were preemptively listed. The 3-year incidence of deceased and living donor transplantation was 5% and 6%, respectively, for patients who initiated dialysis and 26% and 44%, respectively, for patients who were preemptively listed.

CONCLUSIONS: Many patients with ESKD qualifying with top 20% EPTS status are not placed on the transplant waiting list in a timely manner, with significant variation on the basis of demographic and social factors. Patients who are preemptively listed are more likely to receive benefits of top 20% EPTS status. Efforts to expedite care for qualifying candidates are needed, and automated transplant referral for patients with the best prognoses should be considered.

9. *Schold J and Meier-Kreische HU. Which Renal Transplant Candidates Should Accept Marginal Kidneys in Exchange for a Shorter Waiting Time on Dialysis? Clin J Am Soc Nephrol 2006;1:532-538.*

Abstract: Renal transplantation has been established as a life-saving procedure for patients with ESRD. Deceased donor kidneys convey variable life expectancies for recipients. However, limited information is available to guide patients and patient advocates concerning the appropriateness to list for expanded criteria donations (ECD). Half-lives for wait-listed transplant candidates were estimated from the time of ESRD onset on the basis of recipient age, primary diagnosis, and organ quality using survival models. In addition, we evaluated the likelihood of candidates' receiving a transplant on the basis of age and other characteristics by duration of waiting time. Older patients (65) had longer life expectancy when they accepted an ECD within 2 yr of ESRD onset (5.6 yr) compared with waiting for a standard kidney (5.3 yr) or a living donation (5.5 yr) after 4

yr of dialysis. Conversely, younger recipients (18 to 39 yr) had longer life expectancy with a living donation (27.6 yr) or standard kidney (26.4 yr) after 4 yr on dialysis compared with an ECD after 2 yr of dialysis (17.6 yr). Increased candidate age was associated with the likelihood of not receiving a transplant during the period on the waiting list as a result of mortality and separately related to morbidity and delisting. Older and frailer transplant candidates benefit from accepting lower quality organs early after ESRD, whereas younger and healthier patients benefit from receiving higher quality organs even with longer dialysis exposure. These findings are important for transplant candidates and advocates decision-making and for potential further implementation in allocation policy.

Numerator Statement: Number of patients in the practitioner group listed on the kidney or kidney-pancreas transplant waitlist or who received living donor transplants within the first year following initiation of dialysis.

Denominator Statement: The denominator for the FYSWR is the expected number of waitlist or living donor transplant events in the practitioner group according to each patient's treatment history for patients within the first year following initiation of dialysis, adjusted for age, incident comorbidities, dual Medicare-Medicaid eligibility, Area Deprivation Index (from patient's residence zip code) and transplant center characteristics, among patients under 75 years of age who were not already waitlisted and did not have kidney transplantation prior to the initiation of ESRD dialysis.

Denominator Exclusions: Patients who were at age 75 or older on their initiation of dialysis date are excluded. Patients who were admitted to a skilled nursing home facility (SNF) or a hospice during the month of evaluation were excluded. These exclusions represent conditions for which transplant waitlist candidacy is highly unlikely, and which can be identified readily with available data. Patients were also excluded if waitlisted or transplanted prior to initiation of first dialysis. Patients who were attributed to dialysis practitioner groups with fewer than 11 patients or 2 expected events are not excluded from the measure. All patients who meet the denominator inclusion criteria are included and used to model a given dialysis practitioner group's expected waitlist rate. If a dialysis practitioner group has fewer than 11 patients or 2 expected events, then the dialysis practitioner group is excluded from reporting outcomes.

Measure Type: Outcome

Data Source: Claims, Registry Data

Level of Analysis: Clinician: Group/Practice

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a ***health outcome*** measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a new outcome measure at the clinician group/practice level that tracks the number of incident dialysis patients in a practitioner (inclusive of physicians and advanced practice providers) group who

are under the age of 75 and were listed on the kidney or kidney-pancreas transplant waitlist or received a living donor transplant within the first year of initiating dialysis.

- The developer provides a logic model that depicts a link between when a patient with end-stage renal disease (ESRD) initiates dialysis and then subsequently receives education, referral to a transplant center, and assistance from a dialysis practitioner to complete the transplant evaluation process and optimize their health as well as functional status they are then likely to receive a transplant or be added to the wait list with the potential to receive a transplant.

Summary:

- The developer noted that according to two technical expert panels (TEP) that were convened to discuss measures that improve access to kidney transplantation, there is broad support for the importance of waitlisting and further, that a vote demonstrated that a majority of the TEP members were in favor of the development of measures that targeted waitlisting.
    - The TEP was comprised of dialysis nephrologists, transplant nephrologists, transplant surgeons, social workers, researchers, and patient representatives with a history of end-stage kidney disease
- In addition, the developers also noted empirical support for the value of waitlisting to patients which came from a study published in the American Journal of Transplantation.
    - The participants of the study were primarily patients with advanced chronic kidney disease prior to transplant and those who had transplants.
    - They were asked about their priorities in choice of transplant center. They stated that they were most likely to rank waitlisting characteristics as the most important feature.
- Further, the developer cited several studies that provide strong support for the association between processes under dialysis practitioner control and waitlisting.
    - In the first study at a dialysis facility in Georgia, the authors conducted a correlation analysis between ranking on referral ratios and waitlist rates and found that the correlation was statistically significant.
    - The second study which used national registry data to investigate the association between whether patients were informed about kidney transplantation and access to transplantation found that around 30% of patients were uninformed about kidney transplantation which was associated with the rate of access to transplantation.
    - A similar study noted that patients who reported receiving transplantation information were associated with a three-fold increase in likelihood of waitlisting.
    - The last study that the developers looked at examined transplant education practices. The study found that facilities that used greater than three education practices had 36% higher waitlist rates than facilities that used less education practices.

Question for the Committee:

- *Is a relationship demonstrated between waitlisting and at least one healthcare structure, process, intervention, or service?*

Guidance from the Evidence Algorithm

Outcome measure that assesses performance on a health outcome (Box 1) à the relationship between the measured health outcome and at least one health action is demonstrated by empirical data (Box 2) à Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

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

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

- The developer presented an analysis of descriptive statistics for the first-year standardized waitlist ratio (FYSWR).
    - There were 281,479 patients and 2,168 practitioner groups that had at least 11 patients and at least two expected events included in the analysis.
    - The analysis demonstrated that the mean value of FYSWR was 1.01 and the interquartile range was 0.77. They further stated that the bottom quartile of practitioner groups had 46 percent lower waitlisting or living donor transplant rates among new dialysis patients during their first year of dialysis than the national average. The developer also stated that the top quartile of practitioner groups had 33 percent higher waitlisting or living donor transplant rates among new dialysis patients during their first year of dialysis than the national average. This data suggests a performance gap exists.

Disparities

- The developer presented the FYSWR by race, ethnicity, and sex for the sample used for performance gap.
    - Mean FYSWR was highest for the categories Other (2.88) and Asian Pacific Islander (2.04) and lowest for Black (1.05).
    - Black (1.05) and White (1.13) had similar FYSWRs compared to the mean across the entire sample (1.01).
    - Non-Hispanics (1.09) had a lower FYSWR than Hispanics (1.48).
    - Males (1.12) had a higher FYSWR than females (0.87).
- The developer stated that the data demonstrated wide variation and performance gaps between different race, ethnicity, and sex categories.

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

- Evidence presented supports transplant. Geographic variability is evident between dialysis facilities. Education by nephrologists our facility staff increased waitlisting. Measure is supported by evidence although most of the studies on waitlisting look at facilities.
- The evidence supports the link between timely waitlisting to achieve transplant and positive outcomes.
- Evidence includes, but not limited to, clinical benefit of transplantation earlier during the course of ESRD and impact of delayed waitlisting on prioritization.
- Evidence is moderately strong, though waitlist and death on waitlist may differ amongst transplant centers.

- Adequate data presented showing association between dialysis clinical processes/factors and waitlisting and transplant.
- The evidence presented provides a rationale for waitlisting as an outcome, but is tangential to the intermediate outcome being measured. Specifically, the measure is intended to assess practitioner/group performance, but the measure and supporting evidence fail to acknowledge that waitlisting per se is a decision made by the transplant center and is beyond the locus of control of the providers targeted in these measures. While a referral to a transplant center and initiation or even completion of the waitlist evaluation process might be appropriate performance measures at this level of analysis, these newly proposed practitioner/group level waitlist measures are not. The transplant center decides whether a patient is placed on a waitlist, not the facility, practitioner, or group practice. There are many potential obstacles and delays in the evaluation process with multiple parties that have nothing to do with the practitioner or group. Penalizing a practitioner/group practice each month for these or other delays is inappropriate; transplant measures with an appropriate sphere of control should instead be pursued.
- I am not aware of other studies.
- High
- Renal transplantation has been shown to improve both the quality of life and patient mortality. Although organ availability is the major rate-limiting factor of transplantation, encouraging expansion of the transplant waitlist may help reduce the waitlist demographic disparities reported for some geographic areas. However, as constructed, the evidence supporting this measure appears to be seriously flawed and, as such, the measures is not a valid indicator of dialysis provider quality: 1. The developer states that "in preparing patients for suitability for wait listing, dialysis practitioners optimize their health and functional status, improving their overall state. No evidence is provided by the developer to support the notion that dialysis providers preferentially work to optimize the health, wellness, and functional status of patients who are suitable for transplantation as compared with any, and every, other patient in the dialysis facility. 2.The developer states that “dialysis practitioners exert substantial control over the processes that result in wait listing. This includes proper education of dialysis patients on the option for transplant, referral of appropriate patients to a transplant center for evaluation and assisting patients with the completion of the transplant evaluation process, in order to increase their candidacy for transplant wait listing." (Bold added) The developer has not presented any evidence to support the statement that” dialysis practitioners exert substantial control over the processes that result in wait listing.” Wait listing is a decision made solely by the independently operated, hospital-based, transplant center. The decision to wait list is not within the control of the providers listed in this measure. The developer has not provided any evidence that successful, or unsuccessful, wait listing can be attributed to an individual nephrologist or practice group, and the developer presents no evidence to support this construct, as presented. The wait list criteria of each transplant center are determined by that individual transplant center. They are not published, they are not standardized, they are not regulated, and they are not controlled by any regulatory agency, nor by any organized structure. The transplant center, and only the transplant center, determines whether a patient shall be placed on a transplant list. Thus, there is a fundamental flaw in the structure and evidence provided for this measure. 3. The developer states that the role of the dialysis practitioner “…Includes proper education of dialysis patients on the option of transplant [and] referral of appropriate patients…” It is correct to state that the nephrologist refers appropriate patients for transplantation evaluation by the independent, separate, transplantation group. However, referral and ultimate wait listing are distinct and separate activities 4. The numerator statement of the measure as developed is “the adjusted count of patient months in which the patient at the dialysis practitioners group practice is on the kidney or kidney-pancreas transplant wait list as of the last day of each month during reporting year. The denominator statement "includes all patient months for the patients who are under the age of 75 in the reporting month…” The denominator exclusions for

referral to a transplant center for evaluation for wait listing are limited to: Patients who are at age 75 or older in the reporting month; patients were admitted to a skilled nursing facility during the month of evaluation; patients were admitted to a skilled nursing facility within 1 year of dialysis initiation; patients determined to be in hospice were excluded from month of evaluation and the remainder of the reporting period; patients with dementia. Thus, as constructed, the developer has determined that with the few exceptions noted above all other dialysis patients under the age of 75 are appropriate for referral to a transplant center (by the nephrologist) and, for the nephrologist to be compliant with the measure, the patient will need to be WAITLISTED for transplantation, (BY THE TRANSPLANT CENTER). Thus, patients with prior CVAs, neurologic disorders (excluding dementia) extensive vascular disease including amputations, severe cardiac disease, frailty, etc., who do not wish to seek transplantation will require referral by the nephrologist for evaluation for waitlisting. As designed, the measure has the potential to interfere with patients' ability to make their own decisions regarding transplantation and transplantation evaluation and waitlisting.

- Measure of first year dialysis patients waitlist ratio for transplant. Evidence to support that patients do better with decreased morbidity and mortality when they have had less time on dialysis.
- Evidence is compelling.
- Health outcome provides empirical data to demonstrate a relationship between the outcome and at least one healthcare structure, process intervention or service. New outcome measure at the clinician group/practice level that tracks that number of incident dialysis patients in a practitioner group who are under the age of 75 listed for a kidney or kidney-pancreas transplant. Two TEP panels convened to discuss measure to improve access to kidney transplantation.
- Pass
- True
- High - - it is clearly desirable to incent more patients to be wait-listed in 1st year kidney replacement treatment
- The empirical data provided is sufficient for this outcome measure.
- The data support that transplant is associated with improved outcomes.
- The developer describes a new measure and provides a logic model that shows a link between when a patient with ESRD initiates dialysis and is referred for transplant/ waitlisted and therefore likely to receive transplant. There is clearly a link between referral/wait list and the intervention of transplant as numerous studies cited have shown. The measures passes on evidence.

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

- Deciles range from 0.13 to 2.31. Disparities are evident (race, ethnicity, sex)
- The evidence shows opportunity for improvement. There is variation in bottom quartile and top quartile of practitioner groups. Disparities data also demonstrates variability further supporting the need for measurement to gauge improvement.
- 2,168 practice groups and 281,479 patients included (2016 – 2019). FYSWR performance: median 0.92 (lower quartile 0.56, upper quartile 1.33). Differences in performance based on race, ethnicity, and sex.
- There appears to be a gap, but it is not clear whether some of the difference is due to transplant center factors.
- Moderate performance gap data presented with data linked to subgroup disparities related to race, ethnicity, sex.
- Gaps in both provider performance and between racial and ethnic groups is presented; however, there is no evidence provided to support that performance on the measure is not more significantly linked to transplant center practices than to those of the treating practitioner/group.

- I'm not sure all of the data was provided.
- Moderate
- Moderate wide variation in ultimate transplant listing by transplantation centers.
- There are certain areas and certain populations who are not referred to transplant or waitlisted in a timely fashion. This warrants a national measure to standardize the process and ensure that all patients have the access/option to work towards kidney transplant.
- Compelling evidence of gap.
- Demonstrating quality problems and opportunity for improvement. Disparities presented the FYSWR by race, ethnicity and sex for sample use for performance gap. Asian Pacific Islander 2.04 and lowers for blacks 1.05. Black and white had similar FYUSWRS compared to mean across the entire sample. Non-Hispanics had lower FYSWR than Hispanics. Males had a higher FYSWR than females. Data demonstrated wide variation and performance gaps between different race, ethnicity, and sex categories. Rating for opportunity of improvement was moderate.
- Yes, there is a moderate gap in care.
- True
- Gap clear
- The evidence has sufficiently identified gaps. Waitlists and processes widely vary. Disparities exists by race, sex and ethnicities.
- Not clear; only 6% of those measured are outside the average.
- There is clearly a performance gap between different race, ethnicity and sex categories. This gap is very pronounced especially in indigent areas where patients do not have the same access to care. Thus there is a high level for improvement.

Criteria 2: Scientific Acceptability of Measure Properties

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

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

- The SMP Passed on Reliability with a score of: H-0; M-10; L-0; I-0
- The SMP Passed on Validity with a score of: H-0; M-8; L-2; I-0

2a. Reliability: Specifications and Testing

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

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

Specifications:

- The measure specifications were clear and precise.

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:

- Testing was conducted using the inter-unit reliability (IUR) with a bootstrap approach. This approach utilizes a resampling procedure to estimate the within facility variation that cannot be directly estimated by ANOVA.
    - The developer calculated an IUR value of 0.64 for the measure, which indicates that 64 percent of the variation in the measure can be attributed to the between-facility differences (signal) and 36 percent to the within-facility variation (noise).
    - The developer notes that the IUR suggests a moderate degree of reliability.
    - Dialysis practitioner group practices with less than 11 eligible patients and less than two expected events were excluded from this calculation.

SMP Summary:

- Reliability passed the SMP's preliminary review and was therefore not discussed at the meeting.
- In regard to specifications, SMP noted that the developer should provide clarity around patient attribution. Additionally, they asked that the developer be clear that this is a three-year measure.
- One SMP member raised concern with the measure's ability to identify variation in performance with over 94 percent of facilities classified as "average" / "as expected".

Questions for the Committee regarding reliability:

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

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

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

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

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

Validity Testing

- Validity testing conducted at the Accountable Entity Level:
    - The developer tested the validity of the measure by evaluating the association between the dialysis practitioner group level measure performance, and subsequent mortality and overall transplant rates among all patients attributed to the practitioner groups.
    - The developers examined the Spearman correlation between the practitioner group measure value and each of the outcomes, respectively.
    - The dialysis practitioner group level second year average mortality rates are 15.3, 15.7, and 15.9 deaths per 100 patient-years for T1, T2, and T3, respectively (trend test p=0.0607). The Spearman correlation coefficient is -0.02 (p=0.3151).
    - The dialysis practitioner group level second year average transplant rates are 4.7, 3.2, and 1.8 transplants per 100 patient-years for T1, T2, and T3, respectively (trend test p=<0.01).

- The developer noted that higher FYSWR performance correlated with higher second year transplant rate, with clear separation of transplant rates across practitioner group tertiles of performance. The direction of the relationship with mortality was as expected, with numerically lower mortality with higher performance on the FYSWR measure, though it did not achieve statistical significance.

Exclusions

- The developers evaluated the exclusion criteria by comparing the differences in the number of patients with and without excluding age greater than or equal to 75, nursing home patients, and hospice patients. The developer noted that they do not exclude patients from dialysis practitioner groups with fewer than 11 attributed patients or two expected events.
- The developer reported that the number of patients before exclusions was 410,849 and after exclusions it was 281,479.
- The developer also reported the following frequencies for each excluded variable
  - Age greater than or equal to 75 – 101,658 (24.7 percent)
  - Nursing home from CMS-2728 – 18,178 (4.4 percent)
  - Nursing home from nursing home history file – 9,390 (2.3 percent)
  - Hospice – 144 (0.04 percent)
- The developer presented the distribution of performance scores before and after exclusions. The mean, standard deviation, and interquartile range did not differ greatly. The developer indicated it can be concluded that the exclusions do not significantly impact the distribution of performance scores.
- The developer stated that though performance scores are modestly affected by exclusions, they are important on clinical groups as they represent a group of patients highly unlikely to be suitable for transplant waitlisting. The developer also noted that there is fair degree of variation in the percentage of patients excluded across practitioner groups. Lastly, the developer notes that the data to determine exclusions is readily available and therefore adds no additional burden.

Risk-Adjustment

- The developer conducted a statistical risk model with 18 variables. The risk model is a two-stage Cox model that was first fitted in order to obtain an estimate of age, comorbidities, and transplant center effected to be used as an offset. During the second stage, a national average baseline hazard was estimated. The baseline, age, comorbidities, and transplant center adjustments were then used to compute the probability of an event for each patient, followed by the total expected number of events at each dialysis practitioner group practice.
- The C-statistic for the model was 0.75 which suggests good predictive ability of the risk model.
- Further, the decile plots show the risk factors in the model are discriminating well between patients. There is good separation between among all 10 deciles and the ordering is predicted by the model.
- The developers also provided rationale for the patient-level risk factors chosen for the model. The developers divided the factors into three categories social risk factors, functional risk factors, and clinical risk factors.
  - The developers stated that for social risk factors the included dual eligibility and area deprivation index (ADI) as they were significantly associate with waitlisting.
  - The developer chose to exclude race, sex, and ethnicity from the model because while there were differences in waitlisting by sex and race, it could not be determined if those difference were due to biological or other factors.

- The developer chose to include functional status factors as these are associated with worse outcomes following transplant and may indicate that a patient is inappropriate for waitlisting.

Meaningful Differences

- The dialysis practitioner groups were classified into three categories: 'As Expected, 'Better than Expected', or 'Worse than Expected'. The practitioner groups were split into these categories based on whether observed and expected values are statistically different at the 5% level. The developer reported average values of FYSWR between the groups to determine if there are practically meaningful differences in performance scores. They calculated the p-value using a Poisson approximation.
- To account for the simultaneous monitoring of many dialysis practitioner groups and to account of the unexplained variation among the groups, the developers used an approach described in a paper by Kalbfleisch and Wolfe. This approach converts p-values for each group to a Z-score that is stratified into four groups based on patient-years. They then derive the mean and variance of a normal empirical null distribution. This distribution is then used to calculate the p-value for each practitioner.
- The developers found the following information
    - Four percent of dialysis practitioners were classified as better than expected and two percent were worse than expected
    - The better-than-expected group on average have more than double observed waitlist/living donor transplant rates than that of expected waitlist/transplant rates while the worse than expected group had observed rates less than 1/5 of what was expected.
- The developer states these results suggest that there are meaningful differences.

Missing Data

- The developer stated that missing data occurs rarely for variables included in this measure.
- The element that the developer found had the most missingness was the assignment of dialysis practitioner groups.
    - The developer noted that this missingness occurs for two reasons:
        - Some patients could not be attributed to a dialysis practitioner group because they were missing the National Provider Identifier (NPI) or Unique Physician Identifier Number (UPIN) information on the CMS-2728 form
        - NPI or UPIN could not be matched with the most current and active practitioner group from the provider table.
    - The developer found that 6.2 percent of patients are missing a dialysis practitioner group.
- The developer also aggregated these patients into their own group and calculated performance scores.
    - They found that the FYSWR for this group was 1.05 which was not statistically significant from the average score, suggesting that these patients have similar waitlisting experiences to the average patient.

Comparability

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

SMP Summary:

- One SMP reviewer noted that tertiles are limited in their ability to demonstrate stability vs movement among levels.
    - The developer noted that tertiles were chosen in order to evaluate a gradient in effect but also maintain ample numbers within each group.
- One member noted the developers report a high missing practitioner rate (inability to attribute) of 6.2 percent. It is unclear why this problem does not exist for #3694 and #3695, when these are all measures of waitlisting among patients on dialysis under the care of dialysis practice groups. The same data source (IDR) is used for all three measures. However, it appears with analysis furnished by the measure submitter, these cases "have similar waitlisting experience to the average." This mitigates the concern of the large amount of missingness.
    - The developer noted that a different approach was used in this measure to capture all dialysis patients. In this measure they used CMS Form 2728 for provider attribution so as to not limit the measure to a Medicare-Fee-for-Service patient population.
- Reviewers noted several areas of concern about the risk adjustment model for consideration by the Standing Committee regarding appropriate selection of conditions and social risk factors. The SMP felt that because these factors have a severe impact on patients with end-stage renal disease (ESRD) the factors should be mitigated for, not adjusted for.
    - The developer advised that their decision to propose this measure is in large part motivated by a desire to reduce such disparities, and the factors chosen have a conceptual basis in that they are proxies for financial and social resources that can affect success following transplantation. Additionally, a Technical Expert Panel consisting of a range of stakeholders, including several patients with ESRD, discussed these issues and were in consensus about the need for social risk adjustment. A dominant concern was that in the absence of such adjustment, dialysis practitioners caring for a disproportionate share of socially vulnerable patients may inappropriately be penalized by the measure, leading to unintended adverse consequences in terms of access to care for these patients.
- One SMP member noted that they were unclear if risk factors in the model were present at the onset of measurement period (e.g., data elements from CMS #2728). The reviewer noted that this is important so as to limit the risk factors to those that were present at the start of care.
    - The developer noted that the items from CMS Form 2728 all occur prior to the start of the measurement period.

Questions for the Committee regarding validity:

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

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

Committee Pre-evaluation Comments:

2a1. Reliability – Specifications

- It is not clear if switching nephrology practices is captured.
- Specifications were clear and understandable.

- Concerns about approach for attribution to practitioner group outlined in validity.
- Exclusions were stated, but did not mention cancer or recent cancer history. These would also limit waitlisting at most transplant centers.
- Specifications are clear with no apparent barriers to consistent implementation.
- No concerns with the specifications. Data elements and logic/calculation algorithm are clearly defined. Believe the measure could be consistently implemented.
- I feel like it could help patients; however, I think more data is needed.
- Moderate
- 94 percent of providers were classified as "average/as expected."
- No concerns.
- None
- Reliability was conducted using the inter-unit reliability with bootstrap approach. This approach utilized a resampling procedure to estimate the within facility variation that cannot be directly estimated by ANOVA. Developer calculated an IUR 0.64 for the measure…64 percent of the variation in the measure can be attributed to the between facility differences and 36 percent within facility variation. IUR suggests a moderate degree of reliability. Reliability passed the SMPs preliminary review. No concerns that the likelihood that this measure cannot be consistently implemented.
- Moderately convinced that the providers will report accurate data.
- True
- No concerns.
- No issues.
- Data elements are clear.
- The measures are clear and precise using data from form 2728, CMS and transplant wait list data. Reliability appears moderate to high.

2a. Reliability – Testing

- IUR indicates only moderate reliability.
- I would ask the committee if there is concern with the variability among facilities as well as within facility? To me across facility variability seems high.
- Overall IUR 0.64 – moderate reliability; unclear what the reliability would be for smaller group practices.
- Main concern is not having all exclusion criteria available and accurate.
- IUR of 0.64 so acceptable reliability.
- Yes. While the overall IUR across all facilities is acceptable at 0.64, stratification of reliability scores by provider size was not detailed. Because of this, it's impossible to determine how widely reliability varies across the spectrum of provider/group sizes. As has often been the case with other CMS measures, reliability for small providers might be substantially lower than the overall IUR, effectively rendering the metric meaningless for use in performance measurement in this group. Request CMS provide data demonstrating reliability for all providers by detailing IURs by provider/group size. Additionally, the SMP requested that the developer clarify that this is a "three-year measure." It is not clear from the specifications that this is the case; could the developer confirm?
- Yes.
- Moderate
- The IUR is 0.64 falling below the 0.70 threshold typically applied.
- No concerns.

- No
- No moderate degree of reliability was identified meaning 64% of variation was due to between practice variation and 36% within in practice variation.
- Moderately sure accurate reporting will. occur.
- Appropriate
- None
- The IUR of 64% is acceptable.
- No
- I have no concerns about reliability testing.

2b1. Validity – Testing

- Mortality hardly differed in the 3 tertiles.
- Reviewing the SMP comments on validity does raise some concern regarding the risk adjustment for the social risk factors. Yet I believe this measure has value.
- Second year mortality Spearman correlation coefficient -0.02 (p=0.3151); second year transplant rate Spearman correlation coefficient 0.32 (p<.01).
- As above
- Data presented showing better performance on proposed measure associated with less mortality and higher transplant rates.
- No additional concerns beyond those raised by the SMP on risk adjustment (see Question 9).
- Yes
- Moderate
- 2b2-2b6 the model and measure are not valid at the accountable entity level. The overall TRANSPLANT rate (as distinct from wait listing ratio) used by the developer to test validity at the "accountable entity level" is not attributable to the Nephrologist, but is more linked to the transplant team, the regional wait list size and the match between the donor and the potential candidates on the wait list, the performance of the organ donor network for that region and a host of other factors that are no linked to the quality of the nephrologist's care.
- No concerns
- No
- No concerns with testing results. Accountable entity level Data element validation from literature is acceptable. Patient encounter level testing was not performed.
- No
- Appropriate
- No
- No issue.
- No
- Overall I do not have any concerns about testing results.

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

- Active cancer should be an exclusion.
- SMP had a concern about the risk adjustment for the social risk factors. I think the recognition of Social Risk factors is important.
- Exclusions: age >= 75 at start of dialysis, admission to SNF or hospice during evaluation month, waitlisted prior to initiation of dialysis, or transplanted prior to initiation of dialysis. The detailed information on exclusion states nursing home residence at time of dialysis initiation – clarify which

exclusion criteria apply (only if present at initiation of dialysis or assessed monthly). Clarify why patients waitlisted prior to dialysis initiation are excluded from the measure and would practitioner groups with high pre-emptive waitlisting perform poorly on this measure? Patients are attributed to practitioner group based on the NPI/UPIN entered on CMS 2728 Medical Evidence form (significantly different from the approach used in PPPW and aPPPW). Concern about the attribution approach. Risk adjustment includes age, incident comorbidities (2728), ADI, DE, weighted SRTR mortality ratio and weighted SRTR transplant ratio.

- Transplant centers do have differing wait list criteria. Therefore, regional differences may drive some of the variability and a regional or center adjuster would be ideal.
- Risk adjustment done seems reasonable and logical.
- No concerns with exclusions. The risk model appears to fit well, with a c-statistic of 0.75; however, the SMP's concerns on the inclusion of social risk variables in the final model are noted. In agreement with what appears to be the consensus from the SMP—i.e., because social risk factors have a profound impact on patients with ESRD, they should be mitigated for, not adjusted for. A discussion among Standing Committee members would be helpful on this issue.
- I would need more information.
- Moderate
- This is very difficult to comment upon -- as a process measure there should be no risk adjustment, however as a health outcome measure as listed robust risk adjustment should be present. The approach to risk adjustment put forth by the developer was commented on above. The only additional comment is that social demographics, gender and race may influence ultimate rates of transplantation. The wait list ratio measure is flawed by being attributed to the nephrologists and excluding social and demographic factors raises additional questions regarding the overall robustness and applicability of the measure.
- Exclusions listed are appropriate. Risk adjustment strategy is included.
- Valid measure by all criteria.
- Health outcome---potential social risk factor variables and the measure focus. There is a conceptual relationship between potential social risk factor variables and the measure focus. Developers stated that for social risk factors the included dual eligibility and area deprivation index as they were significantly associated with waitlisting. Chose to exclude race, sex and ethnicity from the model because while there were differences in waitlisting by sex and race, it could not be determined if those difference were due to biological or other factors. Functional status factors as these are associated with worse outcomes following transplant and may indicate that a patient is inappropriate for waitlisting. Missing data is rarely for variables included in this measure. Missing data due to NPI OR UPIN could not be matched with the most current and active practitioner group from provider table.
- The exclusions appear consistent. Most risk adjustment variables are present at the start of care. Risk-adjustment strategy is acceptable.
- True
- No concerns.
- No issues with exclusions. Risk adjustment: the social risk factors may have more consequences than the developers account for.
- Consider race measurement: while race is not a biologic construct, the data may reflect differences in referral pattern or evaluation which are important.
- Exclusions and risk adjustments are appropriate. My only concern throughout the analysis is that wait list may not mean active as patients can be on the list but inactive- and then the practitioner can take a long time to expedite the workup. But this measure is good step in the right direction.

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

- No
- Missing data was not a concern.
- 31.5% excluded for age, nursing home residence and hospice; 6.2% of patients missing dialysis practitioner group assignment.
- Yes, missing data or inaccurate data would alter validity.
- No apparent threats to validity; measure seems to distinguish between very different groups, though overall 94% of practices perform as expected so distinguishing quality within this group may rely on the overall measure performance starting to improve with more appreciable spread between providers now all lumped into expected range.
- Scores differentiated as "as expected," "better than expected," and "worse than expected." No concerns with approach. Missing data unclear. Unclear. Request additional information/clarity around the high missing practitioner/group rate of 6.2% and the impact on appropriately assigning attribution and measure validity.
- I would need more information.
- Moderate
- 2b4 meaningful differences overall 2% of providers were "worse than expected" this suggests the ability of this wait list measure to distinguish meaningful differences to help effect change is not robust.
- No concerns
- Valid measure by all criteria.
- Not implemented yet. Validity of the measure was tested by evaluating the association between dialysis practitioner group level measure performance and subsequent mortality and overall transplant rates among all patients attributed to the practitioner groups.
- Minimal threat to validity.
- Yes
- No concerns
- Missing practitioner data should be discussed further. Appreciate the developer including them in their own group and finding no significant difference. The risk model may suggest negative implications on the data.
- No issues
- I do not believe missing data constitutes a threat to validity. Only one set of specifications is used for this measure.

Criterion 3. Feasibility

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

- The developer states that the data elements are generated or collected and used by healthcare personnel during the provision of care. Further, the data elements are coded by someone other than the person obtaining original information.
- The developer states that the measure relies on data elements that are defined in a combination of electronic sources.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

3. Feasibility

- All feasible
- No concerns.
- No concerns about feasibility.
- Generally appear feasible
- Highly feasible
- No concerns with feasibility for this measure.
- I would need more information.
- Moderate
- Patient level wait listing data should be readily available. However, patient level data regarding co-morbidities are only obtained at the initiation of dialysis. The only subsequent data applied to adjust for new co-morbidities are largely gathered from hospital Medicare claims data. These co-morbid factors all impact the final decision by the nephrologist to refer, and by the transplant team to wait list, a patient. All additional data sources regarding concurrent patient related factors must be easily available, transparent, and complete for the application of the measure to be clinically meaningful.
- No concerns
- All elements routinely collected.
- Data elements are generated or collected and used by health care personnel during the provision of care. Data elements are coded by someone other than the person obtaining original information. The measure relies on data elements that are defined in a combination of electronic sources.
- Collecting electronic data makes this measure feasible.
- Appropriate
- Feasibility high
- No issues with feasibility.
- No issues
- The elements are routinely generated, measure is highly feasible for what it wants to measure.

Criterion 4: Use and Usability

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported?	☐ Yes	☒ No				
Current use in an accountability program?	☐ Yes	☒ No	☐ UNCLEAR			
Planned use in an accountability program?	☒ Yes	☐ No	☐ NA			

Accountability program details

- The developer plans to submit this for use in a public reporting or payment program such as the Quality Payment Program.

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

Feedback on the measure by those being measured or others

- This is a new measure; therefore, no feedback has been obtained by those being measured or measure users. However, during the development phase the developers convened a multistakeholder TEP which noted support for this measure.
    - The TEP specifically stated that this is an important measure given waitlisting is a crucial and necessary step for transplantation.
    - The TEP also stated that dialysis practitioners can directly contribute to the processes required for waitlisting.
    - The TEP also expressed the need for strong risk adjustment in areas of socioeconomic status and comorbidities as well as transplant center effects.
- Additionally, the developer noted that they presented the TEP with two other existing waitlist measures that are publicly reported at the facility level as a starting point for this measure.
    - The TEP advised on how the construction of the facility level measure should be revised to adapt to the practitioner level.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☒ Pass ☐ No Pass

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

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

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

Improvement results

- The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of this waitlisting measure. Once implemented dialysis practitioner group practice performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in waitlisting rates among the target population.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

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

Potential harms

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

Additional Feedback:

- This measure was reviewed by the Measure Applications Partnership (MAP) for the CMS ESRD Quality Improvement Program in 2017. The MAP recommended conditional support for rulemaking. MAP acknowledged this measure addressed an important quality gap for dialysis facilities; however, it discussed several factors that should be balanced when implementing this measure. Therefore, MAP recommended that this measure be reviewed by the Scientific Methods Committee, the Renal Standing Committee, and the need for the Disparities Standing Committee to provide guidance on potential health equality concerns.

***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

- Not in use. Only 2% of practices performed worse than expected.
- Not currently in use but developer plans to submit for inclusion in QPP.
- Potential future use in CMS programs.
- Reporting would presumably be at the provider group level
- No apparent issues with use.
- This is a new measure, not yet being publicly reported. It is intended for use in publicly reported programs; thus results will be disclosed and available to the broader public if the measure is endorsed

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

- I would need more information.
- Moderate
- TEP has commented on this new measure. It has not yet been widely assessed by others. It is not currently publicly reported. The developer states it will be used in accountability program.
- N/A
- Not yet implemented
- This is a new measure therefore no feedback has been obtained by those being measure or measure users. However, during the development phase the developers convened a multistakeholder TEP which noted support for this measure.
- Currently this measure is not reported publicly.
- True
- No concerns
- No concerns with feedback.
- Not reported; potential for future use.
- The developer plans to submit this for use in the public domain-- which is very important and should improve quality. This is a new measure and there is no feedback as of yet. Use is good/pass.

4b. Usability

- Harm if unadjusted confounders result in flagging group as worsted than expected.
- Definite benefit to measuring waitlisting.
- Exclusion of transplant waitlisting prior to dialysis initiation may have unintended consequences.
- I still feel much of the wait list determination is driven by the transplant centers. Providers making the referral often understand the limitations and criteria for waitlisting. This measure may significantly increase the number of inappropriate referrals to transplant centers.
- Reasonable to assume higher performance results would improve patient quality without apparent harms/unintended consequences
- The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. The developer did not provide an assessment of benefits vs. harms. As previously noted, however, a concern with this measure is misattribution and potential penalties levied again practitioners/groups for an outcome that is largely outside their control.
- I think the majority of patients would benefit; however, I would need more information?
- Moderate.
- 4b2 Without appropriate risk adjustment for social demographics, medical issues, a patient's own desires regarding transplantation, and the approach of the transplantation center to each of these issues, the dialysis practitioners caring for a disproportionate share of patients in a particular risk category may be inappropriately penalized by this measure. This in turn, may lead to unintended adverse consequences with regard to access to care for these patients. As currently structured, with few exceptions, all patients under age 75 must be referred by the nephrologist and presumably be waitlisted by the transplant team for this outcome measure to be satisfied. Without additional adjustments, the measure as constructed, could be unduly burdensome for patients who do not wish to undergo evaluation for transplant wait listing and transplantation. Just as we acknowledge patients' autonomy to determine their participation in other programs, such as vaccination programs, their autonomy regarding decisions such as transplant waitlisting and transplantation should be similarly respected.

- In spite of education about transplant, there are some patients who are either not interested in transplant or do not have a social environment/support structure that will allow them to be successful post-transplant. Unintended consequence is pushing someone towards transplant before they are ready and then the transplant is unsuccessful.
- Not yet implemented
- Benefits of the performance measure in facilitating progress toward achieving high quality, efficient healthcare for individual populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).
- This measure will assist in motivating provider groups who interact with Stage 4, 5 and new 5D patients to early Transplant referral. The measure is not reported as yet so we don't yet realize the benefits vs. harms.
- Appropriate
- Here is a major concern for me - - ascription. The unit of measure is the practitioner group. However, that group is NOT THE GROUP WHO MAKES DECISIONS about wait listing!! It is the transplant center that determines who is and who is not on the wait list. The appropriate measure of performance for the practitioner group would be the REFERRAL RATE, not the waitlist rate. The appropriate measure from the patient's perspective would ascribe responsibility to the Transplant team, not the practitioner group.
- No concerns.
- No issues
- There are no potential harms but only good as it is important to advance referral and waitlist of patients. Usability is moderate to high.

Criterion 5: Related and Competing Measures

Related measures

- The developer noted two non-NQF endorsed measures that are related to this measure.
    - Standardized First Kidney Transplant Waitlist Ratio for Incident Dialysis Patients (SWR)
    - Percentage of Prevalent Patients Waitlisted (PPPW)

Harmonization

- The developer stated that they harmonized this measure with other non-NQF endorsed measures to the extent possible.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- Unendorsed SWR, PPPW
- There are several similar waitlisting measures but with various nuances.
- There may be increased inappropriate referrals which may slow down the referral and workup for appropriate patients.
- N/A
- No concerns.
- Unknown
- There are two other competing non NQF endorsed measures which the developer says are harmonized.

- Measure shares features with Facility level Standardized First kidney transplant waitlist ratio for incident dialysis patients.
- N/A
- Harmonized with existing non-NQF measures
- No
- No
- True
- None
- No concerns.
- None
- The developer harmonized this measure to non NQF measures SWR and PPPW.

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

Member Expression of Support

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

Comments

Comment 1 by: Lisa McGonigal, Kidney Care Partners; Submitted by Lisa McGonigal, Kidney Care Partners

Practitioner/Group-Level First Year Standard Waitlist Ratio (NQF 3689, CMS) Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted in Active Status (NQF 3694, CMS) Practitioner/Group-Level Percentage Of Prevalent Patients Waitlisted (NQF 3695, CMS) KCP recognizes the tremendous importance of improving transplantation rates for patients with ESRD, but does not support the attribution of successful or unsuccessful waitlisting to dialysis facilities, individual practitioners, or group practices and thus cannot support these measures. KCP believes that while a referral to a transplant center and initiation or even completion of the waitlist evaluation process might be appropriate measures for these levels of analysis that could be used in CMS's quality programs, the newly proposed practitioner/group level Percentage of Prevalent Patients Waitlisted (PPPW), Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW), and First Year Standardized Waitlist Ratio (FYSWR) measures are not. Waitlisting per se is a decision made by the transplant center and is beyond the locus of control of any of the providers targeted in these measures. In reviewing these measures, we offer the following comments: I. Overarching Concerns Several of KCP's concerns apply to all three proposed transplant access measures: a. Attribution. As above, we strongly object to attributing successful/unsuccessful placement on a transplant waitlist to dialysis facilities, individual clinicians, or practitioner group practices and believe this is a fatal structural flaw with these measures. The transplant center decides whether a patient is placed on a waitlist, not the facility, practitioner, or group practice. KCP patient members who are transplant recipients have noted there are many obstacles and delays in the evaluation process with multiple parties that have nothing to do with the facility or practitioner—e.g., one patient noted their private pay insurance changed the locations where they could be evaluated for transplant eligibility on multiple occasions, repeatedly interrupting the process mid-stream. Penalizing a practitioner/group practice each month through the PPPW, aPPPW, and FYSWR for these or other delays is inappropriate; such misattribution is fundamentally misaligned with NQF's first "Attribution Model Guiding Principle," which states that measures' attribution models should fairly and accurately assign accountability. KCP emphasizes our commitment to improving transplantation access, but we believe other measures with an appropriate sphere of control should be pursued. For instance, our sister

organization, the Kidney Care Quality Alliance (KCQA), has developed a dialysis facility-level Transplant Access Measure Set that will be submitted to NQF for endorsement consideration later this year. The set pairs a referral rate metric with a measure assessing the waitlisting rate specifically among those patients who were referred by the facility within the preceding three years. Because the KCQA waitlisting measure denominator is limited to those patients who were deliberately referred by the dialysis facility within a defined time period, facilities have considerably more agency over the measure than metrics such as the PPPW; this construct will also provide a counterbalance to the referral measure, curbing the tendency to indiscriminately refer patients who are not appropriate transplant candidates, preventing unnecessary patient and transplant center burden. The same approach could be applied at the practitioner/group level. B. Variation in Transplant Center Eligibility Criteria. We also note that criteria indicating a patient is "not eligible" for transplantation can differ by location. For instance, one center might require evidence of an absence of chronic osteomyelitis, infection, heart failure, etc., while another may apply eligibility exclusions differently or have additional or different criteria. The degree to which these biological factors influence waitlist placement must be accounted for in any model for the measure to be a valid representation of waitlisting. C. Stratification of Reliability Results by Group Size and Performance Scores Absent. We also note that CMS has provided no stratification of reliability scores by provider size for the measures; we are thus unable to discern how widely reliability varies across the spectrum of practitioner/group practice sizes. We are concerned that the reliability for small providers might be substantially lower than the overall IURs, as has been the case, for instance, with other CMS standardized ratio measures. This is of particular concern with the FYSWR, for which empiric testing has yielded an overall IUR of only 0.64—interpreted as "moderate" reliability by statistical convention. To illustrate our concern, the Standardized Transfusion Ratio for Dialysis Facilities (STrR) measure (NQF 2979) also was found to have an overall IUR of 0.60; however, the IUR was only 0.3 ("poor" reliability) for small facilities (defined by CMS as <=46 patients for the STrR). Without evidence to the contrary, KCP is concerned that FYSWR reliability is similarly lower for small groups, effectively rendering the metric meaningless for use in performance measurement in this subset of providers. KCP believes it is incumbent on CMS to demonstrate reliability for all providers by stratifying data by practice size.

Comment 2 by: Submitted by David White, American Society of Nephrology

TO: NQF Renal Standing Committee

FR: Tod Ibrahim, Executive Vice President, the American Society of Nephrology Members of the National Quality Forum Renal Standing Committee The more than 37,000,000 Americans living with kidney diseases and the 21,000 nephrologists, scientists, and other kidney health care professionals who are members of the American Society of Nephrology (ASN), thank you for the opportunity to comment on the 5 proposed transplantation, vascular access, and modality education measures under consideration: • Facility-Level Standardized Modality Switch Ratio for Incident Dialysis Patients (SMoSR) • Facility-Level Standardized Fistula Rate for Incident Patients (ISFR) • Practitioner/Group-Level First Year Standard Waitlist Ratio (FYSWR) • Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted (PPPW) • Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW) Based on our review, ASN is concerned by several aspects of the measures and offers comment on all five measures submitted to NQF: • Focus on incident maintenance dialysis populations with "stand alone" measures that are independent of measures targeting patients in other stages of kidney diseases such as non-dialysis advanced chronic kidney disease and prevalent dialysis. • Reliance on CMS-2728 data (End Stage Renal Disease Medical Evidence Report Medicare Entitlement and/or Patient Registration) for any risk adjustment including transplant measures • Attribution of measures to dialysis facilities • Lack of adjustment for variables that are critical for patient equity, such as social determinants of health • Focus on dialysis unit-specific measures, without consideration of advanced CKD care and nephrologist-led care

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☒ No**
2. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - **Reviewer 1:** None
   - **Reviewer 2:** None
   - **Reviewer 3:** The developers should be more specific about patient attribution. Additionally, it will be helpful to say explicitly if this is a 3-year measure.
   - **Reviewer 4:** No concerns
   - **Reviewer 5:** No concerns
   - **Reviewer 7:** sp.06: Add the age limit for adults (75yo)
   - **Reviewer 8:** No concerns
   - **Reviewer 9:** Age exclusion. SNF exclusion. Fewer than 11 patients for a given practitioner group.
   - **Reviewer 11:** NOTE: This measure does not address any patient condition that could be improved with a clinical intervention. I would classify the type of measure as "process: appropriate use" as the measure encourages practitioner groups to quickly place patients needing kidney or pancreas transplants on a waitlist.

RELIABILITY: TESTING

3. **Reliability testing level: ☒ Accountable-Entity Level ☐ Patient/Encounter Level ☐ Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure: ☒ Yes ☒ No**
5. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted? **☒ Yes ☐ No**
6. **Assess the method(s) used for reliability testing:**
   - **Reviewer 1:** The developer used accepted IUR statistics to assess reliability.
   - **Reviewer 2:** Appropriate methods; ANOVA. Calculated inter-unit reliability.
   - **Reviewer 3:** For reliability testing, the developers calculated IUR, which is acceptable.
   - **Reviewer 4:** Reliability by calculating inter-unit reliability (IUR) for the annual performance scores was appropriate.
   - **Reviewer 5:** Developers estimate overall IUR using a bootstrap approach. This seems appropriate, but they are unable to describe how reliability varies according to the practice group's sample size.
   - **Reviewer 7:** No concerns
   - **Reviewer 8:** The use of a ANOVA and the IUR is a logical test of reliability for this given measure.
   - **Reviewer 9:** Data sources are claims and registry based. The First Year Standardized Waitlist Ratio is an observed/expected datapoint based on a risk model ANOVA approach was addressed to get at the between practice variation and the within practice variation. IUR (inter-unit reliability) was used to assess the attribution to between-practice variation. It was estimated by a bootstrap approach utilizing a resampling scheme. It was only utilized for those practices with at least 11 eligible patients.

- **Reviewer 11:** The developer used any ANOVA approach comparing between and within variance across provider groups. They report an IUR (inter-unit reliability) for the between/total ratio.
- **Reviewer 12:** IUR with bootstrap, excl <11. Unclear if risk-adjusted in model. The IUR is 0.64. Dialysis practitioner group practices with < 11 eligible patients and < 2 expected events were excluded from this calculation

7. **Assess the results of reliability testing**
   - **Reviewer 1:** Reliability is adequate with IUR reported at .64.
   - **Reviewer 2:** IUR was 0.64, which indicates a moderate degree of reliability.
   - **Reviewer 3:** The IUR is 0.64. This is common for a risk adjusted outcome measure.
   - **Reviewer 4:** No concerns but IUR only indicated moderate reliability.
   - **Reviewer 5:** IUR=0.64, which is moderate. However, this overall estimate suggests that the measure may have low reliability for smaller practice groups.
   - **Reviewer 7:** A moderate degree of reliability was identified, thus the moderate rating for reliability.
   - **Reviewer 8:** The testing result is an IUR of 0.64. This was the sole test result shared. Unfortunately we did not receive test results at various cut points (e.g. denominator at various percentile placements). The 0.64 figure reflects a reliability figure that is marginally acceptable.
   - **Reviewer 9:** The IUR was 0.64, meaning that 64 % of the variation was due to between practice variation and 36% is within-practice variation, which supports moderate reliability.
   - **Reviewer 11:** The IUR value was 0.64 for groups with >10 patients and >1 expected events. This represents a moderate level of reliability.
   - **Reviewer 12:** The IUR is 0.64. Dialysis practitioner group practices with < 11 eligible patients and < 2 expected events were excluded from this calculation
8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? **NOTE:** If multiple methods used, at least one must be appropriate.

   ☒ **Yes** ☐ **No** ☐ **Not applicable**
9. 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)
10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

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

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

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

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)
11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - **Reviewer 1:** IUR is acceptable.
    - **Reviewer 2:** Appropriate methods; results demonstrated moderate reliability.
    - **Reviewer 3:** IUR is 0.64.
    - **Reviewer 4:** No concerns, just rated it lower than the 2 other waitlist measures because of a lower reliability IUR.

- **Reviewer 5:** Overall measure is moderate, but a higher denominator threshold (>11) may be necessary to avoid reporting on practices for which reliability would be poor.
- **Reviewer 7:** A moderate degree of reliability was identified, thus the moderate rating for reliability.
- **Reviewer 8:** Response to question 7: The testing result is an IUR of 0.64. This was the sole test result shared. Unfortunately we did not receive test results at various cut points (e.g. denominator at various percentile placements). The 0.64 figure reflects a reliability figure that is marginally acceptable.
- **Reviewer 9:** IUR by an accepted methodology was 0.64.
- **Reviewer 11:** The methodologies describe to assess reliability were appropriate. The results for the item reliability were moderate. My concern is that this measure does not fit the definition of a patient outcome, but is better described as a practice process measure.
- **Reviewer 12:** Was model risk adjusted? Threaten false partition of variation to level vs patient.

VALIDITY: TESTING

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

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

13. **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)

14. **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)**

15. **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)

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

- **Reviewer 1:** Testing validity by correlations with subsequent transplant rates and mortality rates is acceptable, but not highly compelling, particularly as mortality rates are affected by many other factors and were not, in fact, correlated with performance on the wait list measure.
- **Reviewer 2:** Tested the association between group-level performance on the measure and group-level mortality and group-level transplant rates. Reasonable relationships to test. Hypotheses of relationships were reasonable.
- **Reviewer 3:** The developers assessed the validity of this measure through evaluating the association between this measure and two related quality measure. One is subsequent mortality measure and another is overall transplant rate. The developers provided conceptual rationale on what to expect.
- **Reviewer 4:** Validity of the measure was tested by evaluating the association between the dialysis practitioner group level measure performance, and subsequent mortality and overall transplant rates among all patients attributed to the practitioner groups

- **Reviewer 5:** Construct validity is assessed by estimating practice-level associations between this measure and subsequent-year mortality rates and subsequent-year transplant rates. This is an elegant approach that removes autocorrelation between measures.
- **Reviewer 7:** The first hypothesis, if I understand correctly, is about a correlation between two measures that assess basically the same thing: performance on the FYSWR measure and transplant rates, which are the basis of the FYSWR measure. A significant correlation is expected, as described, but provides low evidence of empirical validity given the circularity of this test. The second hypothesis is a better test for empirical validity, as it assesses the correlation between two different measure that are expected to be associated with each other: the FYSWR measure and mortality rates.
- **Reviewer 8:** The test of the relationship between the FYSWR measure and two other measures (i.e. transplant rate and mortality rate) are appropriate for testing validity. However, we were not presented with information as to the methodology for calculating transplant rates and mortality rates.
- **Reviewer 9:** Validity testing was correlation to another metric, namely the subsequent mortality and overall transplant rates for the applicable patients in a given practice group. Assumption was that higher results on this metric would be associated with a higher transplant rate and lower mortality. Groups were divided into turtles to maintain reasonable numbers. A Cochran-Armitage trend test was utilized to evaluate the relationship between these two outcomes and then a Spearman correlation coefficient.
- **Reviewer 11:** Tertile comparison of measure scores is rather limited in its ability to demonstrate validity (i.e., stability vs movement among levels).
- **Reviewer 12:** Validity of the measure was tested by evaluating the association between the dialysis practitioner group level measure performance, and subsequent mortality and overall transplant rates among all patients attributed to the practitioner groups.

17. **Assess the results(s) for establishing validity**
    - **Reviewer 1:** The correlation with subsequent transplant rate is a reasonable basis for validity of the measure, but not a particularly strong basis.
    - **Reviewer 2:** Relationship between measure and transplant rates were correlated. Non-statistically significant result with mortality.
    - **Reviewer 3:** The results on the association between this measure and overall transplant rate provide support to the validity of this measure. Higher measure score correlated with higher overall transplant rate.
    - **Reviewer 4:** Adequate results to distinguish performance between practice groups.
    - **Reviewer 5:** Associations are clinically significant and in the expected direction for both comparisons, although the association with subsequent-year mortality is not statistically significant.
    - **Reviewer 7:** Focusing on the second hypothesis, no correlation was found [-0.02 (p=0.3151)], providing low evidence for the measure's empirical validity, driving the low validity rating. Given this, it is advised that other forms of validity are tested, possibly face validity or data element validity.
    - **Reviewer 8:** The test result regarding mortality rate was very weak: -0.02. The result regarding the transplant rate was modest: 0.32.
    - **Reviewer 9:** Results were as expected for the tertiles but clearer for the second year transplant rate than for the mortality. This was expected given the many potential causes for mortality. 81 (4%) were better than predicted from the model, 2,022 (94%) were as expected, and only 54 (2%) were worse than expected. These results were deemed meaningful and statistically significant.
    - **Reviewer 11:** Measure scores at highest level were highly skewed. No difference in mortality; transplant rate showed positive relationship with measure score.

- **Reviewer 12:** The dialysis practitioner group average second year mortality is 15.3, 15.7, 15.9 deaths per 100 patient-years for $TT1$, $TT2$, $TT3$ groups, respectively (trend test p=0.0607). The Spearman correlation coefficient is: -0.02 (p=0.3151). The dialysis practitioner group average second year transplant rate is 4.7, 3.2, 1.8 transplants per 100 patient-years for the $TT1$, $TT2$, $TT3$ groups, respectively (trend test p<.01). The Spearman correlation coefficient is: 0.32 (p<.01).

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**
    - **Reviewer 1:** None
    - **Reviewer 2:** None. All exclusions are clinically important.
    - **Reviewer 3:** No concern
    - **Reviewer 4:** Appropriate exclusions
    - **Reviewer 5:** The developers report a surprisingly high missing-practitioner rate (inability to attribute) of 6.2%. It is extremely unclear why this problem does not exist for #3694 and 3695, when these are all measures of waitlisting among patients on dialysis under the care of dialysis practice groups. The same data source (IDR) is used for all three measures.
    - **Reviewer 7:** No concerns
    - **Reviewer 8:** No concerns
    - **Reviewer 9:** Missing data elements were present in 6.2% due to patients mismatch to a given practitioner NPI. Excluded patients were tested for performance scores and were modestly affected by the exclusions.
    - **Reviewer 11:** Measure exclusions identified but did not make a difference in measure score.
    - **Reviewer 12:** Must exclude cancer patients, scleroderma patients within first 2 years of diagnosis, ask a nephrologist.
19. **Risk Adjustment**

19a. **Risk-adjustment method**

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

☐ Other method assessing risk factors (please specify)

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

☒ Yes ☒ No ☒ Not applicable

19c. **Social risk adjustment:**

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

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

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

19d. **Risk adjustment summary:**

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

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

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

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No

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

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

- **Reviewer 1:** Risk adjustment decisions were reasonable, with dual-eligibility and ZIP-level ADI included as social risk factors.
- **Reviewer 2:** Appropriate methods. C-statistic of 0.75.
- **Reviewer 3:** It is much preferred if the developers validated the risk model using an independent dataset. Adjusting for transplant center mortality ratio is another concern. It is easier to see adjusting for transplant ratio, and yet the effect of mortality ratio is much stronger than transplant ratio.
- **Reviewer 4:** Very thorough analysis of risk adjustment and appropriate inclusion of risk variables.
- **Reviewer 5:** Please see comments on #3694 and 3695. Including social risk factors (ADI, dual eligibility) in the risk-adjustment model for this process measure is shocking and unconscionable when overwhelming evidence indicates that social disparities in access to transplantation have a severe impact on patients with ESRD. These social factors are in the causal pathway and must be mitigated, not adjusted for. For some reason, the developers are comfortable omitting sex and race, but they insist on including dual eligibility and ADI.
- **Reviewer 8:** One concern / question I had was whether a number of the risk factors were present at the onset of the measurement period. There are several data elements in the risk model that could occur before or after onset of care are captured on the CMS form 2728. There's a lack of detail as to when the form is completed that is then used in the risk adjustment. Example: Let's say the measurement period is CY20. Regarding the risk factors captured in the CMS form 2728: What is the date (range) of the completed CMS form 2728 used in the extracting of risk factors? If the answer is the instruction is to pull the most recently available completed CMS form 2728 for use in risk adjustment, that becomes an issue completed forms are used after Jan 1, 2020.
- **Reviewer 9:** Risk adjustment included Medicare Dual eligibility and Area Deprivation Index. Ethnicity and race were also included.
- **Reviewer 11:** Developer identifies the measure score as an outcome and makes a valiant effort to describe how this "outcome" could be risk adjusted using a statistical process (see pg. 30 in their document). I think the measure score is a process measure and should not be risk adjusted.
- **Reviewer 12:** I did not see data describing SES and outcome in model.

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

    For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?

    - **Reviewer 1:** The measure can only identify extreme outliers. The vast majority of groups fall in the "as expected" category.
    - **Reviewer 2:** None. 94% of clinician groups were "as expected".
    - **Reviewer 3:** No concern
    - **Reviewer 4:** No concerns
    - **Reviewer 5:** In this case, there is some basis for concern. Only 81 group practices performed statistically better than expected, and their median FYSWR was 2.59. Only 54 practices performed statistically worse than expected, and their median FYSWR was 0.19, indicating that this measure cannot distinguish clinically meaningful differences in performance (e.g., an FYSWR of 0.5).
    - **Reviewer 7:** No concerns
    - **Reviewer 8:** In response to 2b.06, 94% of facilities perform "as expected" per statistical testing. In turn, only 6% of facilities are "high" or "low" outliers. This low rate of identifying outlier facilities means the measure is of low value in aiding consumers in their decision making based on quality (as this measure defines quality).
    - **Reviewer 11:** The measure score is compared with an expected score. The result is reporting provider group as "Better than—As—Worst than" Expected. The measure is not a patient outcome. Hence, the comparison among provider groups is based on dubious predictive standards. The developer/owner could simply establish "industry standard" strata and report provider group performance based on these strata.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    - **Reviewer 1:** N/A
    - **Reviewer 2:** Not applicable.
    - **Reviewer 5:** Not applicable.
    - **Reviewer 7:** Not applicable.
    - **Reviewer 8:** No concern as the measure does not employ multiple data sources nor methods.
    - **Reviewer 9:** None
22. **Please describe any concerns you have regarding missing data.**
    - **Reviewer 1:** None
    - **Reviewer 2:** None. Patients with a missing dialysis practitioner had scores that were not statistically significantly different from the average score.
    - **Reviewer 3:** No concern
    - **Reviewer 4:** No concerns
    - **Reviewer 5:** 6.2% missingness on NPI/provider group is surprisingly high, although it does not appear to have a substantial effect overall. Why is this problem so much worse for this measure than for #3694/3695?
    - **Reviewer 7:** No concerns
    - **Reviewer 8:** Some concern that over 6% of cases have no NPI stated on the CMS form 2728 or there's no match to the current/active practitioner group from the IDR table. However, it appears with analysis furnished by the measure submitter, these cases "have similar waitlisting experience to the average". This mitigates the concern of the large amount of missingness.
    - **Reviewer 9:** None
    - **Reviewer 11:** Minimal missing data.

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 25.

23. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**
24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**
25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **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 that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)
26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
    - **Reviewer 1:** The one empirical test of validity was significant and in the predicted direction.

- **Reviewer 2:** Solid empirical validity approach. Measure only had a statistically significant relationship with transplant rates.
- **Reviewer 4:** No concerns
- **Reviewer 5:** Major threat to validity is the inappropriate inclusion of social factors in the risk-adjustment model, which obscures the impact of discrimination by adjusting for factors in the quality pathway.
- **Reviewer 7:** Low evidence of validity given the empirical validity results. As noted above, it is recommended that other forms of validity are tested.
- **Reviewer 8:** Response to question 17: The test result regarding mortality rate was very weak: -0.02. The result regarding the transplant rate was modest: 0.32. Response to question 19: One concern / question I had was whether a number of the risk factors were present at the onset of the measurement period. There are several data elements in the risk model that could occur before or after onset of care are captured on the CMS form 2728. There's a lack of detail as to when the form is completed that is then used in the risk adjustment. Example: Let's say the measurement period is CY20. Regarding the risk factors captured in the CMS form 2728: What is the date (range) of the completed CMS form 2728 used in the extracting of risk factors? If the answer is the instruction is to pull the most recently available completed CMS form 2728 for use in risk adjustment, that becomes an issue completed forms are used after Jan 1, 2020. Response to question 20: In response to 2b.06, 94% of facilities perform "as expected" per statistical testing. In turn, only 6% of facilities are "high" or "low" outliers. This low rate of identifying outlier facilities means the measure is of low value in aiding consumers in their decision making based on quality (as this measure defines quality).
- **Reviewer 9:** C-statistic of 0.75. Rationale of analysis makes sense, given the data available.
- **Reviewer 11:** The measure is not a patient outcome and use of risk adjustment is not advised for process measures. Measure has some potential as a process measure, although reporting of performance should be reconsidered.
- **Reviewer 12:** Good; just address exclusions question please.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**
    - N/A

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - **Reviewer 5:** Again, the inappropriate inclusion of social factors in risk-adjustment models for PROCESS measures such as these waitlist measures must be discussed.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

---

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

2021 Submission:

Updated evidence information here.

2018 Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]

This measure tracks the outcomes of new placement on the kidney or kidney-pancreas transplantation waitlist or receipt of a living donor transplant within the first year after dialysis initiation, with the intended objective of improving the overall health of patients on dialysis. Being waitlisted or receiving a living donor kidney transplant are outcomes as they represent a desirable change in health status for patients on dialysis, indicating achievement of a health condition conducive to kidney transplantation. These outcomes result from specific activities directed by dialysis practitioners with the particular goal of achieving suitability for kidney transplantation by addressing the specific healthcare needs of patients on dialysis. These activities can include, but are not limited to, ensuring an ideal dialysis prescription and care, correction and optimization of common underlying chronic health conditions such as heart failure, coronary artery disease, diabetes mellitus, hyperparathyroidism, and obesity, and as needed, optimizing mental health and social support systems. In addition, dialysis practitioners support the path for patients towards waitlisting or living donor transplantation through proper education about the transplantation option, referral to a transplant center and assistance with completion of the transplant evaluation process. The logic model for the steps involved is diagrammed below (with the outcome measure in bold):

Patients with ESRD are initiated on dialysis -> Patients not already on the wait list are educated about the option of kidney transplantation and assessed for eligibility for transplant referral by a dialysis practitioner -> Patients are referred to a transplant center for evaluation of candidacy for kidney or kidney-pancreas transplantation -> Dialysis practitioners assist patient with completion of the transplant evaluation process and optimizing their health and functional status -> **Patients deemed to be candidates for transplantation**

who have compatible living donors receive living donor transplant; otherwise they are placed on the wait list with the potential to receive a deceased donor transplant.

[Response Ends]

1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

[Response Begins]

Two previous Technical Expert Panels (TEP) have been convened to discuss potential measures directed at improving access to kidney transplantation, in 2015 and most recently, in 2021 (2015 TEP Report: https://dialysisdata.org/sites/default/files/content/ESRD_Measures/Access_To_Kidney_Transplantation_TEP_Summary_Report.pdf; 2021 TEP Report: https://dialysisdata.org/content/esrd-measures, please see Practitioner Level Measurement of Effective Access to Kidney Transplantation under Ongoing Technical Expert Panels section). Both were comprised of relevant stakeholders, including dialysis nephrologists, transplant nephrologists, transplant surgeons, social workers, researchers, and notably, patient representatives with a history of end-stage kidney disease. Discussions during both TEPs revealed broad support for the importance of waitlisting, and formal voting demonstrated a majority of TEP members were in favor of the development of quality measures targeting waitlisting (at the dialysis facility level for the 2015 TEP, and the practitioner level for the 2021 TEP).

In addition to the above, empirical support for the value of waitlisting to patients comes from a published study reporting on a large survey of 409 patients or family members who agreed to receiving emails from the National Kidney Foundation (Husain S.A. et al, Am. J. Transplant 2018;18(11):2781-2790). Participants included both patients with advanced chronic kidney disease prior to transplant, and recipients of transplants, who were asked about their priorities in choice of a transplant center. Notably, participants were most likely (a plurality of participants) to rank waitlisting characteristics (such as ease of getting on the waitlist) as the most important feature, in contrast to other transplant center characteristics such as post-transplant outcomes and practical considerations (e.g. distance to center).

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]

National or large regional studies provide strong empirical support for the association between processes under dialysis practitioner control and subsequent waitlisting. In one large regional study conducted on facilities in the state of Georgia, a standardized dialysis facility referral ratio was developed, adjusted for age, demographics and comorbidities (Paul S. et al, Clin J Am Soc Nephrol 2018;13:282-289). There was substantial variability across dialysis facilities in referral rates, and a Spearman correlation performed between ranking on the referral ratio and dialysis facility waitlist rates was highly significant ($r=0.35$, $p<0.001$). A national study using registry data (United States Renal Data System) from 2005-2007 examined the association between whether patients were informed about kidney transplantation (based on reporting on the Medical Evidence Form 2728) and subsequent access to kidney transplantation (waitlisting or receipt of a live donor transplant) (Kucirka LM et al. Am J Transplant 2012;12:351-357). Approximately 30% of patients were uninformed about

kidney transplantation, and this was associated with half the rate of access to transplantation compared to patients who were informed. In a related survey study of 388 hemodialysis patients, whether provision of information about transplantation by nephrologists or dialysis staff occurred was directly confirmed with patients (Salter ML et al, J Am Soc Nephrol 2014;25:2871-2877). Patient report of provision of such information was associated with a three-fold increase in likelihood of waitlisting. Finally, a large survey study of 170 dialysis facilities in the Heartland Kidney Network (Iowa, Kansas, Missouri and Nebraska) was conducted to examine transplant education practices (Waterman AD et al, Clin J Am Soc Nephrol 2015;10:1617-1625). Facilities employing multiple (>3) transplant education strategies (e.g. provision of brochures, referral to formal transplant education program, distribution of transplant center contact information) had 36% higher waitlist rates compared to facilities employing fewer strategies.

[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]

A measure focusing on the outcome of waitlisting is appropriate for several reasons. First, in preparing patients for suitability for waitlisting, dialysis practitioners optimize their health and functional status, improving their overall health state. Second, waitlisting is a necessary step prior to potential receipt of a deceased donor kidney transplant (receipt of a living donor kidney is also accounted for in the measure), which is known to be beneficial for survival and quality of life [1]. Third, dialysis practitioners exert substantial control over the processes that result in waitlisting. This includes proper education of dialysis patients on the option for transplant, referral of appropriate patients to a transplant center for evaluation, and assisting patients with completion of the transplant evaluation process, in order to increase their candidacy for transplant waitlisting. These types of activities are included as part of the conditions for coverage for Medicare certification of ESRD dialysis facilities. Finally, wide regional and facility variations in waitlisting rates highlight substantial room for improvement for this measure [2-5].

Additionally, this measure focuses specifically on the population of patients incident to dialysis, examining for waitlist or living donor transplant events occurring within a year of dialysis initiation. This will evaluate and encourage rapid attention from dialysis practitioner groups to the optimization of health of patients to ensure early access to the waitlist, which has been demonstrated to be particularly beneficial [6-9]. This measure contrasts with the other proposed waitlisting measures, which focus on a prevalent population of dialysis patients and encourage maintenance of patients on the waitlist (Percent of Prevalent Patients Waitlisted and Percent of Prevalent Patients Waitlisted in Active Status).

1. *Tonelli M, Wiebe N, Knoll G, et al. Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. American Journal of Transplantation 2011;11:2093-2109.*

Abstract: Individual studies indicate that kidney transplantation is associated with lower mortality and improved quality of life compared with chronic dialysis treatment. We did a systematic review to summarize the benefits of transplantation, aiming to identify characteristics associated with especially large or small relative benefit. Results were not pooled because of expected diversity inherent to observational studies. Risk of bias was assessed using the Downs and Black checklist and items related to time-to-event analysis techniques. MEDLINE and EMBASE were searched up to February 2010. Cohort studies comparing adult chronic dialysis patients with kidney transplantation recipients for clinical outcomes were selected. We

identified 110 eligible studies with a total of 1 922 300 participants. Most studies found significantly lower mortality associated with transplantation, and the relative magnitude of the benefit seemed to increase over time ($p < 0.001$). Most studies also found that the risk of cardiovascular events was significantly reduced among transplant recipients. Quality of life was significantly and substantially better among transplant recipients. Despite increases in the age and comorbidity of contemporary transplant recipients, the relative benefits of transplantation seem to be increasing over time. These findings validate current attempts to increase the number of people worldwide that benefit from kidney transplantation.

2. *Ashby VB, Kalbfleisch JD, Wolfe RA, et al. Geographic variability in access to primary kidney transplantation in the United States, 1996-2005. American Journal of Transplantation 2007; 7 (5 Part 2):1412-1423.*

Abstract: This article focuses on geographic variability in patient access to kidney transplantation in the United States. It examines geographic differences and trends in access rates to kidney transplantation, in the component rates of wait-listing, and of living and deceased donor transplantation. Using data from Centers for Medicare and Medicaid Services and the Organ Procurement and Transplantation Network/Scientific Registry of Transplant Recipients, we studied 700,000+ patients under 75, who began chronic dialysis treatment, received their first living donor kidney transplant, or were placed on the waiting list pre-emptively. Relative rates of wait-listing and transplantation by State were calculated using Cox regression models, adjusted for patient demographics. There were geographic differences in access to the kidney waiting list and to a kidney transplant. Adjusted wait-list rates ranged from 37% lower to 64% higher than the national average. The living donor rate ranged from 57% lower to 166% higher, while the deceased donor transplant rate ranged from 60% lower to 150% higher than the national average. In general, States with higher wait-listing rates tended to have lower transplantation rates and States with lower wait-listing rates had higher transplant rates. Six States demonstrated both high wait-listing and deceased donor transplantation rates while six others, plus D.C. and Puerto Rico, were below the national average for both parameters.

3. *Satayathum S, Pisoni RL, McCullough KP, et al. Kidney transplantation and wait-listing rates from the international Dialysis Outcomes and Practice Patterns Study (DOPPS). Kidney Intl 2005 Jul; 68 (1):330-337.*

Abstract: BACKGROUND: The international Dialysis Outcomes and Practice Patterns Study (DOPPS I and II) allows description of variations in kidney transplantation and wait-listing from nationally representative samples of 18- to 65-year-old hemodialysis patients. The present study examines the health status and socioeconomic characteristics of United States patients, the role of for-profit versus not-for-profit status of dialysis facilities, and the likelihood of transplant wait-listing and transplantation rates.

METHODS: Analyses of transplantation rates were based on 5267 randomly selected DOPPS I patients in dialysis units in the United States, Europe, and Japan who received chronic hemodialysis therapy for at least 90 days in 2000. Left-truncated Cox regression was used to assess time to kidney transplantation. Logistic regression determined the odds of being transplant wait-listed for a cross-section of 1323 hemodialysis patients in the United States in 2000. Furthermore, kidney transplant wait-listing was determined in 12 countries from cross-sectional samples of DOPPS II hemodialysis patients in 2002 to 2003 (N= 4274).

RESULTS: Transplantation rates varied widely, from very low in Japan to 25-fold higher in the United States and 75-fold higher in Spain (both P values <0.0001). Factors associated with higher rates of transplantation included younger age, nonblack race, less comorbidity, fewer years on dialysis, higher income, and higher education levels. The likelihood of being wait-listed showed wide variation internationally and by United States region but not by for-profit dialysis unit status within the United States.

CONCLUSION: DOPPS I and II confirmed large variations in kidney transplantation rates by country, even after adjusting for differences in case mix. Facility size and, in the United States, profit status, were not associated with varying transplantation rates. International results consistently showed higher transplantation rates for younger, healthier, better-educated, and higher income patients.

4. *Patzer RE, Plantinga L, Krisher J, Pastan SO. Dialysis facility and network factors associated with low kidney transplantation rates among United States dialysis facilities. Am J Transplant. 2014 Jul; 14(7):1562-72.*

Abstract: Variability in transplant rates between different dialysis units has been noted, yet little is known about facility-level factors associated with low standardized transplant ratios (STRs) across the United States End-stage Renal Disease (ESRD) Network regions. We analyzed Centers for Medicare & Medicaid Services Dialysis Facility Report data from 2007 to 2010 to examine facility-level factors associated with low STRs using multivariable mixed models. Among 4098 dialysis facilities treating 305 698 patients, there was wide variability in facility-level STRs across the 18 ESRD Networks. Four-year average STRs ranged from 0.69 (95% confidence interval [CI]: 0.64-0.73) in Network 6 (Southeastern Kidney Council) to 1.61 (95% CI: 1.47-1.76) in Network 1 (New England). Factors significantly associated with a lower STR (p<0.0001) included for-profit status, facilities with higher percentage black patients, patients with no health insurance and patients with diabetes. A greater number of facility staff, more transplant centers per 10,000 ESRD patients and a higher percentage of patients who were employed or utilized peritoneal dialysis were associated with higher STRs. The lowest performing dialysis facilities were in the Southeastern United States. Understanding the modifiable facility-level factors associated with low transplant rates may inform interventions to improve access to transplantation.

5. *Melanson TA, Gander JC, Rossi A, et al. Variation in Waitlisting Rates at the Dialysis Facility Level in the Context of Goals for Improving Kidney Health in the United States. Kidney International Reports 2021;6:1965-1968.*

No abstract.

6. *Meier-Kriesche, Herwig-Ulf, and Bruce Kaplan. "Waiting time on dialysis as the strongest modifiable risk factor for renal transplant outcomes: A Paired Donor Kidney Analysis." Transplantation 74.10 (2002): 1377-1381.*

Abstract: BACKGROUND: Waiting time on dialysis has been shown to be associated with worse outcomes after living and cadaveric transplantation. To validate and quantify end-stage renal disease (ESRD) time as an independent risk factor for kidney transplantation, we compared the outcome of paired donor kidneys, destined to patients who had ESRD more than 2 years compared to patients who had ESRD less than 6 months.

METHODS: We analyzed data available from the U.S. Renal Data System database between 1988 and 1998 by Kaplan-Meier estimates and Cox proportional hazards models to quantify the effect of ESRD time on paired cadaveric kidneys and on all cadaveric kidneys compared to living-donated kidneys.

RESULTS: Five- and 10-year unadjusted graft survival rates were significantly worse in paired kidney recipients who had undergone more than 24 months of dialysis (58% and 29%, respectively) compared to paired kidney recipients who had undergone less than 6 months of dialysis (78% and 63%, respectively; P<0.001 each). Ten-year overall adjusted graft survival for cadaveric transplants was 69% for preemptive transplants versus 39% for transplants after 24 months on dialysis. For living transplants, 10-year overall adjusted graft survival was 75% for preemptive transplants versus 49% for transplants after 24 month on dialysis.

CONCLUSIONS: ESRD time is arguably the strongest independent modifiable risk factor for renal transplant outcomes. Part of the advantage of living-donor versus cadaveric-donor transplantation may be explained by waiting time. This effect is dominant enough that a cadaveric renal transplant recipient with an ESRD time less than 6 months has the equivalent graft survival of living donor transplant recipients who wait on dialysis for more than 2 years.

7. *Meier-Kriesche, H. U., Port, F. K., Ojo, A. O., Rudich, S. M., Hanson, J. A., Cibrik, D. M., ... & Kaplan, B. (2000). Effect of waiting time on renal transplant outcome. Kidney international, 58(3), 1311-1317.*

Abstract: BACKGROUND: Numerous factors are known to impact on patient survival after renal transplantation. Recent studies have confirmed a survival advantage for renal transplant patients over those waiting on dialysis. We aimed to investigate the hypothesis that longer waiting times are more deleterious than shorter waiting times, that is, to detect a "dose effect" for waiting time.

METHODS: We analyzed 73,103 primary adult renal transplants registered at the United States Renal Data System Registry from 1988 to 1997 for the primary endpoints of death with functioning graft and death-

censored graft failure by Cox proportional hazard models. All models were corrected for donor and recipient demographics and other factors known to affect outcome after kidney transplantation.

RESULTS: A longer waiting time on dialysis is a significant risk factor for death-censored graft survival and patient death with functioning graft after renal transplantation ($P < 0.001$ each). Relative to preemptive transplants, waiting times of 6 to 12 months, 12 to 24 months, 24 to 36, 36 to 48, and over 48 months confer a 21, 28, 41, 53, and 72% increase in mortality risk after transplantation, respectively. Relative to preemptive transplants, waiting times of 0 to 6 months, 6 to 12 months, 12 to 24 months, and over 24 months confer a 17, 37, 55, and 68% increase in risk for death-censored graft loss after transplantation, respectively.

CONCLUSIONS: Longer waiting times on dialysis negatively impact on post-transplant graft and patient survival. These data strongly support the hypothesis that patients who reach end-stage renal disease should receive a renal transplant as early as possible in order to enhance their chances of long-term survival.

8. *Schold JD, Huml AM, Poggio ED et al. Patients with High Priority for Kidney Transplant Who Are Not Given Expedited Placement on the Transplant Waiting List Represent Lost Opportunities. J Am Soc Nephrol 2021;32:1733-1746.*

Abstract: BACKGROUND: Kidney transplantation is associated with the best outcomes for most patients with ESKD. The national Kidney Allocation System prioritizes patients with Estimated Post-Transplant Survival (EPTS) scores in the top 20% for expedited access to optimal deceased donor kidneys.

METHODS: We studied adults aged 18 years in the United States Renal Data System with top 20% EPTS scores who had been preemptively waitlisted or initiated dialysis in 2015–2017. We evaluated time to waitlist placement, transplantation, and mortality with unadjusted and multivariable survival models.

RESULTS: Of 42,445 patients with top 20% EPTS scores (mean age, 38.0 years; 57% male; 59% White patients, and 31% Black patients), 7922 were preemptively waitlisted. Among 34,523 patients initiating dialysis, the 3-year cumulative waitlist placement incidence was 37%. Numerous factors independently associated with waitlisting included race, income, and having noncommercial insurance. For example, waitlisting was less likely for Black versus White patients, and for patients in the lowest-income neighborhoods versus those in the highest-income neighborhoods. Among patients initiating dialysis, 61% lost their top 20% EPTS status within 30 months versus 18% of patients who were preemptively listed. The 3-year incidence of deceased and living donor transplantation was 5% and 6%, respectively, for patients who initiated dialysis and 26% and 44%, respectively, for patients who were preemptively listed.

CONCLUSIONS: Many patients with ESKD qualifying with top 20% EPTS status are not placed on the transplant waiting list in a timely manner, with significant variation on the basis of demographic and social factors. Patients who are preemptively listed are more likely to receive benefits of top 20% EPTS status. Efforts to expedite care for qualifying candidates are needed, and automated transplant referral for patients with the best prognoses should be considered.

9. *Schold J and Meier-Kreische HU. Which Renal Transplant Candidates Should Accept Marginal Kidneys in Exchange for a Shorter Waiting Time on Dialysis? Clin J Am Soc Nephrol 2006;1:532-538.*

Abstract: Renal transplantation has been established as a life-saving procedure for patients with ESRD. Deceased donor kidneys convey variable life expectancies for recipients. However, limited information is available to guide patients and patient advocates concerning the appropriateness to list for expanded criteria donations (ECD). Half-lives for wait-listed transplant candidates were estimated from the time of ESRD onset on the basis of recipient age, primary diagnosis, and organ quality using survival models. In addition, we evaluated the likelihood of candidates' receiving a transplant on the basis of age and other characteristics by duration of waiting time. Older patients (65) had longer life expectancy when they accepted an ECD within 2 yr of ESRD onset (5.6 yr) compared with waiting for a standard kidney (5.3 yr) or a living donation (5.5 yr) after 4 yr of dialysis. Conversely, younger recipients (18 to 39 yr) had longer life expectancy with a living donation (27.6 yr) or standard kidney (26.4 yr) after 4 yr on dialysis compared with an ECD after 2 yr of dialysis (17.6 yr). Increased candidate age was associated with the likelihood of not receiving a transplant during the period on the waiting list as a result of mortality and separately related to morbidity and delisting. Older and frailer

transplant candidates benefit from accepting lower quality organs early after ESRD, whereas younger and healthier patients benefit from receiving higher quality organs even with longer dialysis exposure. These findings are important for transplant candidates and advocates decision-making and for potential further implementation in allocation policy.

[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]

After applying all exclusion criteria, we evaluated the FYSWR performance scores for all dialysis practitioner group practices that had at least 11 patients and at least 2 expected events in the evaluation period 2016 through 2019. The mean value of FYSWR was 1.01. The interquartile range (Q3-Q1) is 0.77, with the bottom quartile of practitioner group practices having 46% lower, versus the top quartile having 33% higher, waitlisting or living-donor transplant rates among new dialysis patients during their first year of dialysis than the national average. Dates of data: January 1, 2016 – December 31, 2020 (inclusive of data for one year of follow-up beyond last assessment year of 2019).

Number of patients: 281,479

Number of practitioner groups: 2,168

Table 1: Descriptive statistics of FYSWR overall and by decile, 2016-2019

*	Mean	Std Dev	Minimum	Maximum	Median	Lower Quartile	Upper Quartile
Overall	1.01	0.62	0	4.85	0.92	0.56	1.33
*	*	*	*	*	*	*	*
Decile	*	*	*	*	*	*	*
1	0.13	0.12	0.00	0.31	0.14	0.00	0.25
2	0.40	0.05	0.31	0.48	0.41	0.36	0.45
3	0.56	0.05	0.48	0.63	0.56	0.52	0.60
4	0.71	0.04	0.64	0.78	0.72	0.67	0.75
5	0.86	0.04	0.78	0.92	0.86	0.82	0.89
6	0.99	0.04	0.92	1.07	1.00	0.96	1.03
7	1.15	0.05	1.07	1.24	1.14	1.10	1.19
8	1.34	0.06	1.24	1.47	1.33	1.28	1.39
9	1.62	0.10	1.47	1.82	1.59	1.53	1.70
10	2.31	0.45	1.82	4.85	2.16	1.98	2.53

Table 1: Descriptive statistics of FYSWR overall and by decile, 2016-2019

*Cell intentionally left blank.

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

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

[Response Begins]

Dates of data: January 1, 2016 – December 31, 2020 (inclusive of data for one year of follow-up beyond last assessment year of 2019).

Number of patients: 281,479

Number of practitioner groups: 2,168

Table 2: Descriptive statistics of FYSWR, by race, ethnicity and sex, 2016-2019

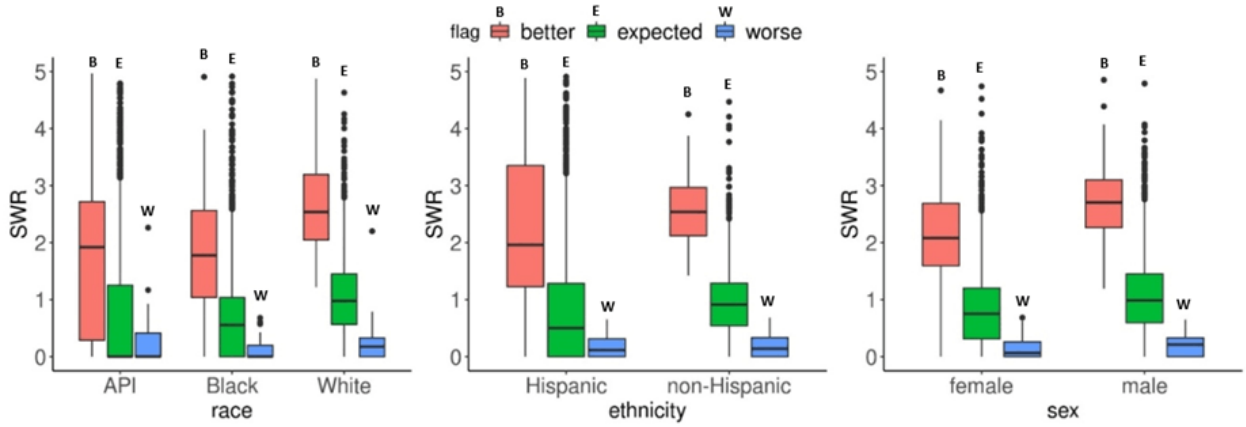
Group	Mean	Std Dev	Minimum	Maximum	Median	Lower Quartile	Upper Quartile
Race	*	*	*	*	*	*	*
White	1.13	0.94	0.00	18.21	0.99	0.55	1.52
Black	1.05	4.97	0.00	195.92	0.58	0.00	1.13
Asian Pacific Islander	2.04	9.52	0.00	286.21	0.00	0.00	1.87
Native American/ Alaskan Native	1.89	12.32	0.00	176.88	0.00	0.00	0.00
Other	2.88	17.33	0.00	296.55	0.00	0.00	0.00
*	*	*	*	*	*	*	*
Ethnicity	*	*	*	*	*	*	*
Hispanic	1.48	13.48	0.00	584.30	0.56	0.00	1.43

Group	Mean	Std Dev	Minimum	Maximum	Median	Lower Quartile	Upper Quartile
Non-Hispanic	1.09	3.49	0.00	158.03	0.93	0.53	1.35
*	*	*	*	*	*	*	*
Sex	*	*	*	*	*	*	*
Female	0.87	0.78	0.00	5.60	0.76	0.30	1.25
Male	1.12	0.79	0.00	5.73	1.00	0.58	1.50

Table 2: Descriptive statistics of FYSWR, by race, ethnicity and sex, 2016-2019

*Cell intentionally left blank.

Figure 1: Performance of FYSWR, by race, ethnicity and sex, 2016-2019



Note: Race groups Native American/Alaskan Native and Other have only small number of patients and were not included in Figure 1.

The data presented in Table 2 and Figure 1 above demonstrate wide variation and performance gaps within strata of race, ethnicity and sex categories.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

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

sp.01. Provide the measure title.

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

[Response Begins]

First Year Standardized Waitlist Ratio (FYSWR)

[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]

The FYSWR measure tracks the number of incident patients in a practitioner (inclusive of physicians and advanced practice providers) group who are under the age of 75 and were listed on the kidney or kidney-pancreas transplant waitlist or received a living donor transplant within the first year of initiating dialysis. For each practitioner group, the First Year Standardized Waitlist Ratio (FYSWR) is calculated to compare the observed number of waitlist events in a practitioner group to its expected number of waitlist events. The FYSWR uses the expected waitlist events calculated from a Cox model, adjusted for age and patient comorbidities at incidence of dialysis. For this measure, patients are assigned to the practitioner group based on the National Provider Identifier (NPI)/Unique Physician Identifier Number (UPIN) information entered on the CMS Medical Evidence 2728 form.

[Response Ends]

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

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

Please do not select:

- *Surgery: General*

[Response Begins]

Renal

Renal: End Stage Renal Disease (ESRD)

[Response Ends]

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

[Response Begins]

Care Coordination

Disparities Sensitive

[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)

Children (Age < 18)

[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

[Response Ends]

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

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

[Response Begins]

Outpatient Services

[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

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]

Number of patients in the practitioner group listed on the kidney or kidney-pancreas transplant waitlist or who received living donor transplants within the first year following initiation of dialysis.

[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

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]

Organ Procurement and Transplant Network (OPTN) Kidney or Kidney-Pancreas waitlist or transplant dates populated and within one year of first dialysis date per Standardized Analysis Files (SAF).

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

The denominator for the FYSWR is the expected number of waitlist or living donor transplant events in the practitioner group according to each patient's treatment history for patients within the first year following initiation of dialysis, adjusted for age, incident comorbidities, dual Medicare-Medicaid eligibility, Area Deprivation Index (from patient's residence zip code) and transplant center characteristics, among patients under 75 years of age who were not already waitlisted and did not have kidney transplantation prior to the initiation of ESRD dialysis.

[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

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]

The practitioner groups were identified based on the attending physician identified on the CMS ESRD Medical Evidence Report (CMS Form 2728). The attending practitioners were linked to the IDR Medicare Virtual Data Mart (VDM) Part B Provider Extract File view table (view name: V2_MDCR_PRVDR_PBX, view database: $SYS_VDM_VIEW_MDCR_$ENVNAME) to identify the most current, active practitioner groups.

Information regarding first ESRD service date, modality, death, waitlist status, and transplant are obtained from Medicare claims, EQRS (formerly CROWNWeb), Organ Procurement and Transplant Network (OPTN), and the Social Security Death Master File.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Patients who were at age 75 or older on their initiation of dialysis date are excluded. Patients who were admitted to a skilled nursing home facility (SNF) or a hospice during the month of evaluation were excluded. These exclusions represent conditions for which transplant waitlist candidacy is highly unlikely, and which can be identified readily with available data. Patients were also excluded if waitlisted or transplanted prior to initiation of first dialysis. Patients who were attributed to dialysis practitioner groups with fewer than 11 patients or 2 expected events are not excluded from the measure. All patients who meet the denominator inclusion criteria are included and used to model a given dialysis practitioner group's expected waitlist rate. If a dialysis practitioner group has fewer than 11 patients or 2 expected events, then the dialysis practitioner group is excluded from reporting outcomes.

[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]

The Nursing Home Minimum Dataset (MDS) and the Questions 17u and 22 on CMS Medical Evidence Form 2728 were used to identify patients in skilled nursing facilities. For hospice patients, a separate CMS file that contains final action claims submitted by Hospice providers was used to determine the hospice status. Patients are excluded if they are nursing home patients according to their Medical Evidence Form 2728 or according to the Minimum Dataset (MDS) data on their initiation of dialysis date. Patients with Medicare Hospice claims on their initiation of dialysis date are also excluded. Patients that were on the kidney or kidney-pancreas waitlist or had a transplant prior to initiation of dialysis were excluded.

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

sp.19. Select the risk adjustment type.

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

[Response Begins]

Statistical risk model

[Response Ends]

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

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

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

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

[Response Begins]

Better quality = Higher score

[Response Ends]

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

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

[Response Begins]

See attached flowchart.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

Claims

Registry Data

[Response Ends]

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

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

[Response Begins]

EQRS (formerly CROWNWeb), Medicare Claims, and the CMS Medical Evidence Form 2728 were used as the data source for establishing the denominator. CMS Medical Evidence Form 2728 was used for the age risk adjustment and exclusion of patients age 75 or older, and comorbidity condition adjustments. Organ Procurement and Transplant Network (OPTN) is the data source for the numerator (waitlisting or living donor kidney transplantation). Medicare claims were used for the hospice exclusion criteria. The Nursing Home Minimum Dataset and Questions 16u and 21 on the CMS Medical Evidence Form were used to identify SNF patients. Additionally, Medicare claims and a payment history file were used to determine dual eligibility status. The Medicare Provider Files from the CMS Integrated Data Repository (IDR) were used to identify practitioner's group practice. Area Deprivation Index (ADI) was obtained from Census data (2011-2015) based on patient zip code. In order to assess the transplant center characteristics, Scientific Registry of Transplant Recipients (SRTR) data was used.

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

2a. Reliability

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

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 are derived from a combination of EQRS (formerly CROWNWeb), CMS Medical Evidence Form (CMS Form 2728), the Nursing Home Minimum Dataset, transplant registries (OPTN, SRTR), and Medicare claims from CMS.

[Response Ends]

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

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

[Response Begins]

01-01-2016 – 12-31-2020 (allowing for one year of follow-up beyond the last year of performance assessment in 2019)

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Clinician: Group/Practice

[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]

Over the reporting period from 2016 through 2019, there were 2,168 practitioner groups included in these analyses, after restricting to practitioner group practices that had at least 11 eligible patients and at least 2 expected events.

[Response Ends]

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

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

[Response Begins]

There were 281,479 patients in total. The average age at their initiation of dialysis was 57.9 years old, 40.9% were female, 63.3% were White, 29.7% were Black, 5.5% were Asian/Pacific Islander, 1.2% were American Indian/Alaskan Native, 0.3% were Other/Multi-racial/Unknown and 17.7% were Hispanic.

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

Patient level:

- Sex (we acknowledge that sex is less recognized as a social risk factor but it is being increasingly considered as such especially given its relationship to gender [see for example, O'Neil et al. Gender/Sex as a social determinant of cardiovascular risk. Circulation 2018;137:854], and have therefore chosen to include an assessment of it in our analysis)
- Race
- Ethnicity
- Medicare-Medicaid dual eligibility

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

Zipcode level – ADI from 2015 Census data.

[Response Ends]

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

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

Choose one or both levels.

[Response Begins]

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

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

We used data from the start of 2016 through the end of 2019 to calculate the First Year Standardized Waitlist Ratio. Our approach for determining measure reliability aligns with one-way analysis of variance (ANOVA), in which the between dialysis practitioner group practice variation (σ_b^2)

and the within- dialysis practitioner group practice variation ($\sigma_{t,w}^2$) in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of the measure (i.e., $\sigma_b^2 + \sigma_{t,w}^2$) that is attributed to the between – dialysis practitioner group practice variation, the true signal that reflects the differences across dialysis practitioner group practices. We assessed reliability by calculating inter-unit reliability (IUR) for the annual performance scores. If the measure were an average of the individuals' measurements under the care of one dialysis practitioner group practice, the usual ANOVA approach would be used. The yearly based measure, however, is not a simple average and we instead estimate the IUR using a bootstrap approach, which uses a resampling scheme to estimate the within dialysis practitioner group practice variation that cannot be directly estimated by ANOVA. A small IUR (near 0) reveals that most of the variation of the measures between dialysis practitioner group practices is driven by random noise, indicating the measure would not be a good characterization of the differences among dialysis practitioner group practices. A large IUR (near 1) indicates that most of the variation between dialysis practitioner group practices is due to true differences between dialysis practitioner group practices.

Below is our approach to calculate IUR.

Let $T_1,\ldots,T_N$ be the FYSWRs for N practitioner groups. Since the variation in $T_1,\ldots,T_N$ is mainly driven by the estimates of dialysis practitioner group practice-specific intercepts ($\alpha_1,\ldots,\alpha_N$), we use their asymptotic distributions to estimate the within-dialysis practitioner group practice variation in FYSWRs. Applying the delta method, we estimate the variance of T_i and denote the estimate as S_i^2. Calling on formulas from the one-way ANOVA, the within-dialysis practitioner group practice variance in FYSWRs can be estimated by

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

and the total variation in FYSWRs can be estimated by

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

where n_i is the number of subjects in the ith practitioner group, $\bar{T} = \sum n_i T_i \Big/ \sum n_i$, and

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

is approximately the average dialysis practitioner group practice size (number of patients per dialysis practitioner group practice). Thus, the $IUR = \sigma_b^2/(\sigma_b^2 + \sigma_{t,w}^2)$ can be estimated by $(s_t^2 - s_{t,w}^2)\Big/ s_t^2$.

The reliability of FYSWR calculation only included dialysis practitioner group practices with at least 11 patients and at least 2 expected events during the entire year.

[Response Ends]

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

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

[Response Begins]

The IUR is 0.64. Dialysis practitioner group practices with < 11 eligible patients and < 2 expected events were excluded from this calculation.

[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 value of IUR indicates that about 64% of the variation in the FYSWR measure can be attributed to the between-dialysis practitioner group practice differences (signal) and about 36% of variation to within-dialysis practitioner group practice variation (noise). The value of IUR implies a moderate degree of reliability.

[Response Ends]

2b. Validity

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

[Response Begins]

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

Empirical validity testing

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

Validity of the measure was tested by evaluating the association between the dialysis practitioner group level measure performance, and subsequent mortality and overall transplant rates among all patients attributed to the practitioner groups. We hypothesized that practitioner groups with higher performance on the FYSWR measure would have subsequently higher transplant rates among their patients. This would be expected to follow from activities these practitioner groups conducted to improve the health and therefore suitability of their patients for transplant candidacy. Along similar lines, we hypothesized that practitioner groups with higher performance on the FYSWR measure would demonstrate lower subsequent mortality among their patients. However, we expected this to be a more modest association given the many other factors that can affect mortality within the dialysis population.

To evaluate the associations, we first divided dialysis practitioner groups into 3 tertiles (T1 to T3) based on their performance on the FYSWR (T1 to T3, from highest to lowest waitlisting). Tertiles were chosen in order to evaluate a gradient in effect, but still maintain sufficient numbers within each group for statistical precision. We then computed the corresponding second year mortality rate and transplant rate among patients assigned to each practitioner group. We then applied the Cochran-Armitage trend test to evaluate the relationship between the tertile grouping and these practitioner group-level outcomes. Finally, we examined the Spearman correlations between FYSWR and the second year mortality rate or second year transplant rate.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

The tertile groups based on the performance scores were defined as:

T_1 (best performance): 1.17 - 4.85

T_2: 0.69 - 1.17

T_3 (worst performance): 0 - 0.69

The dialysis practitioner group average second year mortality is 15.3, 15.7, 15.9 deaths per 100 patient-years for T_1, T_2, T_3 groups, respectively (trend test p=0.0607). The Spearman correlation coefficient is: -0.02 (p=0.3151).

The dialysis practitioner group average second year transplant rate is 4.7, 3.2, 1.8 transplants per 100 patient-years for the T_1, T_2, T_3 groups, respectively (trend test p<.01). The Spearman correlation coefficient is: 0.32 (p<.01).

[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]

As expected, higher FYSWR performance correlated with higher second year transplant rate, with clear separation of transplant rates across practitioner group tertiles of performance. The direction of the relationship with mortality was as expected, with numerically lower mortality with higher performance on the FYSWR measure, though it did not achieve statistical significance.

[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]

Dialysis practitioner groups were classified as 'As Expected,' 'Better than Expected', or 'Worse than Expected' based on whether observed and expected values are statistically different at the 5% level. Average values of FYSWR between these groups are listed to determine if there are practically meaningful differences in performance scores. Specifically, the p-value is computed using a Poisson approximation under which the distribution of the observed number, O, in the dialysis practitioner group is Poisson with a mean value equal to the expected number, E, computed from the Cox model. Accordingly, if the observed number, O, is greater than E, then the mid p-value = Pr(X>=O)+ Pr(X>O) where X has a Poisson distribution with mean E. Similarly, if O<E, the mid p-value = Pr(X <=O)+ Pr(X<O) where X has a Poisson distribution with mean E. To address the problem of simultaneously monitoring a large number of dialysis practitioner groups and to take account of the intrinsic unexplained variation among practitioner groups, we used the approach described in Kalbfleisch and Wolfe (see full citation below). Specifically, to implement this method, the p-value for each dialysis practitioner group is converted to a Z-score, stratified into four groups based on patient-years within each practitioner group. Within each group, using robust estimates of location and scale based on the normal curve fitted to the center of the z-scores, we derive the mean and variance of a normal empirical null distribution. This empirical null distribution is then used to calculate the p-value for each dialysis practitioner. Finally, dialysis practitioner group practices are flagged if they have outcomes that are extreme when compared to the variation in the national waitlist rate. This method aims to separate underlying intrinsic variation in dialysis practitioner group outcomes from variation that might be attributed to poor (or excellent) care.

Reference:

Kalbfleisch, J. and Wolfe, R. (2013). On monitoring outcomes of medical providers. Statis- tics in Biosciences, 5(2):286–302.

[Response Ends]

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

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Table 3: Count (%) of dialysis practitioner group practices and median FYSWR, stratified by classification category.

Classification Category	Count	Percent	Median FYSWR
Better Than Expected	81	4%	2.59
As Expected	2,033	94%	0.97
Worse Than Expected	54	2%	0.19
Total	2,168	100%	0.92

Table 3: Count (%) of dialysis practitioner group practices and median FYSWR, stratified by classification category.

[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]

Four percent of dialysis practitioner group practices were classified as better than expected and 2% as worse than expected. Better than expected physician group on average have observed waitlist/living donor transplant rates more than double that of expected waitlist/transplant rates while worse than expected dialysis practitioner group practices had observed rates less than 1/5 what was expected. These differences are therefore both practically meaningful and statistically significant.

[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]

Many data elements can be obtained from multiple sources and missing data occurs rarely for variables included in this measure. One element with missingness is with respect to assignment of dialysis practitioner groups, which occurs for one of following two reasons: 1) some patients could not be assigned to a dialysis practitioner group due to missing National Provider Identifier (NPI)/Unique Physician Identifier Number (UPIN) information on the CMS-2728 form, or 2) because the NPI/UPIN could not be matched with the most current and active practitioner group from the IDR provider table.

[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]

Table 4: Distribution of missing data among 281,479 patients in the performance assessment period 2016-2019

*	Count	Percent
Patients with no practitioner NPI on 2728, or without a match to the most current and active practitioner group from the IDR provider table	17,654	6.2

Table 4: Distribution of missing data among 281,479 patients in the performance assessment period 2016-2019

*This cell is intentionally left blank.

[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]

Only 6.2% of patients are missing a dialysis practitioner group. Patients with missing dialysis practitioner assignment were aggregated into their own group and also included in the statistical model used to obtain the performance scores. The FYSWR was 1.05 for patients in the missing dialysis practitioner group, which was not statistically significant from the average score, suggesting that these missing patients have similar waitlisting experience to the average.

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.

[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]

In order to evaluate the exclusion criteria, the differences in the number of patients with and without excluding age >= 75, nursing home patients, and hospice patients, were compared. We show the frequency of patients excluded due to each criteria. Additionally, we compared the performance scores before and after exclusions. We do not exclude patients from dialysis practitioner groups with fewer than 11 attributed patients or 2 expected events.

[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 5: Overall number and percentage of patients excluded

*	**Before age, nursing home, and hospice exclusion**	**After age, nursing home, and hospice exclusion**	**Percentage excluded**
Number of patients	410,849	281,479	31.5%

Table 5: Overall number and percentage of patients excluded

*This cell is intentionally left blank.

Table 6: Frequency distribution of patient-months excluded based on each exclusion criteria

Variable excluded	**Frequency (%)**
Age >=75	101,658 (24.7)
Nursing Home from CMS-2728	18,178 (4.4)
Nursing home from Nursing home history file	9,390 (2.3)
Hospice	144 (0.04)

Table 6: Frequency distribution of patient-months excluded based on each exclusion criteria

Table 7: Distribution of performance scores before and after exclusion

FYSWR	**Mean**	**Standard Deviation**	**Minimum**	**Q1**	**Median**	**Q3**	**Maximum**
Before exclusion	1.006	0.622	0	0.563	0.924	1.330	5.238
After exclusion	1.007	0.623	0	0.560	0.922	1.331	4.846

Table 7: Distribution of performance scores before and after exclusion

Figure 2: Distribution of FYSWR before exclusions

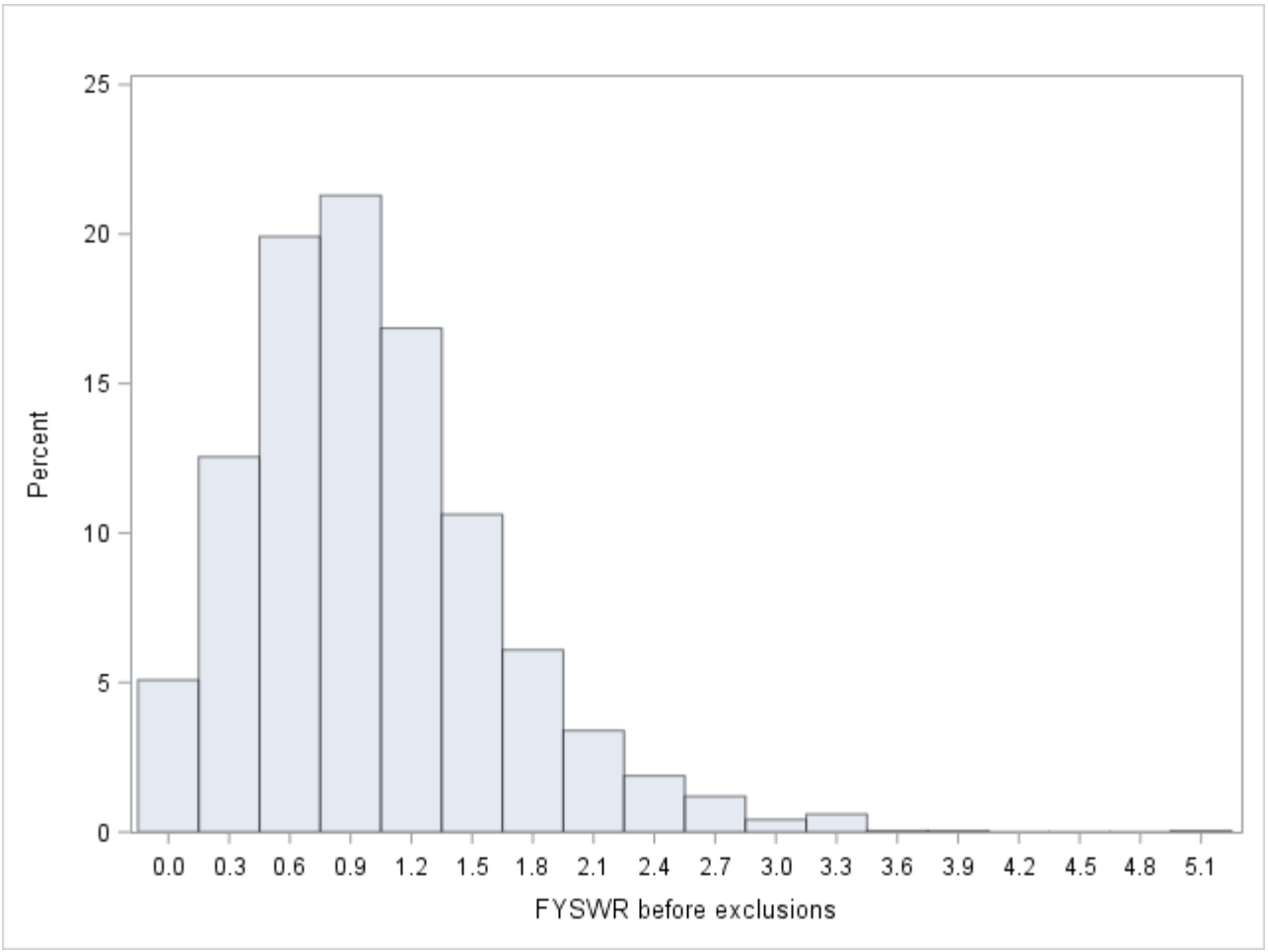


Figure 3: Distribution of FYSWR after exclusions

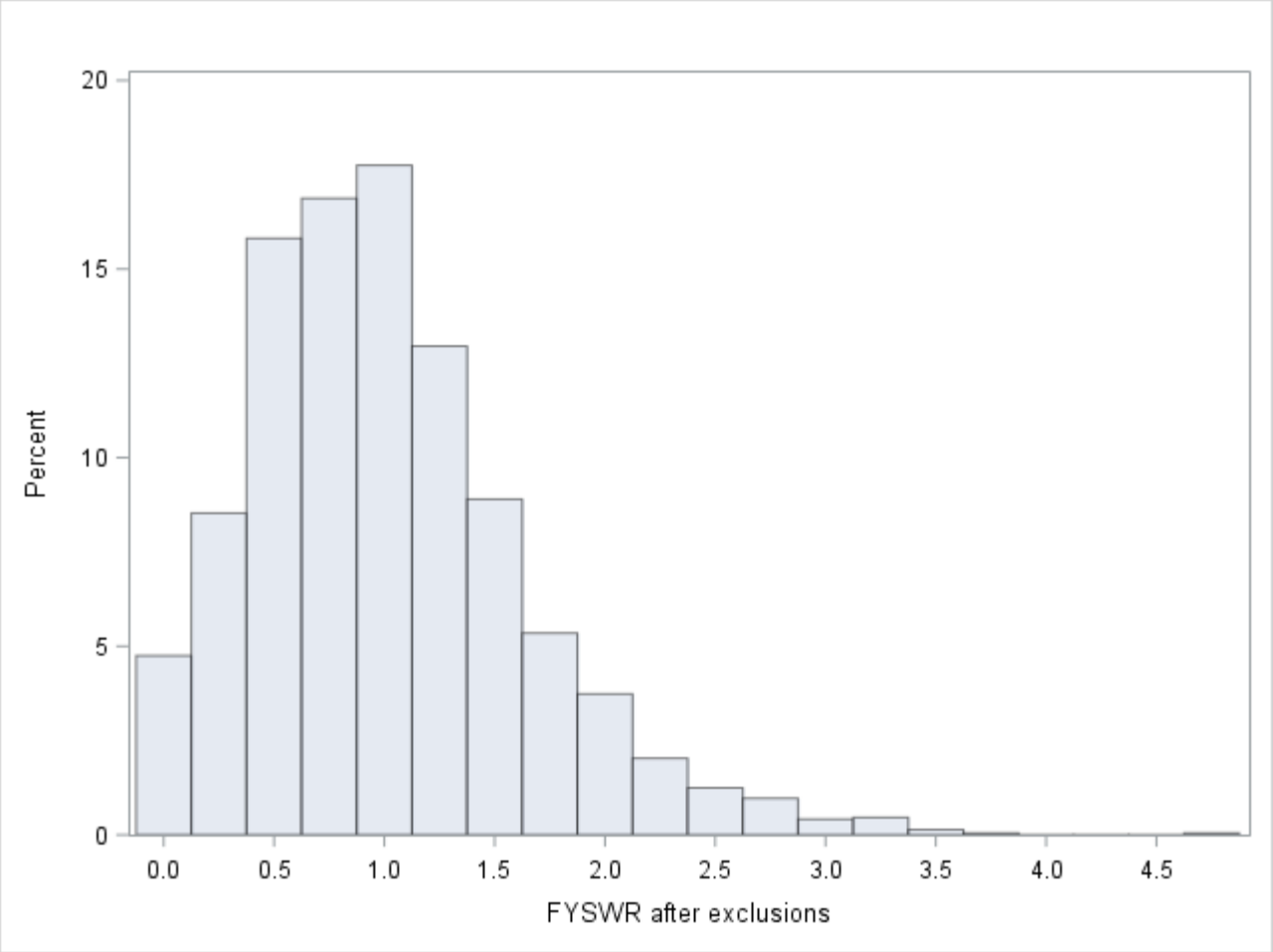
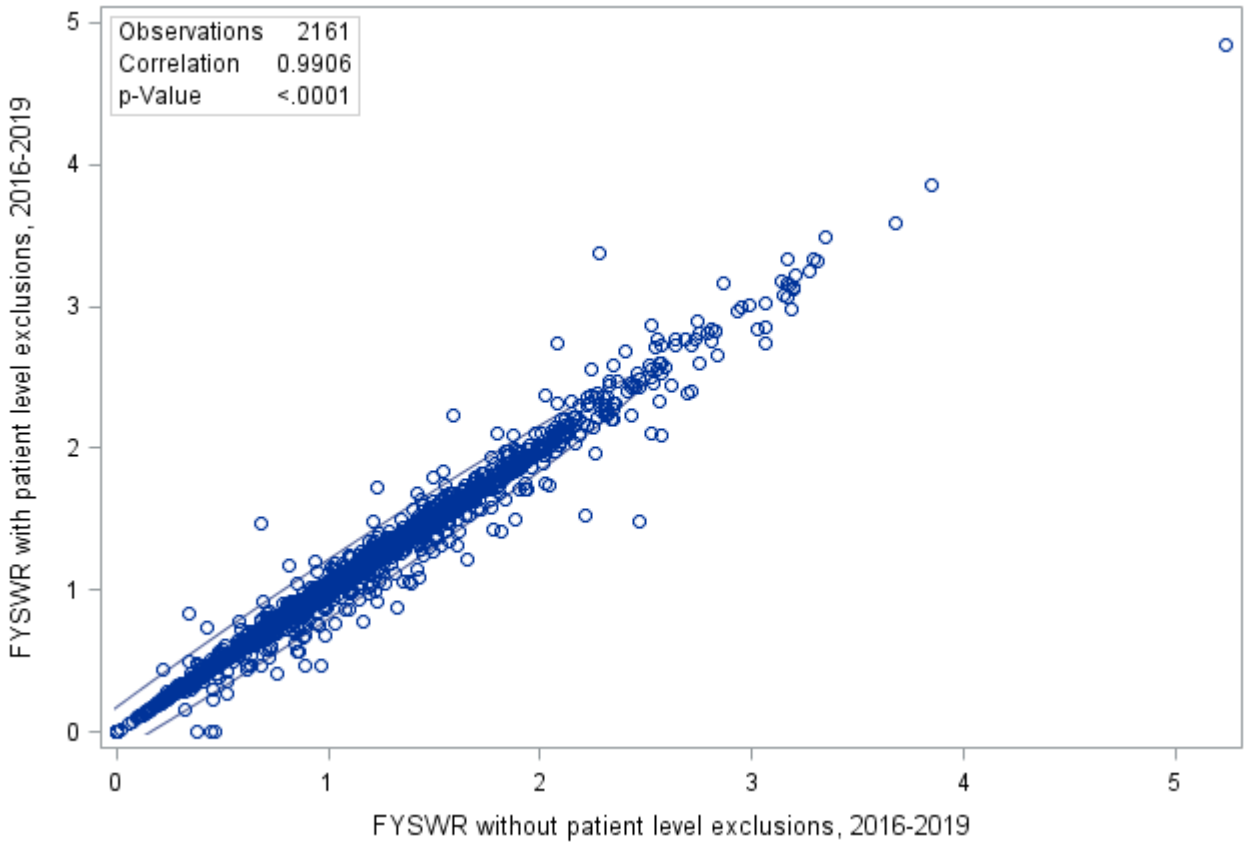


Figure 4: Scatterplot of FYSWR with and without exclusions



The correlation coefficient is 0.9906 (p<.0001).

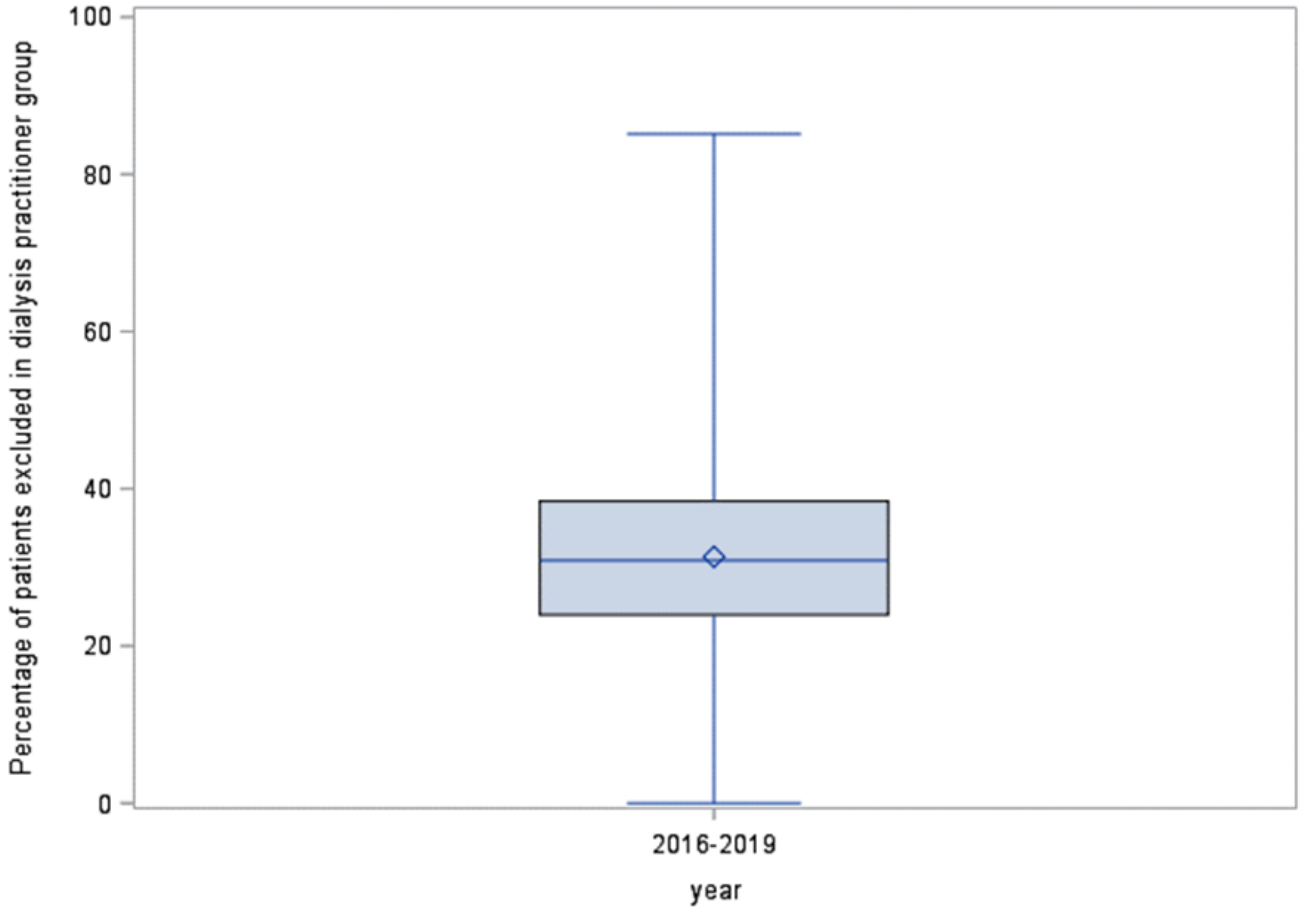
Table 8: Comparison of performance scores with and without excluded patients

*	*	**FYSWR without patient-level exclusion**	**FYSWR without patient-level exclusion**	**FYSWR without patient-level exclusion**	**FYSWR without patient-level exclusion**
*	*	**Better than Expected**	**As Expected**	**Worse than Expected**	**Total**
FYSWR with patient-level exclusion	**Better than Expected**	66	15	0	81
FYSWR with patient-level exclusion	**As Expected**	7	2,005	14	2,026
FYSWR with patient-level exclusion	**Worse than Expected**	0	4	50	54
FYSWR with patient-level exclusion	**Total**	73	2,024	64	2,161

Table 8: Comparison of performance scores with and without excluded patients

*This cell is intentionally left blank.

Figure 5: Distribution of excluded patients (%) across dialysis practitioner group practices



[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]

Although performance scores are modestly affected by the exclusions (table 7-8, and figures 2-4), the exclusions are deemed important on clinical grounds as they represent a group of patients highly unlikely to be suitable for transplant waitlisting. Furthermore, there is a fair degree of variation in the percentage of patients excluded across practitioner groups, as shown in Figure 5. Finally, as the data to determine the exclusions is readily available, there is minimal additional burden for analysis anticipated by using these exclusion criteria.

[Response Ends]

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

[Response Begins]

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

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

See Table 9 below.

[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]

Table 9. Patient characteristics included in the model as covariate variables and the data source.

Variable	Data Source	Notes
Age	EQRS (formerly CROWNWeb)/SAF	Age of Patient
Age Spline 12	EQRS (formerly CROWNWeb)/SAF	Changes Slope of Age Coefficient at 12
Age Spline 18	EQRS (formerly CROWNWeb)/SAF	Changes Slope of Age Coefficient at 18
Age Spline 64	EQRS (formerly CROWNWeb)/SAF	Changes Slope of Age Coefficient at 64
Heart Disease	CMS Form 2728 Box 16	Binary Variable for Atherosclerotic Heart Disease, Other Cardiac Disease, Cardiac Failure
Do not Ambulate	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
COPD	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Do not Transfer	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Cancer	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Peripheral Vascular Disease	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Cerebral Vascular Accident	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Drug Use	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Amputation	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Assistance with Daily Activities	CMS Form 2728 Box 16	Binary Variable '1' has disease, '0' does not
Area Deprivation Index (ADI)	U.S. Census Data	Continuous 0 to 1 Index; Higher Values Indicate Higher Neighborhood Disadvantage

Variable	Data Source	Notes
Medicare Dual Eligible	CMS Form 2728 Box 11/ EQRS (formerly CROWNWeb) Enrollment Database	Binary Variable '1' is Medicare and Medicaid at Incidence, '0' is not
Weighted SRTR mortality ratio	SRTR Program-Specific Reports	Ratio is weighted by the percentage of dead patients in each transplant center in the same patient residential area
Weighted SRTR transplant ratio	SRTR Program-Specific Reports	Ratio is weighted by the percentage of waitlisted patients placed on each transplant center in the same patient residential area

Table 9. Patient characteristics included in the model as covariate variables and the data source.

For the measure outcome, the event was defined as waitlisting or living-donor transplantation. Time zero was defined as the first initiation of dialysis. Patients were followed until waitlisting, living donor transplantation, death, or one-year anniversary since first dialysis (i.e., the earliest thereof). A two-stage Cox model was fitted to calculate the expected number of events. At the first stage, a Cox model stratified on dialysis practitioner group practices was fitted in order to obtain an estimate of the age, comorbidities, and transplant center effects (unconfounded by dialysis practitioner group practices) to be used as an offset. At the second stage, a national average baseline hazard was estimated. The national average baseline (from the second stage), age, comorbidities, and transplant center adjustments (from the first stage) were then used to compute the probability of an event for each patient, followed by the total expected number of events at each dialysis practitioner group practice.

Let p denote the number of patient characteristics in the model and X_{ij}be the specific value of the j^{th} characteristic for the i^{th} patient-record. At the first stage, for patient-record *i*, we denote the measured characteristics or covariates as $X_i = (\chi_{i1}, \chi_{i2}, \ldots, \chi_{ip},)$ and use this to define the regression portion of a Cox model in which dialysis practitioner group practices define the strata. Note that for a categorical characteristic, the X_{ij} value is 1 if the patient falls into the category and 0 otherwise. The output of the first stage is a set of regression coefficients, $\beta_1, \beta_2, \ldots, \beta_p$ and the corresponding predicted value for the i^{th} patient-record is given by

$$\chi_i\beta = (\beta_{1\chi_{i1}} + \beta_{2\chi_{i2}} + \ldots + \beta_{p\chi_{ip}}) \text{ (1)}$$

At the second stage, the relative risk estimates from the first stage were used as an offset, without stratification. After the second stage, the linear prediction is

$$A_i = \beta_{0\chi_{i0}} + X_i\beta = \beta_{0\chi_{i0}} + \beta_{1\chi_{i1}} + \beta_{2\chi_{i2}} + \ldots + \beta_{p\chi_{ip}} \text{ (2)}$$

Suppose that t_i is the end of follow-up time for patient-record i, so that $S_0(t_i)$ is the baseline survival probability at time t_i. The survival probability for this patient-record i at time t_i is:

$$S_i(t_i) = [\mathrm{S}_0(t_i)]^{exp(A_i)}. \text{ (3)}$$

The expected number of waitlistings for this patient-record during follow-up time t_i arises from considerations in the Cox model and can be written as

$$-\ln(S_i(t_i)) = -e^{A_i}\ln(S_0(t_i)). \text{ (4)}$$

The expected number of waitlistings at a given dialysis practitioner group practice can now be computed simply by summing these expected values over the totality of patient-records at that dialysis practitioner group practices.

Specifically, the expected value is the sum over the N patient-records at the dialysis practitioner group practices giving

$$E = \sum_{i=1}^{N} - \ln(S_i(t_i)) = -\sum_{i=1}^{N} e^{A_i} \ln(S_0(t_i)). \quad (5)$$

Let O be the total number of waitlisting observed at the dialysis practitioner group practice during the total four years follow-up period. As stated above, the FYSWR is the ratio of the total number of observed waitlisting to the expected number

FYSWR= O/E. (6)

[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]

[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]

Variables chosen for inclusion in the model were based on a conceptual rationale that included theoretical/clinical considerations (discussed for each set of factors below) and existing literature (see brief list of references including large national or regional datasets, and clinical practice guidelines for kidney transplant candidate evaluation), for factors affecting kidney transplant waitlisting. We considered variables in three categories: social risk, functional risk, and medical/clinical risk. Choices were also discussed with a Technical Expert Panel held in 2021.

Social Risk Factors:

Under conceptual considerations, and as supported by the TEP, it was deemed important to adjust for social risk on the basis that it could affect suitability for transplant waitlisting. This could occur, for example, through difficulty with ability to pay for transplant immunosuppression medications, or lacking the resources to travel

to a transplant center for care, which are considerations taken into account for suitability for transplant waitlisting. For this purpose, dual Medicare-Medicaid eligibility (at the patient level, representing socioeconomic disadvantage) and Area Deprivation Index (ADI) were investigated and included in our model. Dual eligibility was obtained from Medicare claims and could also be obtained from the CMS-2728 form for incident patients within the first year of ESRD. ADI was obtained based on patient zip code of residence and used as a proxy to adjust for potential differences in waitlisting for neighborhoods of different ranking of socioeconomic disadvantage (see Patzer et al reference below).

Functional Risk Factors:

Given that poor functional status and frailty are associated with worse outcomes following kidney transplantation (see McAdams-Demarco et al, below), patients with low functional status may be less appropriate for waitlisting. We therefore included items available on the CMS Form 2728, indicating whether Assistance with Daily Activities is needed, Inability to transfer, and Inability to ambulate.

Clinical/Medical Risk Factors:

Age adjustment was deemed necessary on clinical grounds and supported by the Technical Expert Panel (TEP) held in 2021. Although age alone is not a contraindication to transplantation, older patients are likely to have more comorbidities and be generally more frail thus making them potentially less suitable candidates for transplantation. This may affect waitlisting rates for dialysis provider group practices with a substantially older age composition than the average. A linear spline was used to model the effect of (continuous) age. Knot placements were determined empirically based on a preliminary model that categorized age.

In addition, incident comorbidities were selected for adjustment into the FYSWR model based on demonstration of a higher associated mortality (hazard ratio above 1.0) and statistical significance (p-value <0.01) in a first year mortality model, thus reflecting patients at higher risk of early mortality and therefore potentially unsuitable for transplant waitlisting.

Finally, the TEP deemed it important to adjust for elements affecting waitlisting that may be partially outside control of dialysis practitioners, such as transplant center behavior. Two transplant center characteristics were chosen for adjustment in the model, including transplant center waitlist mortality rate, and transplant center transplant rate. The former is a reflection in part of transplant center criteria for waitlisting, as centers with more liberal criteria (i.e. less selective) will tend to accept sicker patients and therefore have higher waitlist mortality, whereas centers with more restrictive criteria will tend to have lower waitlist mortality rates. The transplant center transplant rate reflects both local organ availability and center behavior with regards to how quickly they are able to transplant waitlisted patients (e.g. by aggressively pursuing living donation). For adjustment in the model, patients were assigned to transplant center based on historical waitlisting patterns in their zip code of residence.

References:

1. Jesse D. Schold, Sumit Mohan, Anne Huml, Laura D. Buccini, John R. Sedor, Joshua J. Augustine and Emilio D. Poggio. Failure to Advance Access to Kidney Transplantation over Two Decades in the United States. JASN 2021;32:913

Abstract:

Background: Extensive research and policies have been developed to improve access to kidney transplantation among patients with ESKD. Despite this, wide variation in transplant referral rates exists between dialysis facilities.

Methods: To evaluate the longitudinal pattern of access to kidney transplantation over the past two decades, we conducted a retrospective cohort study of adult patients with ESKD initiating ESKD or placed on a transplant waiting list from 1997 to 2016 in the United States Renal Data System. We used cumulative incidence models accounting for competing risks and multivariable Cox models to evaluate time to waiting list placement or transplantation (WLT) from ESKD onset.

Results: Among the study population of 1,309,998 adult patients, cumulative 4-year WLT was 29.7%, which was unchanged over five eras. Preemptive WLT (prior to dialysis) increased by era (5.2% in 1997–2000 to 9.8% in 2013–2016), as did 4-year WLT incidence among patients aged 60–70 (13.4% in 1997–2000 to 19.8% in 2013–2016). Four-year WLT incidence diminished among patients aged 18–39 (55.8%–48.8%). Incidence of WLT was substantially lower among patients in lower-income communities, with no improvement over time. Likelihood of WLT after dialysis significantly declined over time (adjusted hazard ratio, 0.80; 95% confidence interval, 0.79 to 0.82) in 2013–2016 relative to 1997–2000.

Conclusions: Despite wide recognition, policy reforms, and extensive research, rates of WLT following ESKD onset did not seem to improve in more than two decades and were consistently reduced among vulnerable populations. Improving access to transplantation may require more substantial interventions.

2. Jesse D. Schold, Jon A. Gregg, Jeffrey S. Harman,‡ Allyson G. Hall, Pamela R. Patton, and Herwig-Ulf Meier-Kriesche. Barriers to Evaluation and Wait Listing for Kidney Transplantation. CJASN 2011;6:1760.

Abstract:

Background and objectives: Many factors have been shown to be associated with ESRD patient placement on the waiting list and receipt of kidney transplantation. Our study aim was to evaluate factors and assess the interplay of patient characteristics associated with progression to transplantation in a large cohort of referred patients from a single institution.

Design, setting, participants, & measurements: We examined 3029 consecutive adult patients referred for transplantation from 2003 to 2008. Uni- and multivariable logistic models were used to assess factors associated with progress to transplantation including receipt of evaluations, waiting list placement, and receipt of a transplant.

Results: A total of 56%, 27%, and 17% of referred patients were evaluated, were placed on the waiting list, and received a transplant over the study period, respectively. Older age, lower median income, and noncommercial insurance were associated with decreased likelihood to ascend steps to receive a transplant. There was no difference in the proportion of evaluations between African Americans (57%) and Caucasians (56%). Age-adjusted differences in waiting list placement by race were attenuated with further adjustment for income and insurance. There was no difference in the likelihood of waiting list placement between African Americans and Caucasians with commercial insurance.

Conclusions: Race/ethnicity, age, insurance status, and income are predominant factors associated with patient progress to transplantation. Disparities by race/ethnicity may be largely explained by insurance status and income, potentially suggesting that variable insurance coverage exacerbates disparities in access to transplantation in the ESRD population, despite Medicare entitlement.

3. Rachel E. Patzer, Sandra Amaral, Haimanot Wasse, Nataliya Volkova, David Kleinbaum, and William M. McClellan. Neighborhood Poverty and Racial Disparities in Kidney Transplant Waitlisting. JASN 2009;20:1333.

Abstract:

Racial disparities persist in the United States renal transplantation process. Previous studies suggest that the distance between a patient's residence and the transplant facility may associate with disparities in transplant waitlisting. We examined this possibility in a cohort study using data for incident, adult ESRD patients (1998 to 2002) from the ESRD Network 6, which includes Georgia, North Carolina, and South Carolina. We linked data with the United Network for Organ Sharing (UNOS) transplant registry through 2005 and with the 2000 U.S. Census geographic data. Of the 35,346 subjects included in the analysis, 12% were waitlisted, 57% were black, 50% were men, 20% were impoverished, 45% had diabetes as the primary etiology of ESRD, and 73% had two or more comorbidities. The median distance from patient residence to the nearest transplant center was 48 mi. After controlling for multiple covariates, distance from patient residence to transplant center did not predict placement on the transplant waitlist. In contrast, race, neighborhood poverty, gender, age, diabetes, hypertension, body mass index, albumin, and the use of erythropoietin at dialysis initiation was associated with waitlisting. As neighborhood poverty increased, the likelihood of waitlisting decreased for blacks compared with whites in each poverty category; in the poorest neighborhoods, blacks were 57% less likely to

be waitlisted than whites. This study suggests that improving the allocation of kidneys may require a focus on poor communities.

4. Mara A. McAdams-DeMarco, Andrew Law, Megan L. Salter, Eric Chow, Morgan Grams, Jeremy Walston, and Dorry L. Segev. Frailty and Early Hospital Readmission after Kidney Transplantation. American Journal of Transplantation 2013;13:2089.

Abstract:

Early hospital readmission (EHR) after kidney transplantation (KT) is associated with increased morbidity and higher costs. Registry-based recipient, transplant, and center-level predictors of EHR are limited, and novel predictors are needed. We hypothesized that frailty, a measure of physiologic reserve initially described and validated in geriatrics and recently associated with early KT outcomes, might serve as a novel, independent predictor of EHR in KT recipients of all ages. We measured frailty in 383 KT recipients at Johns Hopkins Hospital. EHR was ascertained from medical records as ≥1 hospitalization within 30 days of initial post-KT discharge. Frail KT recipients were much more likely to experience EHR (45.8% vs. 28.0%, P=0.005), regardless of age. After adjusting for previously described registry-based risk factors, frailty independently predicted 61% higher risk of EHR (adjusted RR=1.61, 95% CI: 1.18–2.19, P=0.002). In addition, frailty improved EHR risk prediction by improving the area under the receiver operating characteristic curve (P=0.01) as well as the net reclassification index (P=0.04). Identifying frail KT recipients for targeted outpatient monitoring and intervention may reduce EHR rates.

5. Kidney Disease: Improving Global Outcomes (KDIGO) Kidney Transplant Candidate Work Group. KDIGO Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation. Transplantation. 2020;104: S1 – S103.

Abstract:

The 2020 Kidney Disease: Improving Global Outcomes (KDIGO) Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation is intended to assist health care professionals worldwide who evaluate and manage potential candidates for deceased or living donor kidney transplantation. This guideline addresses general candidacy issues such as access to transplantation, patient demographic and health status factors, and immunological and psychosocial assessment. The roles of various risk factors and comorbid conditions governing an individual ' s suitability for transplantation such as adherence, tobacco use, diabetes, obesity, perioperative issues, causes of kidney failure, infections, malignancy, pulmonary disease, cardiac and peripheral arterial disease, neurologic disease, gastrointestinal and liver disease, hematologic disease, and bone and mineral disorder are also addressed. This guideline provides recommendations for evaluation of individual aspects of a candidate's profile such that each risk factor and comorbidity are considered separately. The goal is to assist the clinical team to assimilate all data relevant to an individual, consider this within their local health context, and make an overall judgment on candidacy for transplantation. The guideline development process followed the Grades of Recommendation Assessment, Development, and Evaluation (GRADE) approach. Guideline recommendations are primarily based on systematic reviews of relevant studies and our assessment of the quality of that evidence, and the strengths of recommendations are provided. Limitations of the evidence are discussed with differences from previous guidelines noted and suggestions for future research are also provided.

[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]

Table 10. Model statistics for risk factors in FYSWR model

Variable	Hazard Ratio	95% Confidence Interval
Age (continuous)	1.10	(1.08, 1.12)
Age Spline 12	0.82	(0.79, 0.85)
Age Spline 18	1.07	(1.05, 1.10)
Age Spline 64	0.92	(0.91, 0.93)
Heart Disease	0.55	(0.53, 0.57)
Do not Ambulate	0.41	(0.34, 0.48)
COPD	0.42	(0.38, 0.46)
Do not Transfer	0.67	(0.49, 0.90)
Cancer	0.58	(0.54, 0.63)
Peripheral Vascular Disease	0.67	(0.63, 0.72)
Cerebral Vascular Accident	0.58	(0.54, 0.62)
Drug Use	0.18	(0.15, 0.21)
Amputation	0.50	(0.45, 0.55)
Assistance with Daily Activities	0.57	(0.54, 0.62)
Area Deprivation Index (ADI), per 10% increase on the percentile scale	0.87	(0.86, 0.88)
Medicare-Medicaid Dual Eligible	0.57	(0.54, 0.60)
Transplant center weighted SRTR mortality ratio	1.39	(1.18, 1.64)
Transplant center weighted SRTR transplant ratio	1.04	(0.97, 1.11)

Table 10. Model statistics for risk factors in FYSWR model

[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]

As noted in section 2b.23, we included Medicare-Medicaid dual eligibility and ADI as social risk factors in the model on a clinical and conceptual basis, and as supported by an expert panel. Both factors were significantly associated with the outcome of waitlisting (see Table 10 in 2b.24).

We additionally examined selected variables, including sex, race and ethnicity, fitting models including covariates from the original model and adding each selected variable one at a time.

Table 11. Hazard Ratio and 95% Confidence Interval of model including sex

Sex	**Hazard Ratio**	**95% Confidence Interval**
Female	0.82	(0.80,0.84)
Male	Reference	Reference

Table 11. Hazard Ratio and 95% Confidence Interval of model including sex

Table 12. Hazard Ratio and 95% Confidence Interval of model including race

Race	**Hazard Ratio**	**95% Confidence Interval**
Native American	0.56	(0.49,0.65)
Asian Pacific Islander	1.11	(1.06,1.17)
Black	0.72	(0.70,0.75)
Other	0.89	(0.73,1.07)
White	Reference	Reference

Table 12. Hazard Ratio and 95% Confidence Interval of model including race

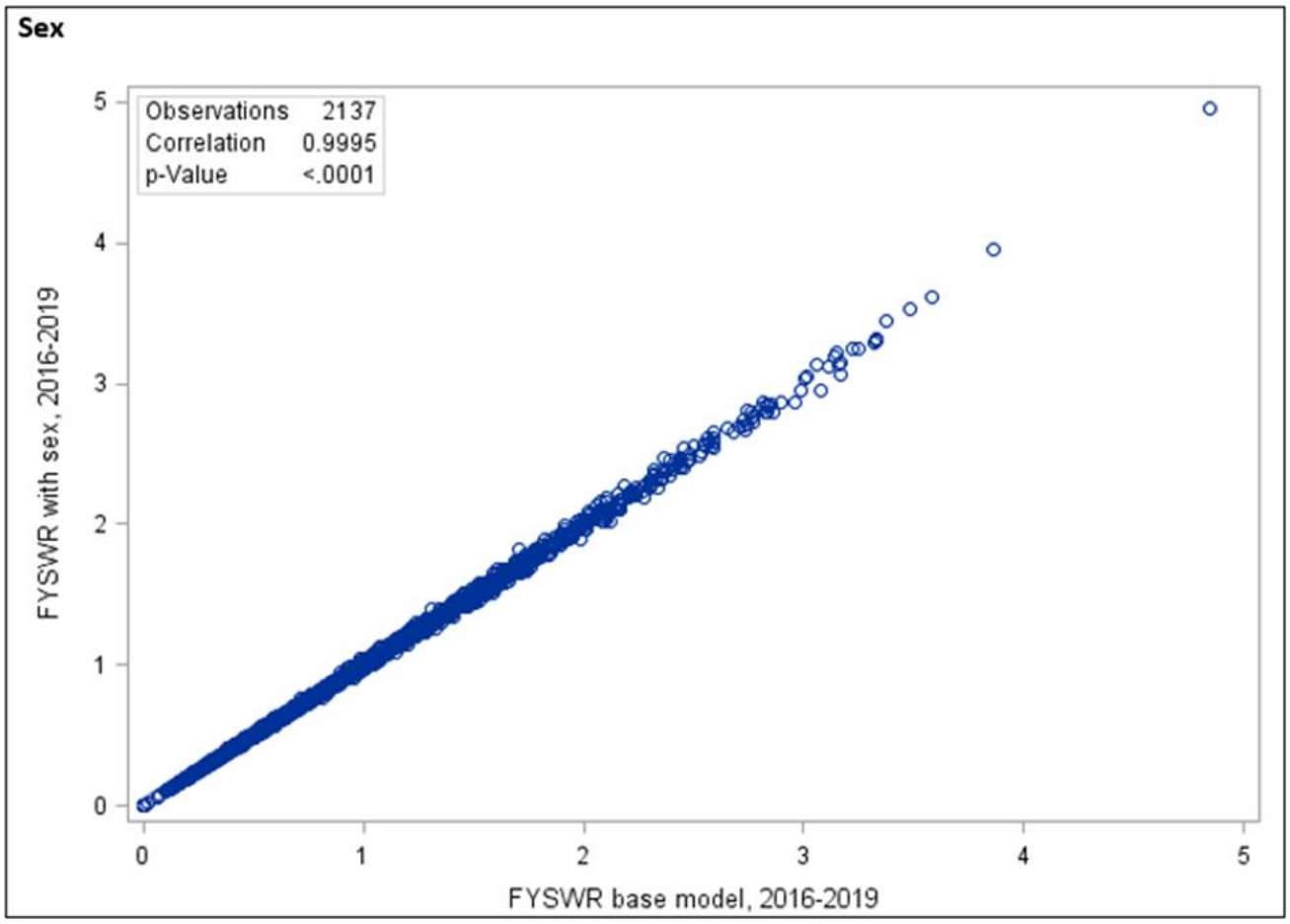
Table 13. Hazard Ratio and 95% Confidence Interval of model including ethnicity

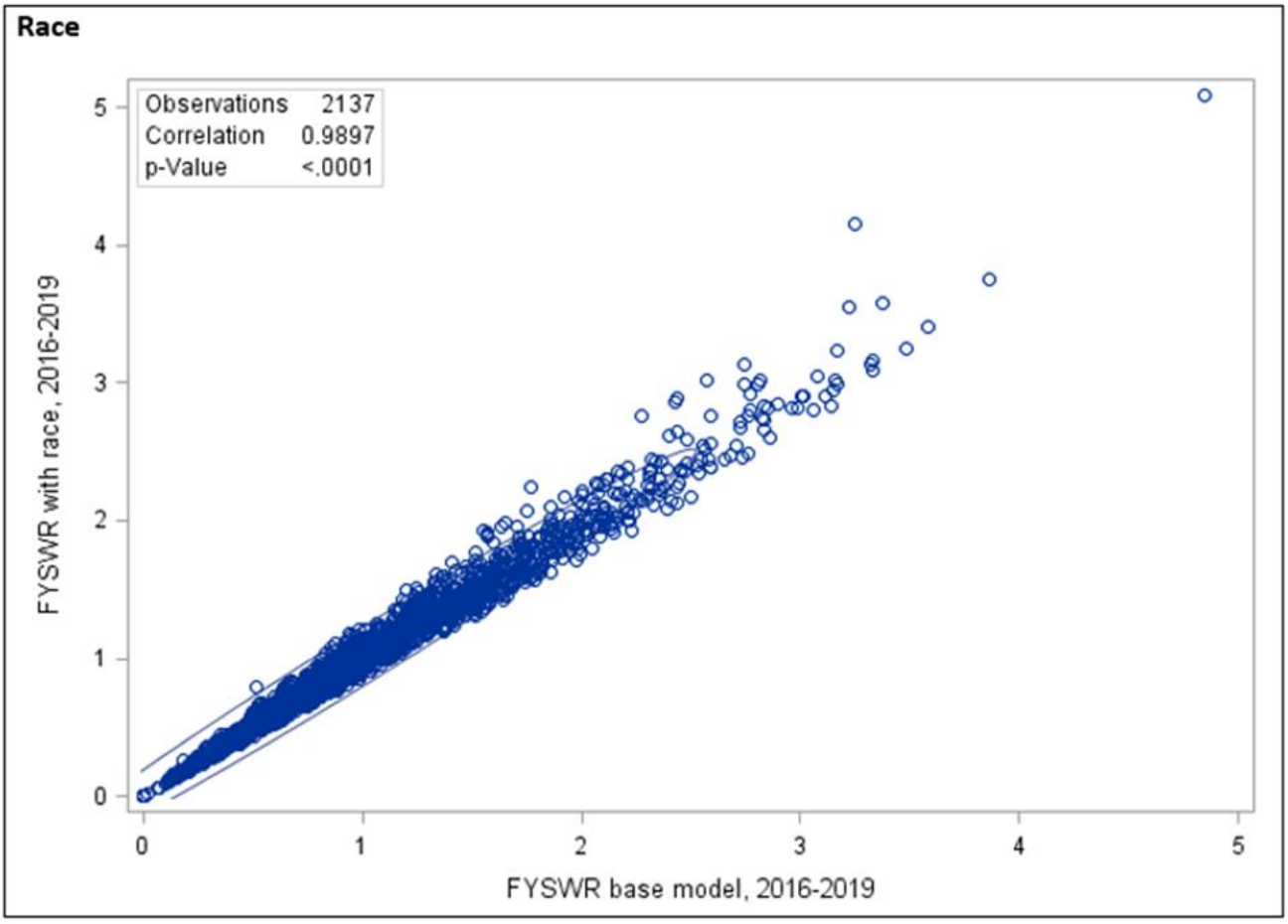
Ethnicity	**Hazard Ratio**	**95% Confidence Interval**
Hispanic	1.00	(0.97,1.04)
Non-Hispanic	Reference	Reference

Table 13. Hazard Ratio and 95% Confidence Interval of model including ethnicity

Compared to men, female patients were less likely to be waitlisted (Hazard ratio = 0.82). Compared to White patients, Asian & Pacific Islanders were more likely to be waitlisted (Hazard ratio = 1.11). Black and Native American patients were less likely to get waitlisted compared with White patients (Hazard ratio = 0.72 and 0.56, respectively). The waitlisting rate for Hispanic patients was not significantly different from Non-Hispanic patients.

Figure 6: Correlation between FYSWR with and without each risk factor


Sex
Observations 2137
Correlation 0.9995
p-Value <.0001
FYSWR with sex, 2016-2019
FYSWR base model, 2016-2019
0
1
2
3
4
5


Race
Observations 2137
Correlation 0.9897
p-Value <.0001
FYSWR with race, 2016-2019
FYSWR base model, 2016-2019
0
1
2
3
4
5

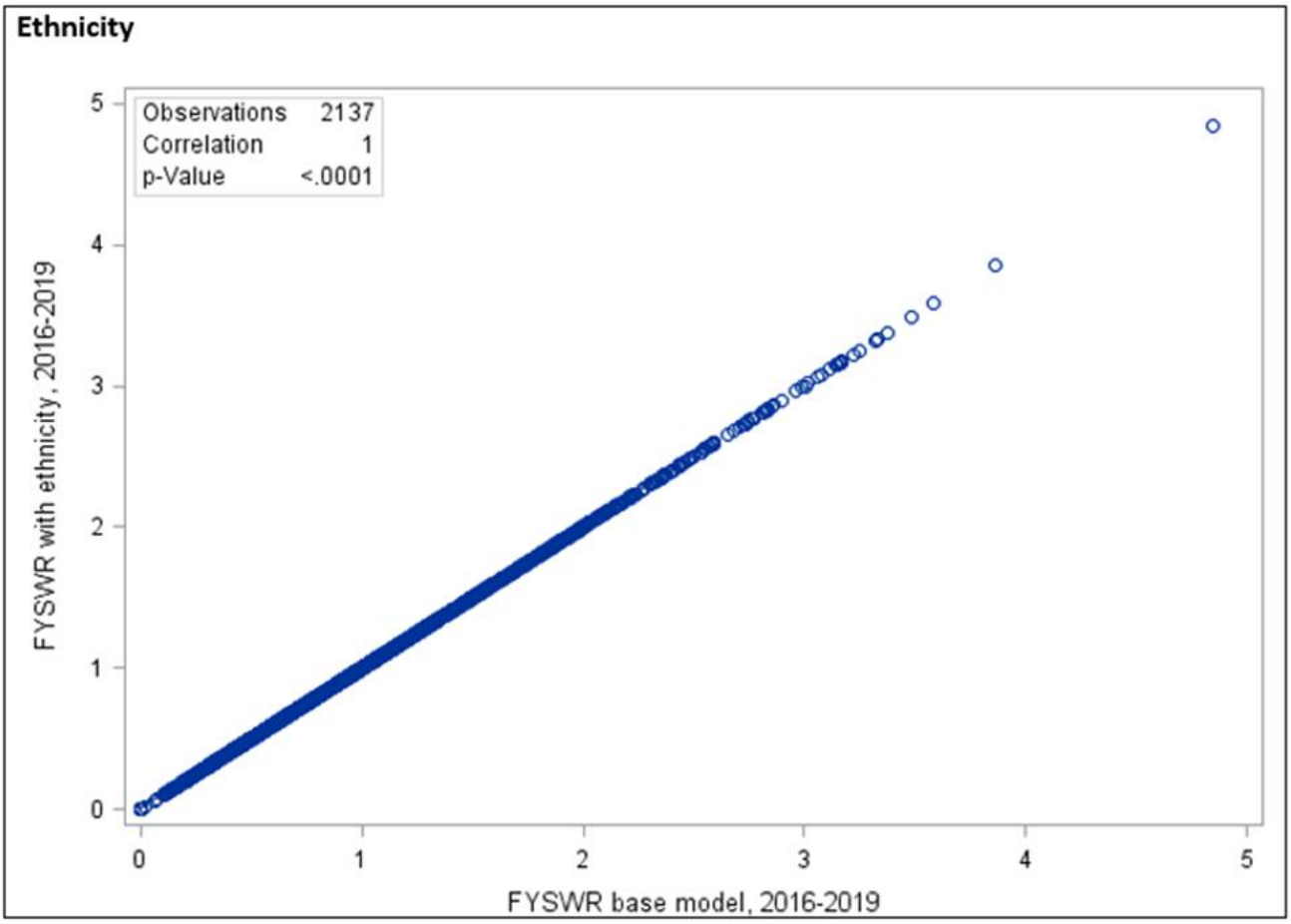


Table 14: Comparison of performances with and without adjusting for risk factors

*	*	**FYSWR without sex**	**FYSWR without sex**	**FYSWR without sex**	**FYSWR without sex**
*	*	**As Expected**	**Better than Expected**	**Worse than Expected**	**Total**
FYSWR with sex	**As Expected**	1,992	2	0	1,994
FYSWR with sex	**Better than Expected**	2	79	0	81
FYSWR with sex	**Worse than Expected**	8	0	54	62
FYSWR with sex	**Total**	2,002	81	62	2,137

Table 14: Comparison of performances with and without adjusting for risk factors

*This cell is intentionally left blank.

*	*	**FYSWR without race**	**FYSWR without race**	**FYSWR without race**	**FYSWR without race**
*	*	**As Expected**	**Better than Expected**	**Worse than Expected**	**Total**
FYSWR with race	**As Expected**	1,985	10	6	2,001
FYSWR with race	**Better than Expected**	5	71	0	76
FYSWR with race	**Worse than Expected**	12	0	48	60
FYSWR with race	**Total**	2,002	81	54	2,137

Table 14: Comparison of performances with and without adjusting for risk factors

*This cell is intentionally left blank.

*	*	**FYSWR without ethnicity**	**FYSWR without ethnicity**	**FYSWR without ethnicity**	**FYSWR without ethnicity**
*	*	**As Expected**	**Better than Expected**	**Worse than Expected**	**Total**
FYSWR with ethnicity	**As Expected**	1,995	1	0	1,996
FYSWR with ethnicity	**Better than Expected**	2	80	0	82
FYSWR with ethnicity	**Worse than Expected**	5	0	54	59
FYSWR with ethnicity	**Total**	2,002	81	54	2,137

Table 14: Comparison of performances with and without adjusting for risk factors

*This cell is intentionally left blank.

Although there are differences in waitlisting by sex and race, it is unclear whether these associations are due to underlying biological or other patient factors, or represent disparities in care. Adjusting for these factors could have the unintended consequence of creating or reinforcing disparities. Furthermore, Tables 11-14 and

Figure 6 show that adjustment for these factors had minimal impact on dialysis practitioner group performance. Therefore, these risk factors were not included in the final risk adjusted model.

[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]

Risk factors were selected for the final model based on the magnitude of the coefficients, evaluation of their statistical significance, and the model C-statistic. The C-statistic measures the discriminative power of the regression model with considered risk factors.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

The C-statistic (also known as the Index of Concordance) was 0.75, meaning that the model correctly ordered 75% of the pairs of patient-months that were discordant with respect to the response variate.

[Response Ends]

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

[Response Begins]

N/A

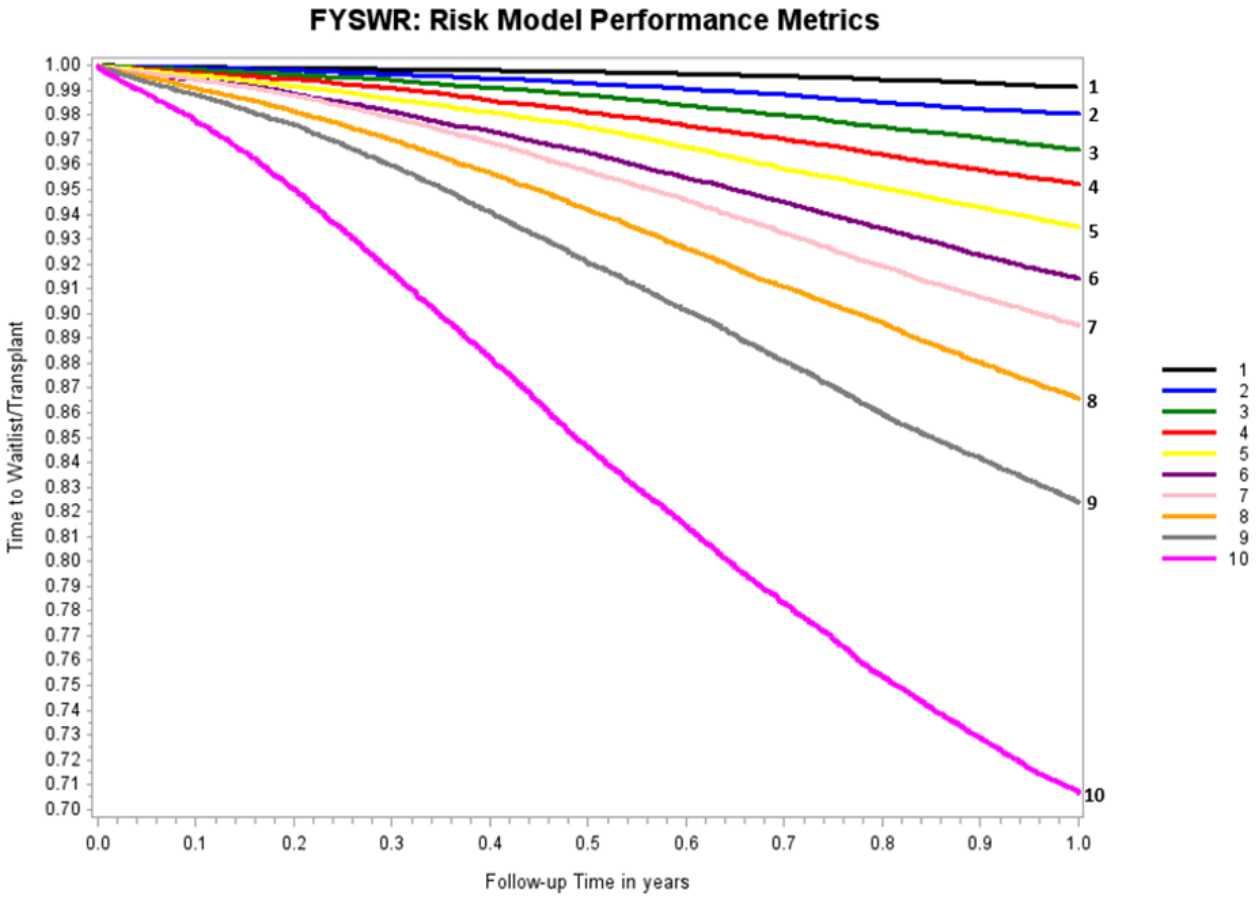
[Response Ends]

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

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

[Response Begins]

Figure 7: Decile plot for FYSWR



[Response Ends]

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

[Response Begins]

N/A

[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]

Figure 7 shows that the risk factors in the model are discriminating well between patients. There is good separation among all 10 groups and the ordering is as predicted by the model. Patients of higher model deciles are much more likely to waitlist or transplant than lower model deciles showing effectiveness of the model to discriminate likelihood of waitlisting.

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

Criterion 3. Feasibility

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

[Response Begins]

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

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

[Response Ends]

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

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

[Response Begins]

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

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

None identified.

[Response Ends]

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

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]

N/A

[Response Ends]

Criterion 4: Use and Usability

4a. Use

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

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

Name of program and sponsor

URL

Purpose

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

Level of measurement and setting

[Response Begins]

Not in use

[Not in use Please Explain]

The measure is undergoing initial endorsement review.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Payment Program

[Response Ends]

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

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

[Response Begins]

The measure is undergoing initial endorsement review.

[Response Ends]

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

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

CMS will determine if/when to report this measure in a public reporting/payment program. One potential application for the measure is in the Quality Payment Program where it would be one of several optional measures that a group practice could select in their evaluation.

[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]

Practitioner group level results have not been disseminated to those being measured as part of the development process. The measure developer sought input from a technical expert panel during development, and those deliberations were open to the public. The TEP summary report was also posted publicly on the CMS website (and is now posted here). The TEP was comprised of stakeholders representing nephrologist (relevant directly to the target of the measure) and dialysis patient perspectives.

[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]

Physician group results have not been disseminated to those being measured as part of the development process.

[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]

Not applicable since the measure is not yet implemented, and results have not been disseminated 4a.08).

[Response Ends]

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

[Response Begins]

As described above, the developer sought input from a technical expert panel during the development of this measure. This group was comprised of stakeholders from nephrologists (those being measured) as well as other stakeholders including a significant number of dialysis/transplant patients. The TEP discussed four waitlisting measures during their deliberations, of which this measure was one.

With respect to the four provisional practitioner level waitlisting measures proposed to the TEP, voting demonstrated majority support for continued development of all of them, including this measure. Support for the measure based on TEP discussions reflected the importance of waitlisting, given it is a crucial and necessary step for transplantation and may confer emotional benefits to patients. In addition, dialysis practitioners can directly contribute to processes necessary for eventual waitlisting, such as educating patients about the benefits of transplantation and assisting with referral to transplant centers for evaluation. TEP members did raise a number of concerns regarding the measure definition, including the need for strong risk adjustment in the areas of social-economic status and comorbid conditions. An adjustment for transplant center effects was also recommended.

The full summary of the TEP feedback can be found here.

[Response Ends]

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

[Response Begins]

See 4a.08.

[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]

As part of the TEP process, the developer presented the TEP with two existing waitlist measures that are currently publicly reported at the facility level as a starting point for development of practitioner-level measures. This measure (one of four resulting from TEP discussion) reflects the input from the TEP on how the construction of the facility level measures should be revised in order to be adapted to the practitioner level and addresses the concerns raised about appropriate risk adjustment.

[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]

The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of this measure. Once implemented practitioner performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in waitlisting rates among the target population.

[Response Ends]

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

[Response Begins]

None.

[Response Ends]

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

[Response Begins]

None.

[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]

Standardized First Kidney Transplant Waitlist Ratio for Incident Dialysis Patients (SWR), Centers for Medicare and Medicaid Services

Percentage of Prevalent Patients Waitlisted (PPPW), Centers for Medicare and Medicaid Services

[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]

Yes

[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]

N/A

[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]

Appendix

Supplemental materials may be provided in an appendix.: Available in attached file

Contact Information

Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Measure Steward Point of Contact: Sardone, Jennifer, jmsto@med.umich.edu

Dollar-Maples, Helen, helen.dollar-maples@cms.hhs.gov

Measure Developer if different from Measure Steward: University of Michigan Kidney Epidemiology and Cost Center

Measure Developer Point(s) of Contact: Sardone, Jennifer, jmsto@med.umich.edu

George, Jaclyn, jaclynrg@med.umich.edu

Parrotte, Casey, parrotte@med.umich.edu

Yaldo, Alexander, yaldo@med.umich.edu

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

Available in attached file

[Response Ends]

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

David Axelrod, MD, MBA
Transplant Surgeon, University of Iowa
Amy Waterman, PhD
Professor of Medicine, Nephrology, UCLA Nephrology
Bobby Howard
Patient, Director, Multicultural Donation Education Program
LifeLink of Georgia
Association of Organ Procurement
Jesse Schold, Mstat, PhD
Research Director, Cleveland Clinic
Emily Watson, MSW, LCSW
Social Worker, Satellite Healthcare, LLC
Krista Lentine, MD, PhD Professor of Medicine
American Society of Nephrology Policy & Advocacy Committee
Saint Louis University ASN Alliance for Kidney Health
Bryan N. Becker, MD, MMM,
Physician, DaVita, Inc.
John T. Ducker, MD, Transplant Nephrologist
Nephrology Associates of Northern Illinois and Indiana
Renal Physicians Association
Teri Browne, PhD, MSW
Associate Dean and Professor
University of South Carolina College of Social Work
Rachel Patzer, PhD, MPH,
Director, Health Services Research Center
Emory University School of Medicine
Della Major, MA
Patient, National Forum of ESRD Networks, member of the Kidney Patient Advisory Council
Sumit Mohan, MD, MPH
Physician and Epidemiologist, Columbia University
American Society of Nephrology Alliance for Kidney Health
Dawn P. Edwards
Patient, National Forum of ESRD Networks Kidney Patient Advisory Council
Geraldine Zingraf, DNP, MBA, RN, CNN, CCTC
Transplant Administrator, Edward Hines, Jr. VA Hospital
Sasha Couch
Patient, Renal Support Network

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]
2022
[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]
01/2022
[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]
Annual
[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]
4/2023
[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]