MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through National Quality Forum's (NQF) Consensus Development Process (CDP). The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Preliminary Analysis* 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 #: 3719

Corresponding Measures:

Measure Title: Prevalent Standardized Waitlist Ratio (PSWR)

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: The PSWR measure tracks the number of prevalent 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. For each practitioner group, the Prevalent Standardized Waitlist Ratio (PSWR) is calculated to compare the observed number of waitlist events in a practitioner group to its expected number of waitlist events. The PSWR uses the expected waitlist events calculated from a Cox model, adjusted for patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI), and transplant center characteristics.

1b.01. 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 prevalent patients on dialysis, examining for the occurrence of new waitlisting or living donor transplant events. 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] Given that many patients may not be ready for transplant candidacy immediately following initiation of dialysis, this measure encourages ongoing attention to transplant candidacy throughout the period following dialysis initiation.

References

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.

---

sp.12. Numerator Statement: Number of prevalent dialysis patients in the practitioner group listed on the kidney or kidney-pancreas transplant waitlist or who received living donor transplants within each calendar year.

sp.14. Denominator Statement: The denominator for the PSWR is the expected number of waitlist or living donor transplant events in the practitioner group according to each patient's treatment history, adjusted for patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI), and transplant center characteristics, among patients under 75 years of age.

sp.16. Denominator Exclusions:

Patients with the below conditions are excluded from the measure:

- Patients were excluded when turning 75.
- Patients who were admitted to a skilled nursing facility (SNF) were excluded from that period.
- Patients were excluded if determined to be in hospice in the prior 365 days
- Patients with dementia

The noted exclusions represent conditions for which transplant waitlist candidacy is highly unlikely, and which can be identified readily with available data.

Patients who were attributed to dialysis practitioner groups with fewer than 11 patients or 2 expected events are not excluded from the measure. If a provider can not be matched to a TIN, patients will be grouped into a

separate 'null' TIN and still included in the models, but are not summarized to any valid individual TINs. 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

sp.28. Data Source:

> Claims

> Registry Data

sp.07. Level of Analysis:

> Clinician: Group/Practice

IF Endorsement Maintenance – Original Endorsement Date:

Most Recent Endorsement Date:

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

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

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

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 can be used, assuming the data are from a robust number of providers and the results are not subject to systematic bias. For measures derived from a patient report, the 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 prevalent dialysis patients in a practitioner 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. For each practitioner group, the Prevalent Standardized Waitlist Ratio (PSWR) is calculated to compare the observed number of waitlist events in a practitioner group to its expected number of waitlist events.
- The developer provides a logic model that depicts a provider can impact if a patient is on the kidney-pancreas transplant list or receives a transplant by educating on transplantation, referring them to a transplant center, assisting with the transplant evaluation process, and optimizing patient's health and functional status.

Summary:

- To demonstrate that the target population finds the measure meaningful, the developer noted that two technical expert panels (TEPs) were convened to discuss potential measures directed at improving access to kidney transplantation. During these TEPs, there was broad support for the importance of waitlisting and formal voting by the TEP demonstrated that a majority of members were in favor of the development quality measures targeting waitlisting at the dialysis facility and practitioner levels.
  - 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.
- To further demonstrate the meaningfulness of the measure to the target population, the developer also cited a study published in the American Journal of Transplantation noting the empirical support for the value of waitlisting to patients.
  - The participants in the study were patients with advanced chronic kidney disease prior to transplant and recipients of transplants.
  - The participants were asked about their priorities in choice of transplant center. They stated that they were most likely to rank waitlisting characteristics, such as ease of getting on the waitlist, as the most important feature.
- The developers provided several studies on the empirical support for connection between processes under the dialysis practitioner's 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 ($r = 0.35$, p-value less than 0.001).
  - The second study which used national registry data from 2005 - 2007 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 half the rate of access to transplantation compared to patients who were informed.
  - A related study of 388 hemodialysis patients noted that patients who reported receiving transplantation information were associated with a three-fold increase in likelihood of waitlisting.
  - The last study of 170 dialysis facilities examined transplant education practices. The study found that facilities that used greater than three education practices had 36 percent higher waitlist rates than facilities that used less education practices.

Question for the Standing 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) → the relationship between the measured health outcome and at least one health action is demonstrated by empirical data (Box 2) → 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 evaluated PSWR performance scores for all practitioner groups that had at least 11 patients and two expected events in the evaluation period between January 1, 2017 and December 31, 2019 resulting in the inclusion of 2,022 groups totaling 362,093 patients.
- The developers found that the mean value of the PSWR was 103 percent with an interquartile range of 63 percent.

Disparities

- The developer presented PSWR scores by race, ethnicity, and sex for the sample.
  - Mean PSWR was highest for the categories Other (437 percent) and Asian Pacific Islander (424 percent) and lowest for White (121 percent).
  - Non-Hispanics had a lower PSWR (107 percent) than Hispanics (164 percent).
  - Females had a higher PSWR (118 percent) than males (98 percent).

Questions for the Standing Committee:

- *With a 103 percent main and 63 percent range in performance, is there enough of a gap in care that warrants a national performance measure?*

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

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**

2a. Reliability: [Specifications](#) and [Testing](#)

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

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

Specifications:

- Measure specifications are clear and precise.

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
  - The developer conducted testing using inter-unit reliability (IUR) with a bootstrap (n=100) approach. This approach utilizes a resampling scheme to estimate the within dialysis provider group practice variation that cannot be directly estimated by ANOVA.
    - The developer calculated a IUR value of 0.56 for the measure, which indicates that 56 percent of the variation in the measure can be attributed to the between-dialysis

practitioner group practice differences and 44 percent to the within-dialysis practitioner group differences.

- The developer noted that this IUR implies a moderate degree of reliability.
- Dialysis practitioner group practices with less than 11 eligible patients and two expected events were excluded from this calculation.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*

Guidance From the Reliability Algorithm

Specifications precise, unambiguous, and complete (Box 1)→ Empirical reliability testing conducted statistical tests as the measure is specified (Box 2) → Empirical reliability testing conducted at the accountable entity level (box 4) → Method described for reliability testing was appropriate (Box 5) → Moderate certainty or confidence that accountable entity levels are reliable (Box 6a) → Moderate

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 that 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 developer hypothesized that for higher PSWR performance, transplant rates would be higher and for higher PSWR performance, mortality would be lower.
  - The developer applied the Cochran-Armitage trend test to evaluate the relationship between tertile groupings and the practitioner group-level outcomes and examined the Spearman correlation between the practitioner group measure rates and each of the outcomes respectively.
    - The dialysis practitioner group level average mortality was 17.7, 17.5, and 18.1 deaths per 100 patient-years for each of the three tertiles (T1 to T3), respectively (trend test p=0.255). The Spearman correlation coefficient was not significant: -0.02 (p=0.264).
    - The dialysis practitioner group level average transplant rate was 4.7, 3.8, and 2.6 transplants per 100 patient-years for each of the three tertiles (T1 to T3), respectively (trend test p<0.001). The Spearman correlation coefficient was significant: 0.41 (p<0.001).
    - The developer noted that higher PSWR performance correlated with higher transplant rates, and the relationship with mortality was also as expected by the developer,

though not statistically significant, with the numerically highest mortality in the lowest performance tertile on the PSWR measure.

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, hospice patients, and dementia patients. The developer noted that exclusions apply to all patients in the models regardless of if the practitioner groups have fewer than 11 patients or two expected events.
    - The developer reported that the number of patients before exclusions was 494,134 and after exclusions it was 362,093; 26.7 percent of patients were excluded.
    - The developer also reported the following frequencies for each excluded variable:
        - Age greater than or equal to 75 – 118,703 patient months (24 percent)
        - Nursing home from nursing home history file – 42,684 patient months (8.6 percent)
        - Hospice – 3,310 patient months (0.7 percent)
        - Dementia-19,369 patient months (3.9 percent)
    - The developer reported that the overall measure scores were changed modestly by the exclusions.
        - The average waitlisting percentage increased from 1.017 percent before exclusions to 1.027 percent after exclusions.
        - The count of "Worse than Expected" practitioner groups increases from 40 to 43 and the count of "Better than Expected" practitioner groups increases from 80 to 84 out of a total of 2,012 groups.
        - The developer stated that though performance scores were modestly affected by exclusions, there is a fair degree of variation in the percentage of patients excluded across practitioner groups. The developer deemed the exclusions as important as they represent a group of patients highly unlikely to be suitable for transplant waitlisting. The developer noted 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 24 variables. The risk model is a two stage Cox model which was fitted to calculate the expected number of events at each facility. At the first stage, a Cox model stratified by dialysis practitioner group practices was fitted in order to obtain an estimate of patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI) and transplant center characteristics 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), patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI) and transplant center characteristics (from the first stage) were then used to compute the probability of an event for each patient, followed by the total expected events for each group practice.
- The developer attested that the variables chosen for inclusion in the model were based on a conceptual rationale that included theoretical/clinical considerations and existing literature for factors affecting kidney transplant waitlisting. The developer stated there were three categories of factors explored and included in the model including:
    - Social risk factors (ADI, dual eligibility status)

- - Functional risk factors (inability to transfer, inability to ambulate, Assistance with ADLs)
    - Clinical/medical risk factors (age, comorbidities, and transplant center characteristics such as waitlist mortality and transplant rate).
- The developer stated that while there are differences in waitlisting by sex, race, and ethnicity, these factors were not included in the final risk model because adjusting for these factors could create or reinforce disparities.
- The developer detailed that 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 model C statistic was 0.72, which means that the model correctly ordered 72 percent of the pairs of patient-months that were discordant with respect to the response variate. The developer attests the decile plots show good separation among all 10 groups and the ordering is as predicted by the model.

Meaningful Differences

- The developer classified dialysis practitioner groups into three groups (As Expected, Better than Expected or Worse than Expected) based on their actual performance compared to their expected performance. The classification into the groups was based on if their observed and expected values were statistically different at the five percent level.
    - Four percent of dialysis practitioner group practices have a PSWR that was "Better than Expected" (median PSWR= 2.19),
    - Ninety-four percent of dialysis practitioner group practices have a PSWR that was "As Expected" (median PSWR=0.93),
    - Two percent of dialysis practitioner group practices have a PSWR that was "Worse than Expected" (median PSWR= 0.18).
- The developer reported that better than expected physician groups 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. The developer reported that these differences are both practically meaningful and statistically significant.

Missing Data

- The developer attested that missing data rarely occurs for the covariates included in the measure. The developer stated that age is calculated using the date of birth and beginning of the treatment period. Further, date of birth is required in the Standard Analysis Data Files therefore, there were no missing values for age, date of birth, and reporting month found.
- The developer assessed missing data for CMS-2728 form, which is used to determine incident comorbidities. The developer found it to be low at 0.51 percent.
    - The developer stated that patients with missing CMS-2728 form were included in the risk model as their own group of missing 2728 and they had a lower hazard of waitlisting compared to those without a missing CMS-2728 form (HR =0.727, 95 percent CI = 0.613, 0.861).

Comparability

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

Questions for the Standing Committee regarding validity:

- *The developer hypothesized that for higher PSWR performance, transplant rates would be higher and for higher PSWR performance, mortality would be lower. The corrections were as expected, however, the Spearman correlation coefficient was not significant for mortality. Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?*

Guidance From the Validity Algorithm

All threats to validity were empirically assessed (Box 1)→ Empirical validity testing conducts on the measure as specified (Box 2) → Empirical testing conducted at the accountable entity level (Box 5) → Method used to test validity was described and appropriate (Box 6) → Moderate certainty or confidence that the accountable entity levels are a valid indicator of quality (Box 7a) → Moderate

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

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 stated 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 attested that all data elements are in defined field in electronic data.
- The developer has not identified any difficulties with data collection, data availability, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, or any other feasibility/implementation issues.

Questions for the Standing 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**

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, and 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames 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

- This measure is a new measure and is not currently in use.The developer stated the measure may be considered for use in a public reporting/payment program in the future.

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; and (3) This feedback has been considered when changes are incorporated into the measure.

Feedback on the measure provided 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.
- The developer also stated that as part of the process, they presented the TEP with two existing waitlist measures at the facility level as a starting point for the practitioner-level measures. The developer attested that this measure reflects the input from the TEP on how construction of the facility level measures should be revised for the practitioner level and addresses the concerns raised about appropriate risk adjustment.

Questions for the Standing 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 (4a1. Improvement; 4a2. Benefits of measure)

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

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

Improvement results

- The developer stated that the measure is not yet implemented in a public reporting program, so improvement could not be evaluated. However, the measure steward, 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 versus harms. The benefits of the performance measure in facilitating progress toward achieving high quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer did not identify any unexpected findings.

Potential harms

- The developer did not report any potential harms.

Additional Feedback:

- N/A

Questions for the Standing 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**

Criterion 5: Related and Competing Measures

Related Measures

- The developer noted one NQF-endorsed measure and two non-NQF-endorsed measures as related measures.
  - NQF #3695 Percentage of Prevalent Patients Waitlisted (PPPW)
  - Standardized First Kidney Transplant Waitlist Ratio for Incident Dialysis Patients (SWR)
  - Percentage of Prevalent Patients Waitlisted

Harmonization

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

Developer Submission

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:

Current Submission:

Updated evidence information here.

Previous (Year) 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 following 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 on maintenance 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 prevalent patients on dialysis, examining for the occurrence of new waitlisting or living donor transplant events. 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] Given that many patients may not be ready for transplant candidacy immediately following initiation of dialysis, this measure encourages ongoing attention to transplant candidacy throughout the period following dialysis initiation.

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 PSWR performance scores for all dialysis practitioner group practices that had at least 11 patients and at least 2 expected events in the evaluation period 2017 through 2019. The mean value of PSWR was 1.03. The interquartile range (Q3-Q1) is 0.63, with the bottom quartile of practitioner group practices having 33% lower, versus the top quartile having 30% higher, waitlisting or living-donor transplant rates among prevalent dialysis patients than the national average. Dates of data: January 1, 2017 – December 31, 2019.

Number of patients: 362,093

Number of practitioner groups: 2,022

Table 1: Descriptive statistics of PSWR overall and by decile, 2017-2019

*	Mean	Std Dev	Minimum	Maximum	Median	Lower Quartile	Upper Quartile
Overall	1.03	0.55	0.00	6.84	0.94	0.67	1.30
Decile	*	*	*	*	*	*	*
1	0.25	0.15	0.00	0.43	0.31	0.16	0.37
2	0.52	0.05	0.43	0.60	0.53	0.49	0.57
3	0.67	0.04	0.60	0.73	0.67	0.64	0.70
4	0.78	0.03	0.73	0.84	0.78	0.76	0.82
5	0.89	0.03	0.84	0.94	0.89	0.87	0.92
6	1.01	0.04	0.94	1.07	1.01	0.98	1.04
7	1.14	0.04	1.07	1.21	1.13	1.10	1.17
8	1.30	0.05	1.21	1.39	1.30	1.25	1.35
9	1.55	0.10	1.40	1.74	1.55	1.47	1.63
10	2.16	0.53	1.74	6.84	1.99	1.86	2.30

*Cells intentionally left blank.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for

improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

Dates of data: January 1, 2017 – December 31, 2019.
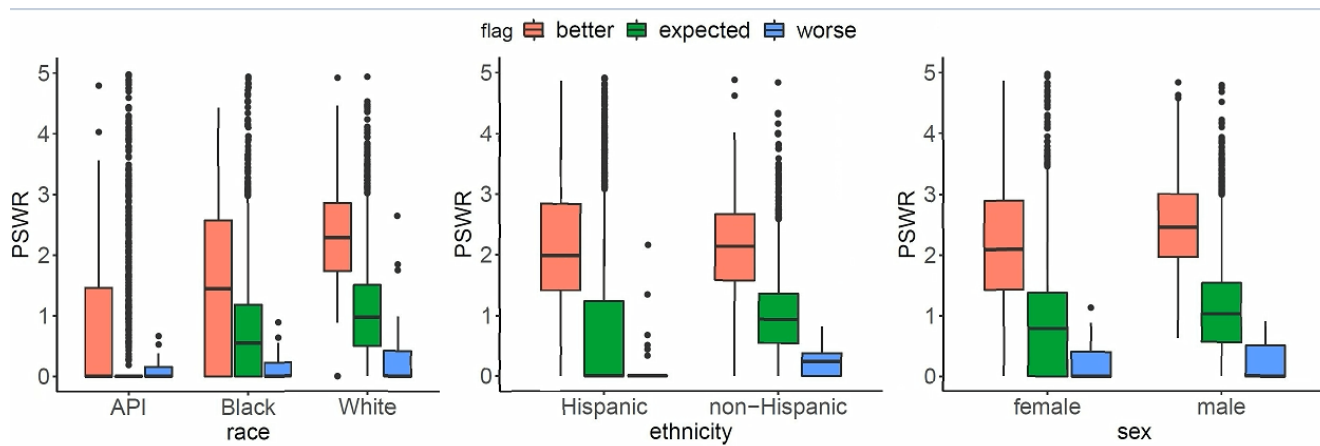
Number of patients: 362,093

Number of practitioner groups: 2,022

Table 2: Descriptive statistics of PSWR, by race, ethnicity and sex, 2017-2019

Group	Mean	Std Dev	Minimum	Maximum	Median	Lower Quartile	Upper Quartile
Race	*	*	*	*	*	*	*
White	1.21	2.42	0.00	97.66	0.99	0.49	1.60
Black	2.13	51.14	0.00	2273.08	0.59	0.00	1.25
Asian Pacific Islander	4.24	67.97	0.00	2544.96	0.00	0.00	1.03
Native American/ Alaskan Native	1.52	11.75	0.00	172.36	0.00	0.00	0.00
Other	4.37	30.90	0.00	529.23	0.00	0.00	0.00
Ethnicity	*	*	*	*	*	*	*
Hispanic	1.64	9.44	0.00	343.23	0.00	0.00	1.55
Non-Hispanic	1.07	0.98	0.00	21.55	0.95	0.54	1.41
Sex	*	*	*	*	*	*	*
Female	1.18	0.96	0.00	8.85	1.04	0.56	1.64
Male	0.98	1.03	0.00	10.56	0.81	0.00	1.44

*Cells intentionally left blank.

Figure 1: Performance of PSWR, by race, ethnicity and sex, 2017-2019

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

[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

2a. Reliability

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]
Prevalent Standardized Waitlist Ratio (PSWR)
[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 PSWR measure tracks the number of prevalent 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. For each practitioner group, the Prevalent Standardized Waitlist Ratio (PSWR) is calculated to compare the observed number of waitlist events in a practitioner group to its expected number of waitlist events. The PSWR uses the expected waitlist events calculated from a Cox model, adjusted for patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI), and transplant center characteristics.
[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

[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]

None available.

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

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

[Response Begins]

 Available in attached Excel or csv file

[Response Ends]

Attachment: 3719_PSWR Data Dictionary_3719_PSWR_data_dictionary-508.xlsx

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

sp.13. 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 prevalent dialysis patients in the practitioner group listed on the kidney or kidney-pancreas transplant waitlist or who received living donor transplants within each calendar year.

[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.14. 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 during the period with the assigned practitioner group. Specifically, date of listing on the kidney or kidney-Pancreas transplant waitlist from OPTN and date of receiving a living donor transplant from OPTN, Form 2728, and Claims are used to identify the event happening during the period with the assigned practitioner group.

[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.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

The denominator for the PSWR is the expected number of waitlist or living donor transplant events in the practitioner group according to each patient's treatment history, adjusted for patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI), and transplant center characteristics, among patients under 75 years of age.

[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.16. 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]

During the target period for eligible Medicare ESRD dialysis patients, Medicare physician claims were used to identify 1) the individual dialysis practitioner that received the monthly capitation payment (MCP) and 2) the dialysis group practice identifier to which that practitioner belongs. Tax identification numbers (TINs) are used to identify the dialysis practitioner group practices on Medicare physician claims. For each period, the patient was assigned to the practitioner, and in turn to that dialysis practitioner's group practice, which as a whole provided dialysis services with the most face-to-face interaction, according to the Healthcare Common Procedure Coding System (HCPCS) codes. Both TIN and MCP present on the physician claims. We do not distinguish in person claims or not.

Monthly capitation payment HCPCS codes included are the following: 90951, 90952, 90953, 90954, 90955, 90956, 90957, 90958, 90959, 90960, 90961, 90962, 90963, 90964, 90965, 90966. Information regarding first ESRD service date, modality, death, waitlist status, and transplant are obtained from Medicare claims, EQRS, Organ Procurement and Transplant Network (OPTN), and the Social Security Death Master File.

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Patients with the below conditions are excluded from the measure:
- Patients were excluded when turning 75.
- Patients who were admitted to a skilled nursing facility (SNF) were excluded from that period.
- Patients were excluded if determined to be in hospice in the prior 365 days
- Patients with dementia

The noted exclusions represent conditions for which transplant waitlist candidacy is highly unlikely, and which can be identified readily with available data.

Patients who were attributed to dialysis practitioner groups with fewer than 11 patients or 2 expected events are not excluded from the measure. If a provider can not be matched to a TIN, patients will be grouped into a separate 'null' TIN and still included in the models, but are not summarized to any valid individual TINs. 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.18. 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 was used to identify patients in skilled nursing facilities. For hospice patients, a separate CMS file that contains final action claims submitted by hospice providers during the past 365 days was used to determine the hospice status.

In addition, we used Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software (CCS) diagnosis categories for prevalent comorbidity selection, including dementia. Patients with evidence of dementia in the prior year were excluded from analysis.

Please refer to the attached data dictionary for more details about the data source, as well as the ICD 10 codes for prevalent comorbidity conditions including dementia.

[Response Ends]

sp.19. 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.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

Yes

[Response Ends]

sp.21. 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.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.23. 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.24. 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 flowchart in attachments.

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*

- *The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*

- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*

- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

N/A

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

Registry Data

[Response Ends]

sp.31. 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 sources for establishing the denominator. EQRS was used for the age risk adjustment and exclusion of patients aged 75 or older. Organ Procurement and Transplant Network (OPTN) is the data source for the numerator (waitlisting or living donor kidney transplantation). Medicare claims from the year prior to the reporting period were used for comorbidity condition adjustments. Medicare claims during the reporting period were used for the hospice exclusion criteria. The Nursing Home Minimum Dataset was used to identify SNF patients. Additionally, Medicare claims during the reporting period 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 dialysis 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.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

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

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

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

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

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

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

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):
- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

- 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 Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) 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 derived from a combination of EQRS (formerly CROWNWeb), the Nursing Home Minimum Dataset, transplant registries (OPTN, SRTR), the CMS Medical Evidence Form (CMS Form 2728), Medicare claims from CMS, and the monthly capitation payment (MCP) from the Integrated Data Repository (IDR).

[Response Ends]

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

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

[Response Begins]

01-01-2017 – 12-31-2019

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

 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 2017 through 2019, there were 2,202 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 362,093 patients in total. The average age at their initiation of dialysis was 58.8 years old, 42.5% were female, 58.4% were White, 35.5% were Black, 4.4% were Asian/Pacific Islander, 1.5% were American Indian/Alaskan Native, 0.2% were Other/Multi-racial/Unknown and 16.6% 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.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

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

Choose one or both levels.

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

We used January 2017 – December 2019 data to calculate the Prevalent Standardized Waitlist Ratio. Our approach for determining measure reliability aligns with one-way analysis of variance (ANOVA), in which the between dialysis provider group practice variation (σ_b^2) and the within- dialysis provider 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 provider group practice variation, the true signal reflects the differences across dialysis provider 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 provider 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 provider 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 provider group practices is driven by random noise, indicating the measure would not be a good characterization of the differences among dialysis provider group practices. A large IUR (near 1) indicates that most of the variation between dialysis provider group practices is due to the real different between dialysis provider group practices.

Below is our approach to calculate IUR.

Let $T_1,...,T_N$ be the PSWRs for *N* provider groups. Within each practitioner group, select at random and with replacement B = 100 bootstrap samples. That is, if the ith practitioner group has n_i subjects, randomly draw with replacement n_i subjects from those in the same practitioner group, find their corresponding PSWRs and repeat the process 100 times. Thus, for the ith practitioner group, we have bootstrapped PSWRs of $T_{i1}^*...,T_{i100}^*$ Let S_i^* be the sample variance of this bootstrap sample. From this it can be seen that

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

and the total variation in PSWRs can be estimated by

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

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

$$n' = \frac{1}{N-1}\left(\Sigma n_i - \left.\Sigma n_i^2 \middle/ \Sigma n_i\right.\right)$$

is approximately the average dialysis provider group practice size (number of patients per dialysis provider group practice). Note that s_t^2 is an estimate of $\sigma_b^2 + \sigma_{t,w}^2$ where σ_b^2 is the between-group variance, the true signal reflecting the differences across practitioner groups. Thus, the IUR, which is defined by IUR = $\left.\sigma_b^2 \middle/ (\sigma_b^2 + \sigma_{t,w}^2)\right.$ and can be estimated by $\left.(s_t^2 + s_{t,w}^2) \middle/ s_t^2\right.$.

The reliability of PSWR calculation only included dialysis provider 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 0.56.
[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 56% of the variation in the PSWR measure can be attributed to the between-dialysis practitioner group practice differences (signal) and about 44% 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 PSWR 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 PSWR 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 PSWR (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 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 PSWR and the mortality rate or 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): 1.16 - 6.84

T_2 0.77 - 1.16

T_3 (worst performance): 0 - 0.77

The dialysis practitioner group average mortality is 17.7, 17.5, 18.1 deaths per 100 patient-years for T_1, T_2, T_3 groups, respectively (trend test p=0.255). The Spearman correlation coefficient is: -0.02 (p=0.264).

The dialysis practitioner group average transplant rate is 4.7, 3.8, 2.6 transplants per 100 patient-years for the T_1, T_2, T_3 groups, respectively (trend test p<.001). The Spearman correlation coefficient is: 0.41 (p<.001).

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

As expected, better PSWR performance correlated with higher transplant rate, with clear separation of transplant rates across practitioner group tertiles of performance. The direction of the relationship with mortality was as expected (modest negative correlation with better PSWR performance), with the numerically highest mortality in the lowest performance tertile of the PSWR measure, though results 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 PSWR 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 empirical null 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 PSWR, stratified by classification category.

Classification Category	Count	Percent	Median PSWR
Better Than Expected	80	4%	2.19

Classification Category	Count	Percent	Median PSWR
As Expected	1902	94%	0.93
Worse Than Expected	40	2%	0.18
Total	2,022	100%	0.94

[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 covariates included in this measure.

Age is calculated using the date of birth and beginning of the treatment period. Date of birth is required in our Standard Analysis Data Files, therefore no missing values were identified in the patient population. We assessed missing data for the CMS-2728 form which is used to determine incident comorbidities.

[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 362,093 patients

Data element	Missing (%)
Patients with missing CMS-2728	1,862 (0.51)

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

There is a low percentage of patients with missing CMS-2728 Forms. Missing CMS-2728 was accounted for with a category for missingness in the model. As shown in Table 10 in section 2b.24, patients with missing CMS-2728 form have a lower hazard of waitlisting compared to those without a missing CMS-2728 form (HR = 0.727 ; 95% CI = 0.613, 0.861).

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, hospice patients, and dementia, 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 practitioners with fewer than 11 attributed patients or 2 expected events in the models. Exclusions apply to all patients in the models, regardless of whether the practitioner groups have fewer than 11 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	494,134	362,093	26.7%

*Cell intentionally left blank.

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

Variable excluded	Frequency (%)
Age >=75	118,703 (24.0%)
Nursing home from Nursing home history file	42,684 (8.6%)
Hospice	3,310 (0.7%)
Dementia	19,369 (3.9%)

Table 7: Distribution of performance scores before and after exclusion

PSWR	Mean	Std Dev	Minimum	Maximum	Lower Quartile	Median	Upper Quartile
Before exclusion	1.017	0.523	0	5.591	0.672	0.947	1.279
After exclusion	1.027	0.55	0	6.843	0.666	0.943	1.297

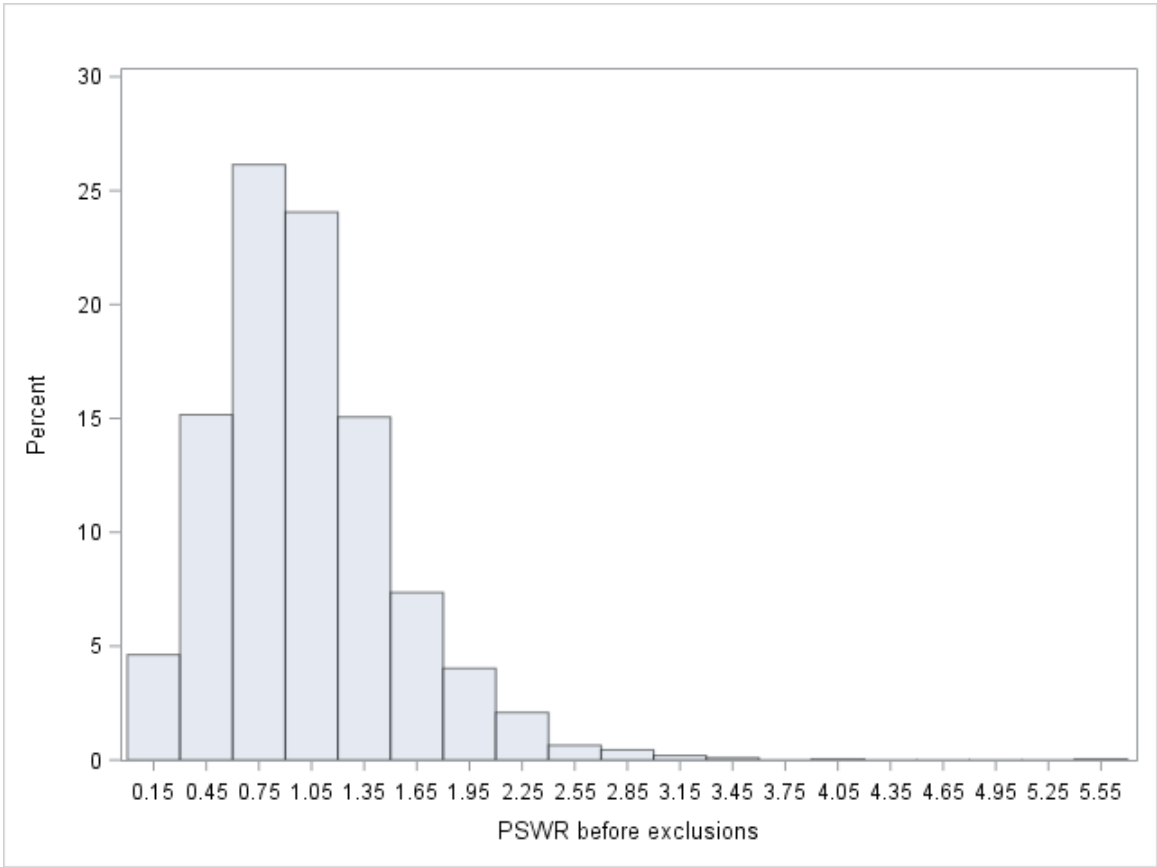**Figure 2:** Distribution of PSWR before exclusions

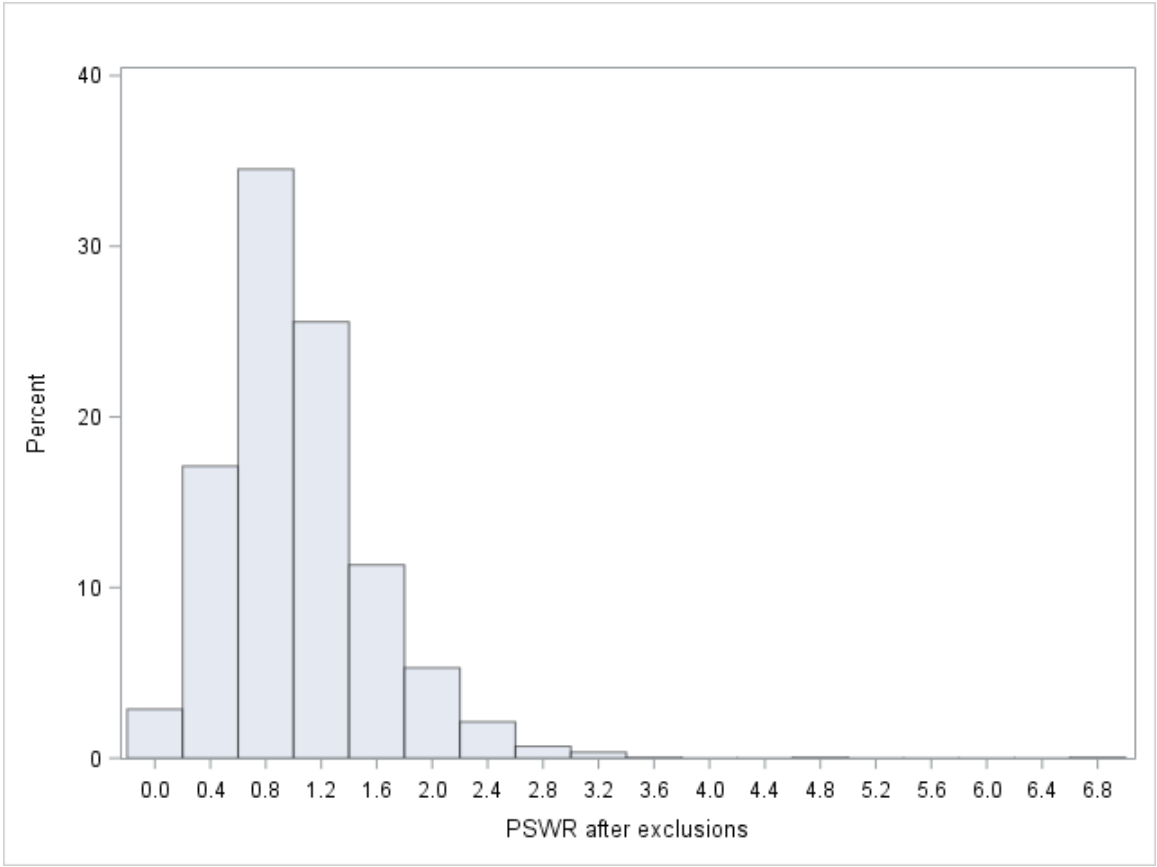**Figure 3:** Distribution of PSWR after exclusions

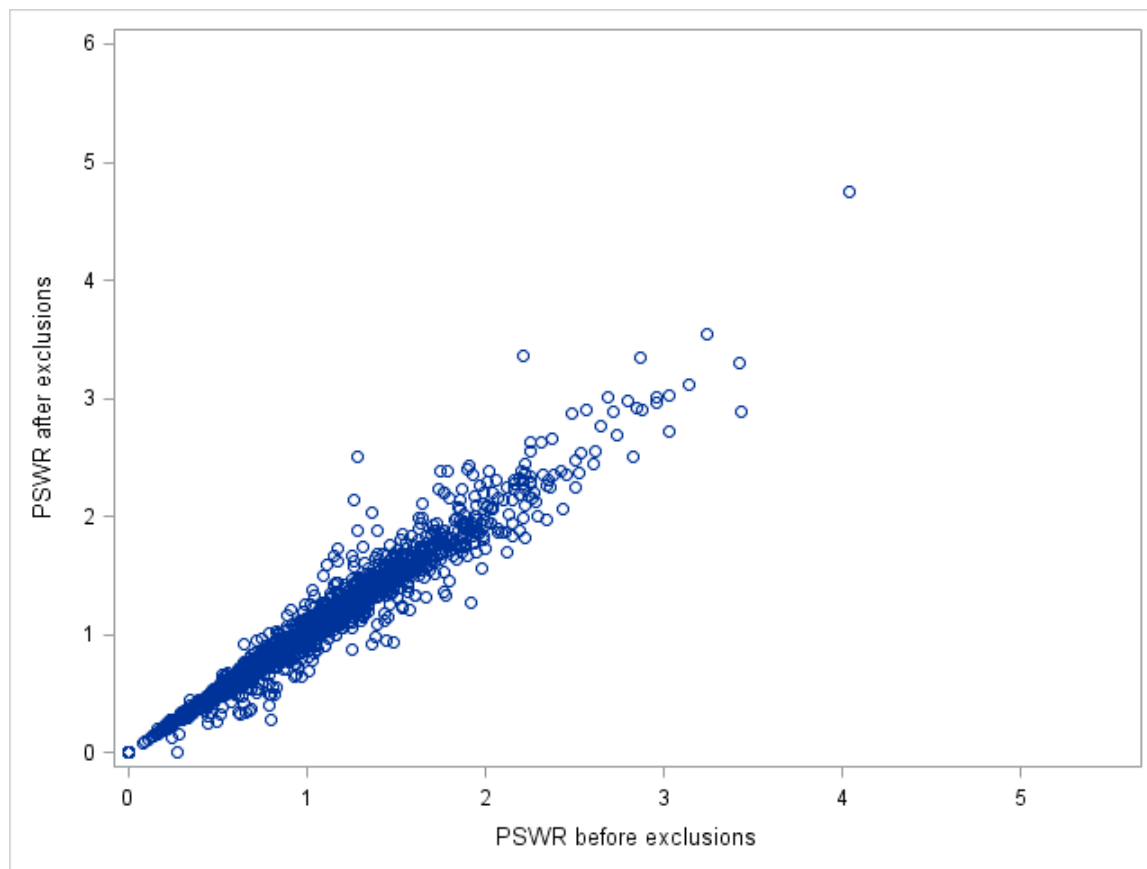**Figure 4:** Scatterplot of PSWR with and without exclusions

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

*	*	PSWR without patient-level exclusion	PSWR without patient-level exclusion	PSWR without patient-level exclusion	PSWR without patient-level exclusion
*	*	Better than Expected	As Expected	Worse than Expected	Total
PSWR with patient-level exclusion	Better than Expected	68	16	0	84
PSWR with patient-level exclusion	As Expected	12	1865	8	1885
PSWR with patient-level exclusion	Worse than Expected	0	11	32	43
PSWR with patient-level exclusion	Total	80	1892	40	2012

*Cell intentionally left blank.

10 practitioner groups do not have a PSWR in the before exclusions due to <2 expected events.

Figure 5: Distribution of excluded patients (%) across dialysis practitioner group practices
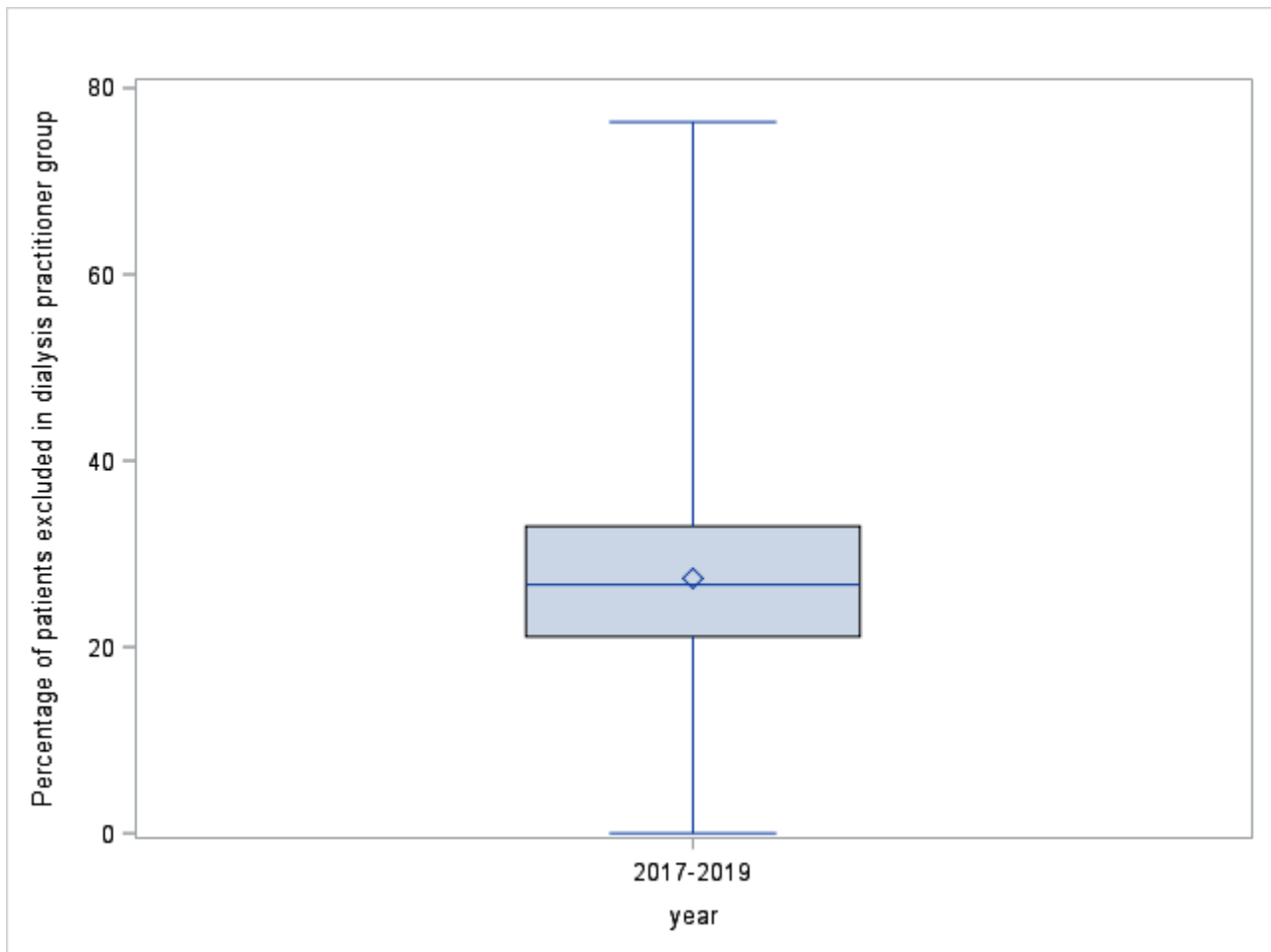
[Response Ends]

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]

Detailed in 2b.20).

[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.

Covariate	Notes
Age categories	0-4; 5-9; 10-14; 15-19; 20-24; 25-29; 30-34; 35-39; 40-44; 45-49; 50-54; 55-59; 60-64; 65-69 (reference); 70-74
Dual eligibility	*
ADI	Area Deprivation Index
Previous waitlisting	*
Previous transplant	*
Year	2017; 2018; 2019 (reference)
Transplant center characteristics	Weighted waitlist mortality ratio; Weighted transplant rate ratio
Comorbidities at incidence	*
Heart disease	*
Other cardiac disease	*
Congestive heart failure	*
Chronic obstruction pulmonary disease	*
Inability to ambulate	*
Inability to transfer	*
Cancer	*
Peripheral vascular disease	*
Cerebrovascular disease	*
Tobacco use	*
Drug use	*
Diabetes, non-primary	*
At least one incident comorbidity listed	*
No Medical Evidence (CMS-2728 Form)	*

Covariate	Notes
Prevalent comorbidities	A set of prevalent comorbidities based on either Medicare inpatient or outpatient claims (individual comorbidities categorized into 64 categories)
At least 6 months of Medicare Coverage in prior year	*

*Cell intentionally left blank.

The event was defined as waitlisting or living-donor transplantation. Patients were followed until waitlisting, living donor transplantation, death, or the end of the follow-up period. 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 patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI), and transplant center characteristics (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), patient age, incident and prevalent comorbidities, previous waitlisting and transplant, dual eligibility, Area Deprivation Index (ADI), and transplant center characteristics (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 adjustment variables in the model and x_{ij} be the specific value of the jth characteristic for the ith patient-record. At the first stage, for patient-record i, we denote the measured characteristics or covariates as

$$\mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{ip}),$$

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 ith patient-record is given by

$$X_i\beta = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}. \quad (1)$$

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

$$A_i = \beta_0 x_{i0} + X_i\beta = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \quad (2)$$

To account for patients starting the follow-up period at different dates during the year, left truncation has been implemented using the counting process syntax in the SAS procedure PROC PHREG. Suppose that t_{iL} and t_{iR} are the starting and end of follow-up time for patient-record i, respectively, so that $S_0(t_{iR})/S_0(t_{iL})$ is the baseline conditional survival probability at time t_{iR} conditional on that the patient start the follow-up period at time t_{iL}. The conditional survival probability for this patient-record i at time t_{iR} is:

$$S_i(t_{iR})/S_i(t_{iL}) = [S_0(t_{iR})/S_0(t_{iL})]^{\exp(A_i)}. \quad (3)$$

The expected number of waitlistings for this patient-record during follow-up time from t_{iL} to t_{iR} arises from considerations in the Cox model and can be written as

$$-\ln(S_i(t_{iR})/S_i(t_{iL})) = -e^{A_i} \ln [S_0(t_{iR})/S_0(t_{iL})]. \quad (4)$$

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

$$E = \sum -\ln[S_0(t_{iR})/S_0(t_{iL})] = -\sum e^{A_i} \ln[S_0(t_{iR})/S_0(t_{iL})]. \quad (5)$$

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 PSWR is the ratio of the total number of observed waitlisting to the expected number

$$\mathbf{PSWR = O/E.} \quad (6)$$

46

[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 (the Kidney Disease Improving Global Outcomes [KDIGO] Clinical Practice Guideline on the Evaluation and Management of Candidates for Kidney Transplantation recommends psychosocial assessment and evaluation of adherence). 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.

In addition, incident and prevalent comorbidities were selected for adjustment into the PSWR 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. For prevalent comorbidities, we used the Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software (CCS) diagnosis categories using Medicare claims. First, we selected comorbidity groupers that were positively and statistically significantly associated with one- year mortality, to again identify conditions associated with early mortality, and therefore potential unsuitability for transplant waitlisting. Then, we included potential candidate conditions that had a prevalence greater or equal to 0.1% in our population to identify a final set of 64 prevalent comorbidities.

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, weighted transplant center waitlist mortality ratio and transplant ratio were calculated on each zip code level. Weight of a transplant center was determined by the likelihood of transplant from neighborhood waitlising patients. patients Patients were then assigned with to transplant center waitlist mortality ratio and transplant ratio based on historical waitlisting patterns in their zip code of residence matched with the corresponding transplant centers.

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 PSWR model

Covariate	Hazard Ratio	95% CI
Age	*	*
0-4	8.396	6.695, 10.529
5-9	9.139	6.664, 12.532
10-14	7.797	6.013, 10.112
15-19	6.705	5.677, 7.918
20-24	4.847	4.418, 5.316
25-29	3.749	3.496, 4.021

Covariate	Hazard Ratio	95% CI
30-34	3.031	2.856, 3.218
35-39	2.638	2.499, 2.785
40-44	2.228	2.119, 2.344
45-49	1.939	1.852, 2.031
50-54	1.573	1.504, 1.644
55-59	1.400	1.341, 1.461
60-64	1.161	1.113, 1.212
65-74	0.431	0.408, 0.455
Dual eligibility	0.779	0.760, 0.799
ADI	0.930	0.923, 0.937
Previous waitlisting	0.859	0.824, 0.897
Previous transplant	1.282	1.219, 1.348
Year	*	*
2017	0.864	0.840, 0.889
2018	0.942	0.916, 0.968
Transplant center characteristics	*	*
Weighted waitlist mortality ratio	1.130	0.936, 1.365
Weighted transplant rate ratio	1.073	0.998, 1.153
Comorbidities at incidence	*	*
Heart disease	0.933	0.885, 0.983
Other cardiac disease	0.952	0.911, 0.993
Congestive heart failure	0.688	0.663, 0.714
Chronic obstruction pulmonary disease	0.433	0.374, 0.500
Inability to ambulate	0.662	0.608, 0.721
Inability to transfer	0.777	0.621, 0.973
Cancer	0.811	0.751, 0.876
Peripheral vascular disease	0.743	0.697, 0.792
Cerebrovascular disease	0.777	0.729, 0.827
Tobacco use	0.545	0.512, 0.579
Drug use	0.509	0.445, 0.582
Diabetes, non-primary	0.714	0.682, 0.747
At least one incident comorbidity listed	0.811	0.770, 0.855
No Medical Evidence (CMS-2728 Form)	1.088	1.038, 1.140

Covariate	Hazard Ratio	95% CI
Prevalent comorbidities	*	*
At least 6 months of Medicare Coverage in prior year	0.727	0.613, 0.861
Candidal esophagitis	0.957	0.733, 1.250
Sarcoidosis	1.165	0.965, 1.407
Cancer of Liver	1.234	0.927, 1.643
Cancer of Lung	0.374	0.249, 0.560
Cancer of Bladder	0.784	0.553, 1.111
Cancer of Bone	0.202	0.101, 0.406
Other Neoplasm	0.942	0.754, 1.175
Non-Hodgkins Lymphoma	0.797	0.589, 1.078
Multiple Myeloma	0.571	0.461, 0.707
Myelodysplastic Syndrome	0.790	0.584, 1.069
Diabetes without complications	1.155	1.108, 1.203
Diabetes with complications	1.159	1.110, 1.211
Glucocorticoid deficiency	1.069	0.905, 1.262
Malnutrition / Cachexia	0.928	0.875, 0.984
Disorders of urea cycle metabolism	0.753	0.524, 1.082
Other amyloidosis	1.165	0.860, 1.580
Other specified disorders of metabolism	0.932	0.815, 1.066
Sickle-cell Anemia	0.756	0.566, 1.008
Pancytopenia	0.978	0.887, 1.078
Neutropenia	1.372	1.169, 1.609
Substance Related Disorders	0.751	0.583, 0.967
Opioid Dependence	0.708	0.622, 0.806
Schizophrenia	0.311	0.226, 0.429
Peripheral autonomic neuropathy in disorders classified elsewhere	0.734	0.498, 1.081
Epilepsy	0.742	0.693, 0.796
Bipolar Disorder	0.808	0.709, 0.920
Major depressive affective disorder	0.839	0.802, 0.879
Alcohol Related Disorders	1.052	0.895, 1.236
Coma	0.900	0.704, 1.150
Cerebral edema	1.174	0.838, 1.644

Covariate	Hazard Ratio	95% CI
Myocardial Infarction	0.695	0.639, 0.757
Coronary Atherosclerosis	1.028	0.970, 1.090
pulmonary embolism and infarction	0.960	0.839, 1.097
Primary pulmonary hypertension	0.807	0.673, 0.967
Pulmonary Heart Disease	0.915	0.859, 0.974
Cardiomyopathy	0.902	0.854, 0.953
Atrioventricular block, complete	0.731	0.582, 0.917
Paroxysmal Tachycardia	0.849	0.754, 0.955
Atrial fibrillation	0.822	0.778, 0.868
Atrial flutter	0.891	0.792, 1.001
Acute Cerebrovascular Disease	0.845	0.775, 0.921
Peripheral and Visceral Atherosclerosis	0.865	0.826, 0.906
Venous Thromboembolism	0.966	0.886, 1.054
Esophageal varices	2.122	1.692, 2.660
Chronic Obstructive Pulmonary Disease	0.602	0.568, 0.638
Aspiration Pneumonitis	0.965	0.827, 1.127
Other Lower Respiratory Diseases	1.348	1.137, 1.597
Respiratory Failure	0.801	0.757, 0.847
Cirrhosis of Liver	1.035	0.941, 1.137
Other Liver Disease	1.268	1.107, 1.454
Pancreatitis	0.821	0.701, 0.961
Chronic Skin Ulcer	0.630	0.591, 0.671
Systemic lupus erythematosus and connective tissue disorders	1.177	1.089, 1.272
Rheumatoid Arthritis	1.211	1.078, 1.359
Pathologic Fracture	0.951	0.725, 1.247
Gangrene	0.662	0.581, 0.754
HIV	0.874	0.778, 0.980
Gastrostomy status	1.194	0.995, 1.433
Other artificial opening of urinary tract status	0.715	0.501, 1.021
Dependence on respirator, status	1.247	1.019, 1.527
Below knee amputation status	0.573	0.508, 0.646
Above knee amputation status	0.446	0.342, 0.582

Covariate	Hazard Ratio	95% CI
Long-term (current) use of insulin	1.14	1.093, 1.188
Inflammatory polyarthropathy	1.012	0.724, 1.415

*Cell intentionally left blank.

[Response Ends]

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

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

[Response Begins]

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

Sex	Hazard Ratio	95% Confidence Interval
Female	0.834	0.814, 0.854
Male	Reference	Reference

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

Race	Hazard Ratio	95% Confidence Interval
Native American	0.625	0.492, 0.794
Asian Pacific Islander	1.186	0.965, 1.458
Black	0.830	0.806, 0.854
Other	1.018	0.833, 1.243
White	Reference	Reference

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

| Ethnicity	Hazard Ratio	95% Confidence Interval
Hispanic	1.173	1.134, 1.214
Non-Hispanic	Reference	Reference

Compared to men, female patients were less likely to be waitlisted (Hazard ratio = 0.834). Compared to White patients, Asian & Pacific Islanders were more likely to be waitlisted (Hazard ratio = 1.186). Black and Native American patients were less likely to get waitlisted compared with White patients (Hazard ratio = 0.830 and 0.625, respectively). The waitlisting rate for Hispanic patients were more likely comparing with Non-Hispanic patients (Hazard ratio = 1.173).

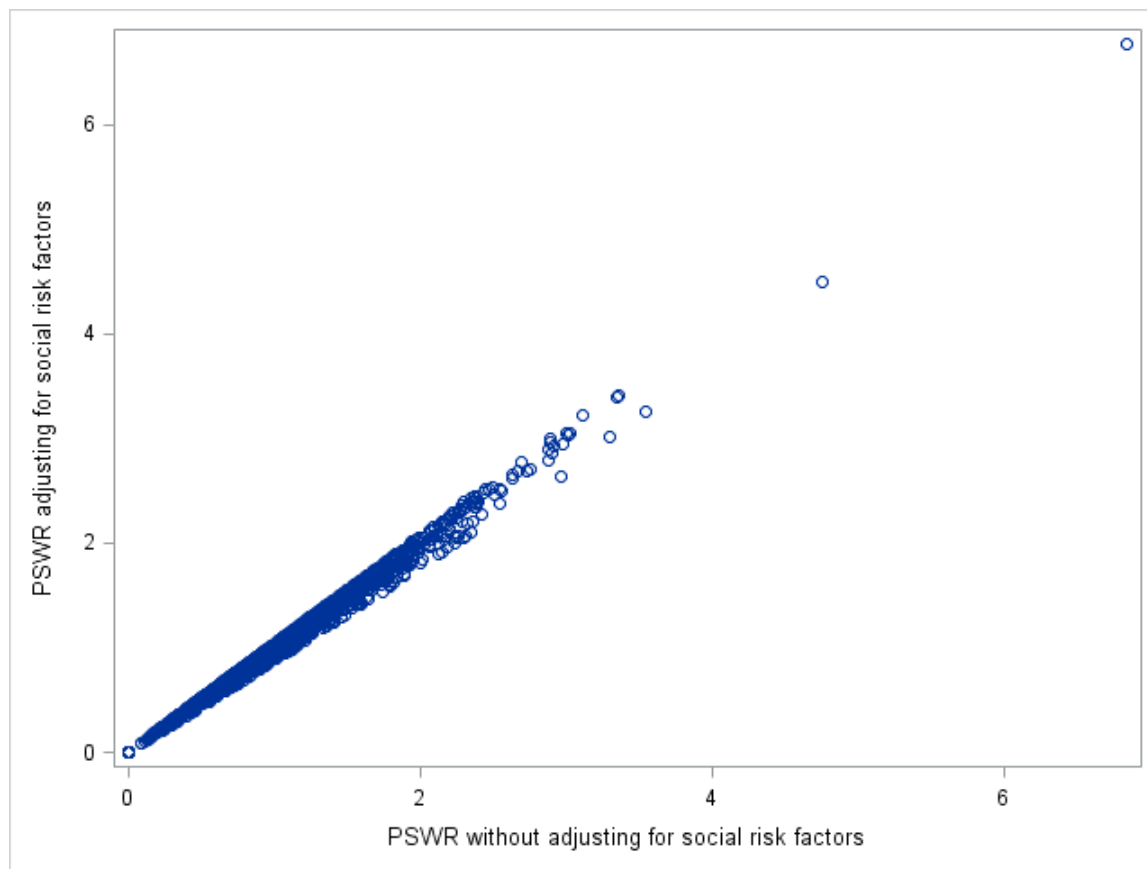**Figure 6:** Correlation between PSWR with and without each risk factor

Sex

Race

55
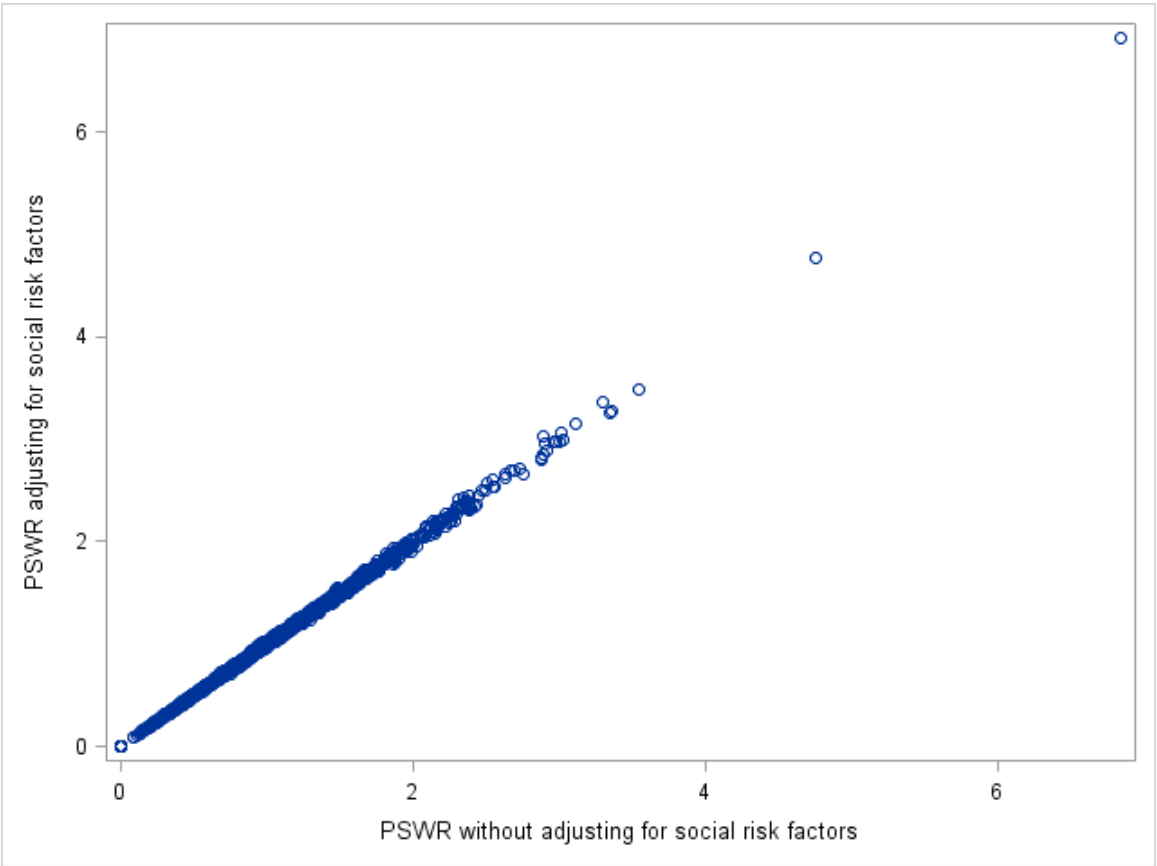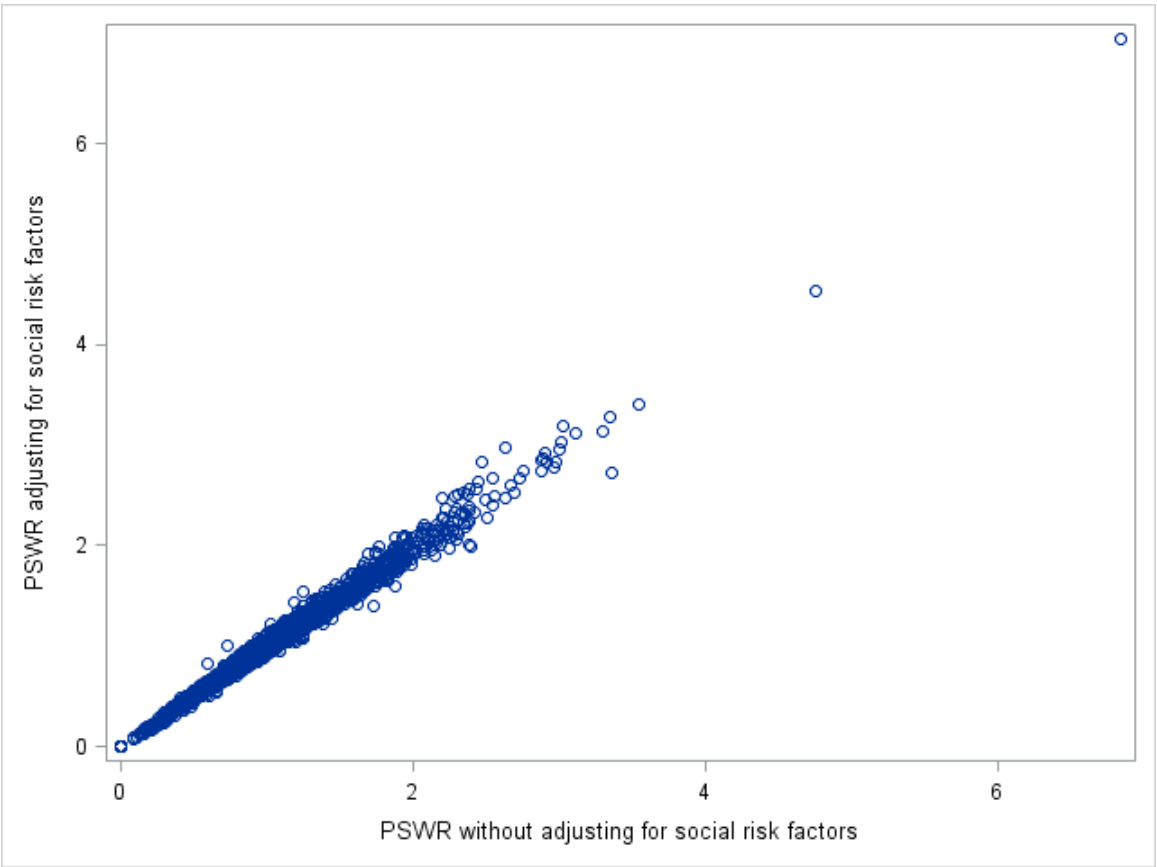
Ethnicity

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

*	*	PSWR without sex	PSWR without sex	PSWR without sex	PSWR without sex
*	*	Better than Expected	As Expected	Worse than Expected	Total
PSWR with sex	Better than Expected	78	0	0	78
PSWR with sex	As Expected	2	1897	1	1900
PSWR with sex	Worse than Expected	0	2	39	41
PSWR with sex	Total	80	1899	40	2019

*	*	PSWR without race	PSWR without race	PSWR without race	PSWR without race
*	*	Better than Expected	As Expected	Worse than Expected	Total
PSWR with race	Better than Expected	70	15	0	85
PSWR with race	As Expected	10	1867	3	1880
PSWR with race	Worse than Expected	0	12	37	49
PSWR with race	Total	80	1894	40	2014

Table 14: Comparison of performances with and without adjusting for race as a risk factor

*	*	PSWR without ethnicity	PSWR without ethnicity	PSWR without ethnicity	PSWR without ethnicity
*	*	**Better than Expected**	**As Expected**	**Worse than Expected**	**Total**
PSWR with ethnicity	**Better than Expected**	66	2	0	68
PSWR with ethnicity	**As Expected**	14	1892	4	1910
PSWR with ethnicity	**Worse than Expected**	0	4	36	40
PSWR with ethnicity	**Total**	80	1898	40	2018

*Cell intentionally left blank.

Although there are differences in waitlisting by sex, race, and ethnicity, 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 0.72, meaning that the model correctly ordered 72% 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]

PSWR is based on time-to-event outcomes and a survival regression model is used instead of logistic regression model for binary outcome. Therefore, we did not provide the Hosmer-Lemeshow statistics. Instead, we provided the risk decile plots based on the statistical risk model in 2b.29. The plot shows that the risk factors in the model are discriminating well between patients. There is good separation among all 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.
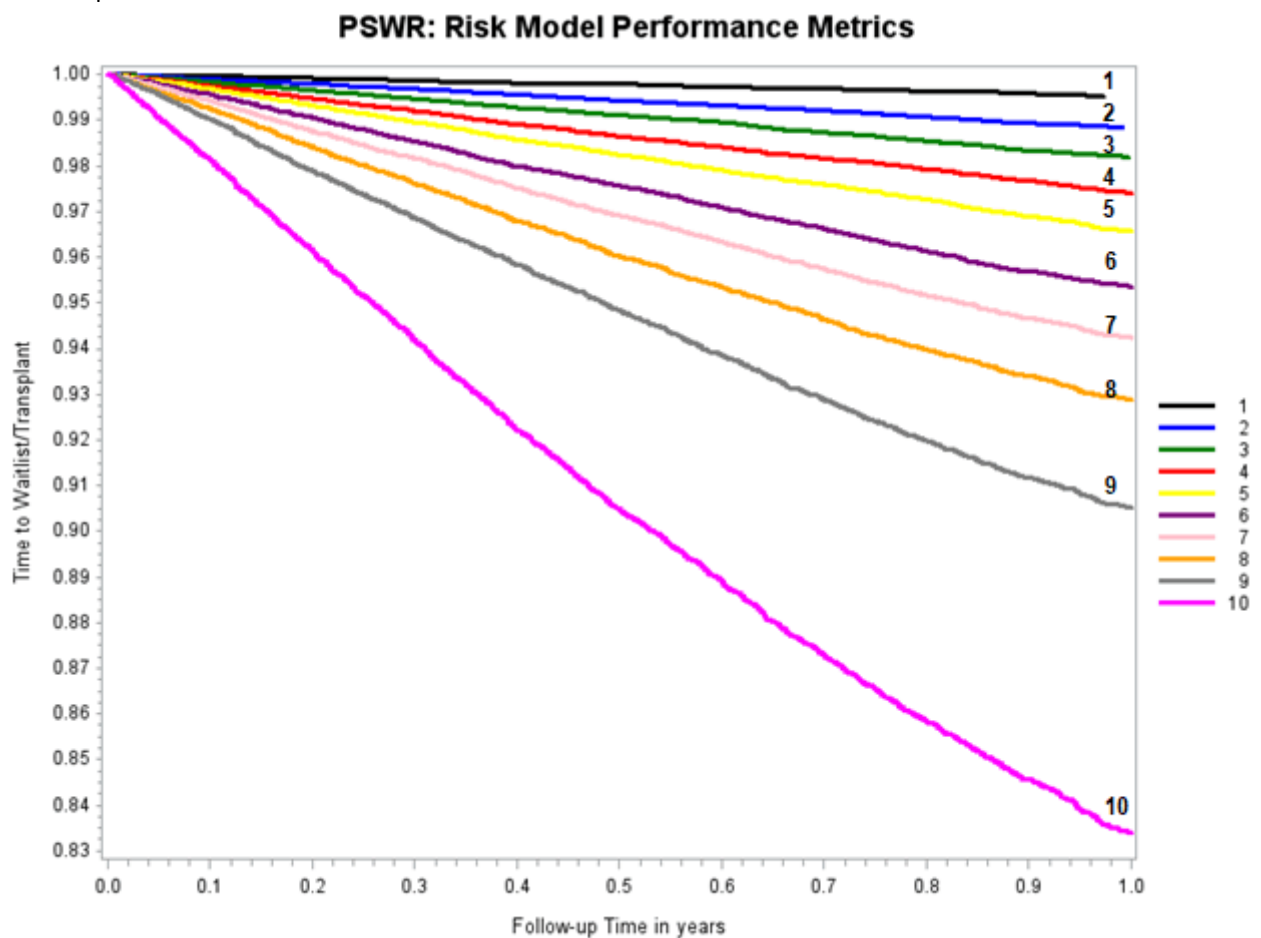
[Response Ends]

2b.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 PSWR

PSWR: Risk Model Performance Metrics

[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

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

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

[Response Begins]

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

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

[Response Ends]

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

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

[Response Begins]

 ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

[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]

Developing an eCQM was outside the scope of this project, based on the available data sources.

[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: https://mmshub.cms.gov/sites/default/files/TEP-Summary-Report-Practitioner-Level-Measurement-Kidney-Transplantation.pdf). 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.

[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: https://mmshub.cms.gov/sites/default/files/TEP-Summary-Report-Practitioner-Level-Measurement-Kidney-Transplantation.pdf.

[Response Ends]

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

[Response Begins]

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: https://mmshub.cms.gov/sites/default/files/TEP-Summary-Report-Practitioner-Level-Measurement-Kidney-Transplantation.pdf). The TEP was comprised of stakeholders representing nephrologist (relevant directly to the target of the measure) and dialysis patient perspectives.

[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 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.

[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]

3695: Percentage of Prevalent Patients Waitlisted (PPPW)

[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]