NATIONAL QUALITY FORUM
Driving measurable health improvements together

# MEASURE WORKSHEET

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

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

## Brief Measure Information

**NQF #:** 3694

**Measure Title:** Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW)

**Measure Steward:** Centers for Medicare & Medicaid Services

**Brief Description of Measure:** This measure tracks the percentage of patients in each dialysis practitioner group practice who were on the kidney or kidney-pancreas transplant waitlist in active status. Results are averaged across patients prevalent on the last day of each month during the reporting year.

The proposed measure is a directly standardized percentage, which is adjusted for covariates (e.g., age and risk factors).

**Developer Rationale:** A measure focusing on 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 kidney transplant, 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].

This measure focuses specifically on the prevalent dialysis population, examining waitlisting monthly in active status for each patient. As this measure assesses monthly waitlisting in active status of patients, it also evaluates and encourages maintenance of patients on the waitlist which is important given the long duration most patients have to wait to eventually access a deceased donor transplant (national median of roughly 4 years) [6]. In particular, maintenance of active status requires ongoing attention by dialysis practitioners to optimizing the health of patients, to ensure sustained suitability for transplant waitlisting. Maintenance of active status on the waitlist is additionally important given demonstrated disparities [7] and positive association with subsequent transplantation [8]. This is an important area to which dialysis practitioners can contribute through ensuring patients remain healthy and complete any ongoing testing activities required to remain active on the wait list. In contrast to this measure, the First Year Standardized Waitlist Ratio focuses solely on new waitlisting and living donor kidney transplants to incentivize early action, rather than ongoing maintenance on the waitlist, as this measure does.

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. United States Renal Data System. 2020 USRDS Annual Data Report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2020.

7. Kulkarni S, Ladin K, Haakinson D, et al. Association of Racial Disparities With Access to Kidney Transplant After the Implementation of the New Kidney Allocation System. JAMA Surg 2019; 154(7):618-625.

8. Grams, M. E., Massie, A. B., Schold, J. D., Chen, B. P., & Segev, D. L. (2013). Trends in the inactive kidney transplant waitlist and implications for candidate survival. American Journal of Transplantation, 13(4), 1012-1018.

Abstract: In November 2003, OPTN policy was amended to allow kidney transplant candidates to accrue waiting time while registered as status 7, or inactive. We evaluated trends in inactive listings and the association of inactive status with transplantation and survival, studying 262,824 adult first-time KT candidates listed between 2000 and 2011. The proportion of waitlist candidates initially listed as inactive increased from 2.3% prepolicy change to 31.4% in 2011. Candidates initially listed as inactive were older, more often female, African American, and with higher body mass index. Postpolicy change, conversion from initially inactive to active status generally occurred early if at all: at 1 year after listing, 52.7% of initially inactive candidates had been activated; at 3 years, only 66.3% had been activated. Inactive status was associated with a substantially higher waitlist mortality (aHR 2.21, 95%CI:2.15-2.28, $p<0.001$) and lower rates of eventual transplantation (aRR 0.68, 95%CI:0.67-0.70, $p<0.001$). In summary, waitlist practice has changed significantly since November 2003, with a sharp increase in the number of inactive candidates. Using the full waitlist to estimate organ shortage or as a comparison group in transplant outcome studies is less appropriate in the current era.

**Numerator Statement:** The numerator is the adjusted count of patient months in which the patient at the dialysis practitioner group practice is on the kidney or kidney-pancreas transplant waitlist in an active status as of the last day of each month during the reporting year.

**Denominator Statement:** All patient-months for patients who are under the age of 75 in the reporting month and who are assigned to a dialysis practitioner group practice according to each patient's treatment history during a given month during the reporting year.

**Denominator Exclusions:** Patients who were at age 75 or older in the reporting month

- Patients who were admitted to a skilled nursing facility (SNF) during the month of evaluation were excluded from that month;
- Patients who were admitted to a skilled nursing facility (SNF) within one year of dialysis initiation according to form CMS-2728
- Patients determined to be in hospice were excluded from month of evaluation and remainder of reporting period
- 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 are not excluded from the measure. All patients who meet the denominator inclusion criteria are included and used to model a given dialysis practitioner group's expected waitlist rate. If a dialysis practitioner group has fewer than 11 patients, then the dialysis practitioner group is excluded from reporting outcomes.

**Measure Type:** Outcome

**Data Source:** Registry Data, Claims

**Level of Analysis: Clinician:** Group/Practice

## Preliminary Analysis: New Measure

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

**The developer provides the following description for this measure:**

- This is a new outcome measure at the group/practice level that tracks the percentage of patients in each dialysis practitioner group practice who were on the kidney or kidney-pancreas transplant waitlist in active status and are in optimal health to receive a transplant.
- The developer provides a logic model that outlines the transplant evaluation process and posits that being waitlisted is an outcome as it represents a desirable change in health status for patients on dialysis, indicating achievement of a health condition conducive for kidney transplantation. The logic model presents the path for patients to achieve waitlisting status which includes eligibility assessment, referral, evaluation and optimizing health and functional status and finally maintaining waitlist status while waiting for a transplant.

**Summary:**

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

***Question for the Committee:***

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *Is a relationship demonstrated between waitlisting and at least one healthcare structure, process, intervention, or service?*
- *Is the evidence directly applicable to the outcome being measured?*

**Guidance from the Evidence Algorithm**

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

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

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

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

- The developer presented an analysis of descriptive statistics for the percentage of prevalent patients waitlisted in active status (aPPW)
    - In 2019, the measure range was 0.0-70.4, with a mean of 12.3 percent, for dialysis practitioner group practices, with at least 11 patients. The interquartile range (Q3-Q1) is 7.3 percent. The bottom quartile of practitioner group practices having 7.3 percent or less of prevalent patients waitlisted vs. the top quartile of practitioner group having 15.6 percent or more of their prevalent patients waitlisted.

**Disparities**

- The developer presented waitlisting status by race, ethnicity and sex for the same sample as presented above for performance gap.
    - Mean waitlisting performance was highest for asian/pacific islanders (18.4 percent) and lowest for native american/alaskan natives (6.9 percent).
    - Non-hispanics had a lower waitlisting percentage on average (11.8 percent) than hispanics (14.5 percent).
    - Black (12.0 percent) and White (12.0 percent) had similar waitlisting percentages compared to the mean across the entire sample (12.3 percent).
    - Females had a mean waitlisting percentage of 10.9 percent and males had a mean of 13.3 percent.

***Questions for the Committee:***

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

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

## Committee Pre-evaluation Comments:

**1a. Evidence**

- Multiple studies provide evidence on the link to positive outcomes and value of waitlisting and processes leading to waitlisting
- Evidence presented includes benefits of kidney transplantation, factors associated with differences in access to waitlisting and/or transplantation, and associations between transplant education and waitlisting. Of note, evidence includes a study of inactive waitlisting published in 2013, prior to 2014 policy revisions.
- The Evidence to support this measure is low. Evidence rating low. The construct of this measure is predicated on the concept that the care provided by the nephrology practice group is the primary driver of a patient being maintained in an active status on the transplant waiting list. However, minimal evidence is presented to substantiate this claim. Other evidence denotes that the transplant waiting list is maintained by the transplant center, not by the nephrology practice group. When classified in an “inactive” status patients continue to accrue waiting time, but are not eligible for an immediate transplantation, and based on available data roughly 1/3 of the total patients listed for transplantation are, at any time, on the inactive list. Numerous publications have evaluated methods that transplant teams can use to systematically review patients on the inactive list (known as status 7 by UNOS) and reclassify them to a more appropriate status. For example, Kataria and colleagues [The impact of systematic review of status 7 patients on the kidney transplant waitlist. BMC Nephrology, April 29, 2019], noted that with close review of their transplant wait list, 18% of the formerly inactive patients were able to be activated and transplanted. The same article noted that reasons for inactive

status included cardio-pulmonary disease, psychosocial factors, and insurance issues. The authors noted that "successful activation of the candidates requires vigorous efforts from the transplant center." These findings were similar to the findings of Shafi and colleagues [Temporary Inactive Status and Renal Transplant Waiting List: Causes, Risk Factors, and Outcomes Transplant Proceedings June 2012], who noted that common causes for transplant inactivation included cardiovascular disease, psychosocial factors and suspected malignancy. Project COIIN (Collaborative Innovation and Improvement Network) is a collaborative project directed by Health Resources and Service Administration for the Organ Procurement and Transplantation Network (OPTN), which focuses on improved waitlist management by transplant centers. This collaborative has been focusing on building a more active waitlist, and according to participants this meant moving some patients to an "inactive status," which resulted in a shorter kidney wait list but one that was more active overall. A 17 percent increase in kidney transplants was reported when active list management, with patient review and reclassification, was performed by the transplant center team. [Featured by UNOS; https://unos.org/news/improvement/achieving a shorter more active waitlist/Sarah Moriarty: Achieving a Shorter, More Active Kidney Waitlist February 7, 2019.] As a corollary, patients can be reclassified as "inactive" for a number of reasons not directly linked to the overall care being provided by the nephrology practice group. For example, patients can be moved to an inactive status because of an acute illness, patient injury, patient travel and inability to be present for transplantation, psychosocial issues and insurance issues. None of these is linked to the care provided by the Nephrology practice. Thus, the evidence suggests that transplant list management is done by the transplant team and that reclassification of the list which can include moving some patients to an "inactive status," can increase the number of transplantations performed

- There appears to be empirical data to support the association between practitioner process and outcome measure. However, I do think there is limited actions that dialysis practitioners can implement to keep patients on the waitlist, if their health deteriorates. This measure makes the assumption that presently, dialysis providers are not already doing everything they can for the health of their patients, an assumption I feel is false.
- Same evidence as offered for 3695 without additional empirical evidence that incentivizing practitioners influences active waitlist status.
- Fail, data is derived from a logical model and there may be unmeasured confounders which are not accounted for
- The evidence presented provides a rationale for waitlisting as an outcome but is tangential to the intermediate outcome being measured. Specifically, the measure is intended to assess practitioner/group performance, but the measure and supporting evidence fail to acknowledge that waitlisting per se is a decision made by the transplant center and is beyond the locus of control of the providers targeted in these measures. While a referral to a transplant center and initiation or even completion of the waitlist evaluation process might be appropriate performance measures at this level of analysis, these newly proposed practitioner/group level waitlist measures are not. The transplant center decides whether a patient is placed on a waitlist, not the practitioner or group practice. There are many potential obstacles and delays in the evaluation process with multiple parties that have nothing to do with the treating practitioner or group. A patient's status on the waitlist (active/inactive) can change frequently within the transplant centers and can be notoriously difficult to track, adding another concern with the aPPPW, in particular; this reality will further compromise the measure's validity and render the information it provides flawed, at best—and potentially harmful, should patients and providers act on the assumption of accuracy. Again, penalizing a practitioner/group practice each month for delays that are beyond their control and/or for inaccuracies in data maintained by a different entity is wholly inappropriate.
- Adequate evidence provided focusing on association between dialysis clinician practice/processes and waitlisting

- Evidence supports the need to get patients waitlisted in order to get transplanted as transplanted patients have improved outcomes
- I am not aware of other studies.
- Health outcome measure. New outcome measure that tracks the percentage of patients in each dialysis practitioner group practice who were on the kidney or kidney-pancreas transplant waitlist in active status and are in optimal health to receive a transplant. Developer provides a logic model that outlines the transplant evaluation process and posits being waitlisted is an outcome as it represents a desirable outcome in patients health status. The logic of the model presents the path for patients to achieve waitlisting status which includes eligibility assessment, referral, evaluation and optimizing health and functional status and finally maintaining a waitlist status while waiting for a transplant. Empirical patient support on the value of waitlisting which came from a study published in the American Journal of Transplantation. Several studies were cited that provide strong support for the association between processes under dialysis practitioner control and waitlisting. Outcome measure that assesses performance on a health outcome and the relationship between the measured health outcome and at least one health action is demonstrated by empirical data. Two tep panels were convened to discuss potential measures directed at improving access to kidney transplantation...in 2015 and 2021. The teps comprised of relevant stakeholders, including dialysis nephrologists, transplant nephrologists, transplant surgeons, social workers, researchers and notably patient representatives with a history of ESRD. 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 and the practitioner level for the 2021 tep.
- Compelling supportive evidence
- Passes but some providers may report incorrectly (referral for transplant and waitlisting are not the same - as seen in some practices).
- True
- No concerns
- No concern.
- The developers provide a logic model that being wait listed is an outcome that leads to the desirable change in health status via transplantation. There are several studies that provide support for the association between waitlist and the intervention of kidney transplant. It is directly applicable to the outcome being measured.
- For the outcome of waitlisting, the measure is supported by evidence.

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

- The developer demonstrates opportunity for improvement in the area of waitlisting. Disparity data provided.
- 2,276 dialysis practitioner groups and 280,855 patients included. Performance on aPPPW (2019): median 11.7% (lower quartile 8.3%, upper quartile 15.6%). Differences in performance based on race, ethnicity, and sex.
- The data presented show that as constructed 92.4% of practices (2104 practices) were preforming "as expected," and 5% (113 practices) were performing lower than expected. This suggest that the measure will not be very effective as a performance improvement outcome measure
- Observational performance gaps are noted
- 15.6% v. 24.7% in lowest to best performing deciles. Same disparities demonstrated as 3695.
- Insufficient due to unmeasured confounders and patient preference.

- Gaps in both provider performance and between racial and ethnic groups is presented; however, there is no evidence provided to support that performance on the measure is not more significantly linked to transplant center practices than to those of the treating practitioner/group.
- Low overall national performance so performance gap exists; data on disparities in care provided based on race, ethnicity, and sex
- Appropriate info provided
- I think it would help patients, but I would need more explanation.
- Developer presented an analysis of descriptive statistics for the percentage of prevalent patients waitlisted in active status...2019 the measure range was 0.0-70.4 with a mean of 12.3 percent for dialysis practitioner group practices with at least 11 patients. The interquartile range is 7.3 percent. The bottom quartile of practitioner group practices having 7.3 percent or less of prevalent patients waitlisted vs the top quartile of practitioner group having 15.6 percent or more of their prevalence patient waitlists. Developer presented waitlisting by race, ethnicity and sex for the same sample as presented for performance gap...Waiting lists performance was highest for Asian/pacific islanders 18.4 percent and lower for native American/Alaskan natives 6.9 percent. Non-Hispanics had a lower waiting listing percentage on average 11.8 % than Hispanics 14.5 %.  Black 12% and white 12% ....mean across the entire sample 12.3%....females had a mean waitlisting percentage of 10.9 percent and males had a mean of 13.3 %
- Meaningful gap demonstrated
- There is a high opportunity for improvement. Unbiased education needed to patients. Opportunities to "interview" multiple facilities needed.
- True
- Insufficient
- Evidence was presented identifying gaps and disparities.
- Performance gap was provided and showed difference in wait list status by race, ethnicity, and sex. There is a moderate gap.
- There is a modest gap in performance: 92.4% of facilities perform as expected

## Criteria 2: Scientific Acceptability of Measure Properties

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

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

- The SMP Passed on Reliability with a score of: H-5; M-3; L-0; I-2
- The SMP Did Not Reach Consensus on Validity with a score of: H-0; M-6; L-4; I-0

### 2a. Reliability: Specifications and Testing

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

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

**Specifications:**

- Measure specifications are clear and precise.

**Reliability Testing:**

- Reliability testing conducted at the Accountable Entity Level:
    - The developer conducted testing using inter-unit reliability (IUR) with a bootstrap (n=100) approach. This approach utilizes a resampling procedure to estimate the within facility variation that cannot be directly estimated by ANOVA.
        - The developer calculated a IUR value of 0.93 for the measure, which indicates that 93 percent of the variation in the measure can be attributed to the between-facility differences and 7 percent to the within-facility variation.
        - The developer notes that this IUR implies a high degree of reliability and can reliably detect differences in performance scores across practitioners.
        - Dialysis practitioner group practices with less than 11 eligible patients were excluded from this calculation.

**SMP Summary:**

- Reliability passed the SMP's preliminary review and was therefore not discussed at the meeting. SMP did not report any significant concerns regarding reliability during their preliminary review.

***Questions for the Committee regarding reliability:***

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

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

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

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

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

**Validity Testing**

- Validity testing was 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 mortality and overall transplant rates among all patients attributed to the practitioner groups.
    - The developers examined the Spearman correlation between the practitioner group measure value and each of the outcomes respectively.
    - The dialysis practitioner group level average mortality was 17.8, 18.3, and 19.2 deaths per 100 patient-years for each of the 3 tertiles (T1 to T3) based on their performance on the PPPW (T1 to T3, from highest to lowest waitlisting), respectively (trend test p=0.002). The Spearman correlation coefficient was: -0.083 (p<0.0001).

  - The dialysis practitioner group level average transplant rate is 5.0, 4.2, and 3.1 transplants per 100 patient-years for T1, T2, T3 groups, respectively (trend test p<0.002). The Spearman correlation coefficient is 0.279 (p<0.0001).
  - The developer noted that higher aPPPW performance correlated with higher transplant rate, and the relationship with mortality was also as expected by the developer, and statistically significant, with numerically lower mortality with higher performance on the aPPPW measure although the magnitude of the association was smaller than for transplant rate.

**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 they do not exclude patients from dialysis practitioner groups with fewer than 11 attributed events.
- The developer reported that the number of patients before exclusions was 3,561,019 and after exclusions it was 2,541,229; 28.6 percent of patients were excluded.
- The developer also reported the following frequencies for each excluded variable:
  - Age greater than or equal to 75 – 766,648 patient months (21.5 percent)
  - Nursing home from CMS-2728 – 26,618 patient months (0.8 percent)
  - Nursing home from nursing home history file – 302,227 patient months (8.5 percent)
  - Hospice – 14,581 patient months (0.4 percent)
  - Dementia-152,951 patient months (4.3 percent)
- Overall measure scores were changed moderately by the exclusions.
  - The average waitlisting percentage increased from 9.0 percent before exclusions to 12.3 percent after exclusions.
- The developer stated that though performance scores are moderately affected by exclusions, practitioner group performance rankings are minimally affected. The developer deemed the exclusions as important as they represent a group of patients highly unlikely to be suitable for transplant waitlisting. The developer also noted that there is fair degree of variation in the percentage of patients excluded across practitioner groups. Lastly, the developer notes that the data to determine exclusions is readily available and therefore adds no additional burden.

**Risk-Adjustment**

- 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. Three categories were chosen: social risk, functional risk, and medical/clinical risk.
- 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.

**Meaningful Differences**

- The developer calculated the nominal p-value and converted them to z-scores to derive the mean and variance of normal empirical null distribution. The empirical null distribution is then used to calculate the p-value for each dialysis practitioner.
  - Dialysis practitioner group practices (92.4 percent) had an aPPPW that was “As Expected”.
  - Dialysis practitioner group practices (2.6 percent) have a aPPPW that was “Better than Expected”,

    - Dialysis practitioner group practices (5.0 percent) have a aPPPW that was “Worse than Expected”.
- Performance on waitlisting in active status varied widely (from 3.4 percent of patients waitlisted in the worse than expected category, to nearly 20 percent in the better-than-expected category), suggesting that differences are also clinically meaningful.

**Missing Data**

- No missing values for age, date of birth, and reporting month were found.
- A missing CMS-2728 form was assessed and found to be low at 1.11 percent.
- Patients with missing CMS-2728 form have a lower odd of waitlisting compared to those without a missing CMS-2728 form (OR =0.510; 95 percent CI = 0.490, 0.531).

**Comparability**

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

**SMP Summary:**

- SMP panel members expressed concerns about the non-independence of patient months in the model. The developer explained that these were accounted for using the empirical null method.
- SMP members also noted that this measure may be better characterized as a process measure. The developer advised that they consider the measure to be an outcome measure, because it represents achievement and maintenance of health suitable for transplant, which is dependent on dialysis practitioner interventions
- SMP discussed the risk adjustment model, specifically the use of concurrent risk factors, transplant center characteristics, and socio-economic factors, such as area deprivation index.
    - SMP noted the potential of adjusting away some of the transplant center effects by including transplant center characteristics in the risk adjustment model. The developer explained that their TEP advised that adjustment was warranted, so that providers disproportionately caring for socially vulnerable patients are not unfairly penalized.
    - SMP also noted the lack of validation using an external data set of the risk adjustment model. The developer advised that they did not perform validations with an external data set, as national data was used which would target the measure population.
    - SMP sought clarity on whether the comorbidities are limited to claims prior to the measurement period. This is important to limit the risk factors to those that were present at the start of care. The developer advised that Medicare claims from the year prior to the reporting period were used for the prevalent comorbidities.
    - SMP members noted an inconsistency between the risk model equation and the description, which includes two-way interaction terms. The developer advised that in the formula, they denoted alpha for transplant center random effects and Z for patient characteristics; Z includes both patient characteristics and transplant center fixed characteristics. The inclusion of the sentence "two-way interactions were examined and selected for the final model based on both the magnitude and statistical significance of the estimates" was an error, as the final model doesn't include interactions.
    - SMP had concerns regarding the inclusion of social risk adjustment in the models. The developer advised that their decision to propose this measure is in large part motivated by a desire to reduce such disparities, and the factors chosen do have a conceptual basis in that they are proxies for financial and social resources that can affect success following

transplantation. Additionally, a Technical Expert Panel consisting of a range of stakeholders, including several patients with ESRD, discussed these issues and were in consensus about the need for social risk adjustment. A dominant concern was that in the absence of such adjustment, dialysis practitioners caring for a disproportionate share of socially vulnerable patients may inappropriately be penalized by the measure, leading to unintended adverse consequences in terms of access to care for these patients.

- SMP members were concerned that the measure may not account for the uncertainty of the estimate if point estimates are used. SMP panel members asked if the score will be used as a point estimate or as a differentiation between categories of provider groups (average, better than average, or worse than average). While the developer responded they will use these to identify those facilities that are significantly different from the average, SMP members were concerned that if point estimates are used, the measure may not account for the uncertainty of the estimate. A SMP member noted that the better-than-expected performance band is not very good on an absolute basis.
- SMP questioned whether having two measures for waitlisting (waitlisted or waitlisted with active status) is necessary. The developer clarified that active status is a subset of waitlisting and requires active maintenance of health status. Furthermore, the developer noted that NQF #3695 is a broader measure that captures the psychological benefit of being on the waitlist.
    - SMP noted that the Standing Committee should consider whether both measures are clinically necessary.

***Questions for the Committee regarding validity:***

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Does the Committee have any concerns about the use of socio-economic factors in the risk adjustment model?*
- *Are there concerns about the overlap between this measure and 3694?*
- *Does the Committee have any concerns about the non-independence of patient months in the model?*
- *Does the Committee believe there are meaningful differences in performance on this measure with 92.4 percent of practices performing as expected?*

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

- SMP did not reach consensus.

## Committee Pre-evaluation Comments:

**2a1. Reliability – Specifications**

- No concerns noted
- Data elements are clearly defined.
- The c score is as expected based on the measure definition
- There should be adjusting for the transplant center most likely used by that practitioner, as waitlist times and percentages over time differ
- No concerns re: definitions implementation
- Insufficient due to unmeasured confounders and patient preference.
- Data elements and logic/calculation algorithm are clearly defined, but it’s unlikely that the measure could be accurately implemented, given the accuracy and tracking issues around waitlisting active/inactive status.

- Specifications seem all OK
- No concerns
- I'm not sure all of the data was provided.
- Measurements specifications are clean and precise. Reliability testing conducted at the accountable entity level....using inter-unit reliability with a bootstrap n=100 approach. This approach utilizes a resampling procedure to estimate the within facility variation that cannot be directly estimated by ANOVA. IUR value of 0-.93 for the measure which indicates that 93 percent of the variation in the measure can be attributed to the between-facility differences and 7 percent to the within-facility variation. This IUR implies a high degree of reliability and can reliably detect differences in performance scores across practitioners. Dialysis practitioner group practices with less than 11 eligible patients were excluded from calculation. The scientific methods panel is satisfied with the reliability testing for the measure.
- All well defined
- Moderate reliability
- True
- No concerns
- "Adjusted" patients months and "given" months are not clear.
- Measure is precise and defined. There should not be any problem in implementing this measure.
- Adequate

**2a2. Reliability – Testing**

- Reliability was achieved by the developers testing methods. showing high reliability of .93
- IUR 0.93.
- As above
- No concerns
- See prior discussion (3695) regarding: specs and risk adjustment.
- Yes, Insufficient due to unmeasured confounders and patient preference.
- Yes. While the overall IUR across all facilities is good at 0.93, stratification of reliability scores by provider size was not detailed. Because of this, it's impossible to determine how widely reliability varies across the spectrum of provider/group sizes. As has often been the case with other CMS measures, reliability for small providers might be substantially lower than the overall IUR, effectively rendering the metric meaningless for use in performance measurement in this group. Request CMS provide data demonstrating reliability for all providers by detailing IURs by provider/group size.
- IUR of 93% so readily meets reliability testing standards
- No concerns
- Yes.
- No
- No
- There is a possibility that provider groups could not report accurately making reliability an issue.
- Appropriate
- Yes
- No concerns.
- No significant concern in reliability of the measure.
- No

**2b1. Validity – Testing**

- It seems like the validity is based solely on the correlation of aPPPW performance and mortality only. It was noted there are other factors that could relate to mortality.
- Higher aPPPW correlated with higher transplant rate and lower mortality rate.
- As constructed the validity of the measure is of low value with regard to the care rendered by the nephrology practice group.
- Some concerns, see below
- Very small number of worse than expected.
- Yes Insufficient due to unmeasured confounders and patient preference.
- No additional concerns beyond those raised by the SMP on risk adjustment (see Question 9).
- No significant concerns -- higher measure performance correlated with higher transplant rates and lower mortality
- No concerns
- Yes.
- Yes....SMP did not reach consensus.
- No
- I don't believe the measure is needed since almost 93% of providers are performing as expected.
- Appropriate
- Yes - uncertain validity
- Mortality and transplant rates are appropriate measures to establish a moderate to low rating.
- I have no concerns with regards to validity
- No

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

- Excluded population is appropriate. Risk adjusted model does not include ethnicity, sex as it was determined they did not show impact on risk adjustment
- Exclusions include age 75 or older, SNF residence (for reporting month), SNF admission in the one year prior to dialysis (2728) [only used for incident patients], hospice (reporting month), dementia in the prior year [AHRQ CCS]; 28.6% of patient-months excluded with age >=75 and nursing home residence being the largest contributors to exclusions. Median performance meaningfully different after application of exclusions (aPPPW median 8.5 before and 11.7 after exclusions). Model included adjustment for age, ADI, DE, diabetes as cause of ESRD, comorbidities at ESRD incidence, 64 prevalent comorbidities from claims (IP or OP), and transplant center characteristics (waitlist mortality rate and transplant rate). Risk adjustment for prevalent comorbidities includes different types of cancer (liver, lung, bladder, bone, lymphoma) and cardiovascular disease. Clarify attribution of transplant center characteristics (per measure specifications, patients assigned to a transplant center based on historical waitlisting patterns in their residential zip code).
- The move of a patient to an "inactive status" is dependent on the actions of the transplant team. The processes around those decisions are not evaluated by this measure. The data drawn from the 2728 forms only reflect the patient's conditions at the time of enrollment for dialysis and are not updated to reflect the patient's current condition. Subsequent health events may, or may not, result in the submission of Medicare claim. Thus, information required for appropriate risk adjustment may not be available
- Excluding new onset or diagnosis of cancer is needed to exclude those who drop off the waitlist due to this
- See response for 3695. Social risk not incorporated

- No
- No concerns with exclusions. The risk model appears to fit well, with a c-statistic of 0.763; however, the SMP's concerns on the inclusion of social risk variables in the final model are noted. A discussion among Standing Committee members would be helpful on this issue.
- Scientific panel raised some issues about risk adjustment but difficult to understand specifically what these concerns were
- No concern
- I would need more information
- Appropriate risk adjustment approach appropriately developed and assessed.
- Exclusions appropriate
- The risk adjustment should be helpful to decrease disparities.
- True
- Some risk adjust made, but unclear to me if all appropriate risk adjust has been proposed.
- Social risk factors of ADI and dual eligibility questionable. Why adjust for missing 2728?
- Exclusions are consistent and reasonable as those patients would not be waitlisted or even referred for transplantation.
- Adjustments are appropriate

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

- Missing data does not seem to be an issue
- SMP did not reach consensus on validity.
- 2b4 meaningful differences in quality will be difficult to detect as most practice groups performed "as expected"
- Transplant center will most likely decide which patients should or should not remain on their waitlist, so transplant center adjustor is needed
- See response for 3695
- Yes, Insufficient due to unmeasured confounders and patient preference.
- Scores differentiated as "as expected," "better than expected," and "worse than expected." No concerns with approach. No concerns around missing data.
- With current very poor national performance, there is less meaningful difference about quality identified by the measure -- but with improvement in national performance (perhaps as a result of the measure?), may become more meaningful
- No concerns
- I'm not sure all of the data was provided. Yes
- No assessment of threats to validity. Exclusions well supported. No missing data does not constitute a threat to the validity of this measure
- No
- High rating on feasibility
- Yes
- Uncertain
- No concerns.
- I do not believe there is any threat to validity, and I believe there would be meaningful differences noted. Missing data should be inconsequential.
- Geographic variation influences performance and is not accounted for

## Criterion 3. Feasibility

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

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

***Questions for the Committee:***

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

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

## Committee Pre-evaluation Comments:

**3. Feasibility**

- Data can be readily obtained and collected.
- No concerns about feasibility
- Moderate as noted up to date data needed for risk assessments may not be readily attainable
- Seems feasible
- No concerns
- Patient preference is not currently recorded
- No concerns with feasibility for this measure.
- Very feasible
- I think it may be challenging to determine waitlisted vs active on waitlist as the dialysis center doesn't always control status of listing
- I'm not sure all of the data was provided.
- Developer states that the data elements are generated or collected and used by healthcare personnel during the provision of care. Further, the data elements are coded by someone other than the person obtaining original information. Measure relies on data elements that are defined in a combination of electronic sources. Rating for feasibility was rated high. Practitioner group level results has not been disseminated to those being measured as part of the development process. Physician group results have not been disseminated to those being measured as part of the development process. The developer sought input from a TEP during development and those deliberations were open to the public.
- High feasibility
- Data elements are easily captured and available in electronic version. High level of feasibility.
- Appropriate
- Feasible

- No concerns.
- The data elements are routinely generated and stored and used by the healthcare personnel in rendering care.
- No issues

## Criterion 4: Use and Usability

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

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

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

**Current uses of the measure**

Publicly reported? ☐ Yes ☒ No

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

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

**Accountability program details**

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

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

**Feedback on the measure by those being measured or others**

- Practitioner group level results have not been disseminated to those being measured as part of the development process.
- Physician group 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 developer advised that this measure reflects the input from the TEP on how the construction of the facility level measures should be revised to be adapted to the practitioner level and addresses the concerns raised about appropriate risk adjustment.

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

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

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

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

**Improvement results**

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

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

**Unexpected findings (positive or negative) during implementation**

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

**Potential harms**

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

***Questions for the Committee***:

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

## Committee Pre-evaluation Comments:

**4a. Use**

- Currently not in use. Plans for use in CMS programs. Developer did indicate the incorporation of TEP input received previously
- May be used in the future within Quality Payment Programs
- New measure TEP done. Other user group feedback not available/completed. Use as accountability measure is expected
- Reporting performance results has usually been at individual practitioner level, at each dialysis facility. There will need to be a better reporting method to provide results at the entire practitioner group level
- Not in use
- Insufficient due to unmeasured confounders and patient preference.
- This is a new measure, not yet being publicly reported. It is intended for use in publicly reported programs, thus results will be disclosed and available to the broader public if the measure is endorsed

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

- Potential implementation seems credible
- Not publicly reported or used in accountability program but planned to be used in accountability program. Feedback is essential for programs to improve. Need to develop method for feedback and opportunity to provide feedback.
- I would need more information
- Used Accountable entity level testing....moderate.
- Feedback incorporated
- Pass
- True
- Unclear
- Not reported publicly. No feedback provided due to the lack of dissemination to the practice groups. Did receive feedback from TEP.
- The measure will be used to report to the public in a quality program- desperately needed as well.
- Not reported; potential for future use

**4b. Usability**

- Developer doesn't state a concern for harms outweighing benefits.
- Discuss factors included in the risk-adjustment and potential for unintended consequences.
- The measure ascribes accountability and responsibly for maintaining active waitlist status to the nephrologist. However, the nephrologist is not involved in that decision-making process and factors apart from the patient's medical condition impact the transplant team's decision to temporarily move a patient to an 'inactive' status. Thus, the nephrologist may be penalized inappropriately which, in turn, could reduce access to care for these patients
- May be useful. There may not be meaningful interventions, even with this measure, for practitioners to keep patients on the waitlist if their health is deteriorating.,
- If ETC, KCC models and current PPPW reporting haven't improved things, will devoting additional resources distract from other quality activities.
- Insufficient due to unmeasured confounders and patient preference.
- The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. The developer did not provide an assessment of benefits vs. harms. As previously noted, however, a concern with this measure is misattribution and potential penalties levied again practitioners/groups for an outcome that is largely outside their control.
- Seems to meet expectations for usability without harm
- Dialysis practitioner may be penalized if patient isn't deemed ready for transplant by the transplant center. Practitioners can educate and optimize health for transplant, but there are multiple factors that determine if the patient is an appropriate candidate for transplant, many of which are out of the dialysis practitioner's control
- I'm not sure all of the data was provided.
- No intended harm.....all benefits.
- Credible applicability to improve outcomes
- This measure is highly usable in provider practice.
- Appropriate

- Two concerns about usability: first, again, ascription. Practitioner group ascription is not a meaningful level, since that group does not determine who is on the active waitlist. It is the transplant center that makes that determination. Ascription at level of Transplant Center OK, or if you want a practitioner level metric, make it referral to transplant center
- Benefits outweigh harm. Have some concern with inaccurate data on 2728 and possible gaming opportunities.
- The measure can be used to further evaluate practitioner for referring patients for transplant and therefore improving the healthcare of these patients in a timely fashion. There are no unintended consequences.
- Not clear

## Criterion 5: Related and Competing Measures

**Related/Competing measures**

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

**Harmonization**

- N/A

## Committee Pre-evaluation Comments:

**5: Related and Competing Measures**

- None noted. There are multiple waitlist measures but all with different nuances.
- None noted
- Presently, percentage on waitlist is at the dialysis provider level only, I seen no competing issues
- SWR, PPPW (not endorsed)
- Is not harmonized with other measures for renal patients.
- Agree with the SMP’s assessment that two prevalent waitlisting measures are not needed.
- Measure 3695 looks at same metrics but focuses on waitlisting vs active waitlisting. Are both measures needed? If goal is to get more patients transplanted and active listing is only way to actually be transplanted, would be better to only consider 3694?
- 3695 may be a related and competing measure -- not sure if need both waitlisted and waitlisted/active measures
- I would need more information
- No competing measures.
- N/A
- None
- True
- Yes, 3695
- None
- There are no competing measures to date.
- 3689, 3695

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

**Member Expression of Support**

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

**Comments**

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

Practitioner/Group-Level First Year Standard Waitlist Ratio (NQF 3689, CMS) Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted in Active Status (NQF 3694, CMS) Practitioner/Group-Level Percentage Of Prevalent Patients Waitlisted (NQF 3695, CMS) KCP recognizes the tremendous importance of improving transplantation rates for patients with ESRD, but does not support the attribution of successful or unsuccessful waitlisting to dialysis facilities, individual practitioners, or group practices and thus cannot support these measures. KCP believes that while a referral to a transplant center and initiation or even completion of the waitlist evaluation process might be appropriate measures for these levels of analysis that could be used in CMS’s quality programs, the newly proposed practitioner/group level Percentage of Prevalent Patients Waitlisted (PPPW), Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW), and First Year Standardized Waitlist Ratio (FYSWR) measures are not. Waitlisting per se is a decision made by the transplant center and is beyond the locus of control of any of the providers targeted in these measures. In reviewing these measures, we offer the following comments: I. Overarching Concerns Several of KCP’s concerns apply to all three proposed transplant access measures: a. Attribution. As above, we strongly object to attributing successful/unsuccessful placement on a transplant waitlist to dialysis facilities, individual clinicians, or practitioner group practices and believe this is a fatal structural flaw with these measures. The transplant center decides whether a patient is placed on a waitlist, not the facility, practitioner, or group practice. KCP patient members who are transplant recipients have noted there are many obstacles and delays in the evaluation process with multiple parties that have nothing to do with the facility or practitioner—e.g., one patient noted their private pay insurance changed the locations where they could be evaluated for transplant eligibility on multiple occasions, repeatedly interrupting the process mid-stream. Penalizing a practitioner/group practice each month through the PPPW, aPPPW, and FYSWR for these or other delays is inappropriate; such misattribution is fundamentally misaligned with NQF’s first “Attribution Model Guiding Principle,” which states that measures’ attribution models should fairly and accurately assign accountability. KCP emphasizes our commitment to improving transplantation access, but we believe other measures with an appropriate sphere of control should be pursued. For instance, our sister organization, the Kidney Care Quality Alliance (KCQA), has developed a dialysis facility-level Transplant Access Measure Set that will be submitted to NQF for endorsement consideration later this year. The set pairs a referral rate metric with a measure assessing the waitlisting rate specifically among those patients who were referred by the facility within the preceding three years. Because the KCQA waitlisting measure denominator is limited to those patients who were deliberately referred by the dialysis facility within a defined time period, facilities have considerably more agency over the measure than metrics such as the PPPW; this construct will also provide a counterbalance to the referral measure, curbing the tendency to indiscriminately refer patients who are not appropriate transplant candidates, preventing unnecessary patient and transplant center burden. The same approach could be applied at the practitioner/group level. b. Variation in Transplant Center Eligibility Criteria. We also note that criteria indicating a patient is “not eligible” for transplantation can differ by location. For instance, one center might require evidence of an absence of chronic osteomyelitis, infection, heart failure, etc., while another may apply eligibility exclusions differently or have additional or different criteria. The degree to which these biological factors influence waitlist placement must be accounted for in any model for the measure to be a valid representation of waitlisting. c. Stratification of Reliability Results by Group Size and Performance Scores

Absent. We also note that CMS has provided no stratification of reliability scores by provider size for the measures; we are thus unable to discern how widely reliability varies across the spectrum of practitioner/group practice sizes. We are concerned that the reliability for small providers might be substantially lower than the overall IURs, as has been the case, for instance, with other CMS standardized ratio measures. This is of particular concern with the FYSWR, for which empiric testing has yielded an overall IUR of only 0.64—interpreted as "moderate" reliability by statistical convention. To illustrate our concern, the Standardized Transfusion Ratio for Dialysis Facilities (STrR) measure (NQF 2979) also was found to have an overall IUR of 0.60; however, the IUR was only 0.3 ("poor" reliability) for small facilities (defined by CMS as <=46 patients for the STrR). Without evidence to the contrary, KCP is concerned that FYSWR reliability is similarly lower for small groups, effectively rendering the metric meaningless for use in performance measurement in this subset of providers. KCP believes it is incumbent on CMS to demonstrate reliability for all providers by stratifying data by practice size.

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

IPercentage of Prevalent Patients Waitlisted In Active Status (NQF 3694, CMS) KCP has identified two concerns specific to the aPPPW measure: a. Rate vs. Ratio. Notwithstanding our concerns described above, consistent with our comments on other standardized ratio measures (e.g., SHR, SMR), KCP prefers normalized rates or year-over-year improvement in rates instead of a standardized ratio. We believe comprehension, transparency, and utility to all stakeholders is superior with a scientifically valid rate methodology. b. Active Status Data. We also note that a patient's status on the waitlist (active/inactive) can change frequently within the transplant centers and can be notoriously difficult to track. We believe this reality will seriously compromise the measure's validity and render the information it provides flawed, at best—and potentially harmful, should patients and providers act on the assumption of accuracy.

### Comment 3 by: Submitted by David White, American Society of Nephrology

TO: NQF Renal Standing Committee FR: Tod Ibrahim, Executive Vice President, the American Society of Nephrology Members of the National Quality Forum Renal Standing Committee The more than 37,000,000 Americans living with kidney diseases and the 21,000 nephrologists, scientists, and other kidney health care professionals who are members of the American Society of Nephrology (ASN), thank you for the opportunity to comment on the 5 proposed transplantation, vascular access, and modality education measures under consideration: • Facility-Level Standardized Modality Switch Ratio for Incident Dialysis Patients (SMoSR) • Facility-Level Standardized Fistula Rate for Incident Patients (ISFR) • Practitioner/Group-Level First Year Standard Waitlist Ratio (FYSWR) • Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted (PPPW) • Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW) Based on our review, ASN is concerned by several aspects of the measures and offers comment on all five measures submitted to NQF: • Focus on incident maintenance dialysis populations with "stand alone" measures that are independent of measures targeting patients in other stages of kidney diseases such as non-dialysis advanced chronic kidney disease and prevalent dialysis. • Reliance on CMS-2728 data (End Stage Renal Disease Medical Evidence Report Medicare Entitlement and/or Patient Registration) for any risk adjustment including transplant measures • Attribution of measures to dialysis facilities • Lack of adjustment for variables that are critical for patient equity, such as social determinants of health • Focus on dialysis unit-specific measures, without consideration of advanced CKD care and nephrologist-led care Practitioner /Group-Level First Year Standard Waitlist Ratio (FYSWR) Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted (PPPW) Practitioner/Group-Level Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW) While ASN is supportive of these measures for ensuring and promoting equitable access to kidney transplantation, it is important to recognize that the actual waitlisting of patients -- active or inactive -- on the waitlist is beyond the control of dialysis units or individual nephrologists as currently structured. While dialysis facilities and managing nephrologists may be able to exert some influence over several of these factors, this influence is dwarfed by the role of the transplant centers, rendering the attribution misdirected. In order to improve these measures, albeit leaving these still without the proper attribution, it is imperative that the following information be easily and readily accessible to referring physicians and dialysis units: 1. Waitlisting criteria at transplant centers including absolute AND relative contraindications. 2. Clear information on the reasons for declining a patient for listing by transplant centers so that nephrologists can determine if patients would benefit from referral to a different transplant center. 3. Active status on the waitlist needs to be made clearly available to nephrologists and dialysis facilities so that centers and dialysis facilities are immediately aware of when (and why) patients are inactivated on the list. If physicians are going to be held accountable for this, they need to be aware of the status and what needs to be done to be re-activate those patients on the waitlist. 4. “Internal holds” placed on a patient by the transplant center while leaving the patient as active on the waitlist. Differences in how transplant centers use this practice can adversely impact the measure and access to transplant for patients who are on extended periods of internal hold unbeknownst to them. The implementation of these measures should be accompanied by easy and timely access to the status of the patient in the evaluation process and waitlist status. A way to shed light on whether transplant centers are inappropriately using “internal hold” for patients is to share organ offer data with nephrologists and dialysis facilities which would help identify patients who are on internal hold instead of being inactivated. The Health Resources and Services Administration (HRSA) and the Organ Procurement and Transplantation Network (OPTN) need to provide access to waitlist data, information on steps to transplantation from centers, and organ offer data in a manner that is timely, easily accessible, and actionable.

## Scientific Acceptability Evaluation

**Measure Number:** *3694*

**Measure Title:** *Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW)*

**Measure is:**

☒ **New** ☐ **Previously endorsed** ***(****NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)*

### RELIABILITY: SPECIFICATIONS

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

   **Submission document:** Items sp.01-sp.30

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

   **Reviewer 1:** None

   **Reviewer 2:** No concerns.

   **Reviewer 3:** My main concern is with the fact that non-independence among patient-months is not accounted for. Additionally, equation 1 (page 29) for the risk model is not consistent with the specifications. For example, no transplant center fixed effects are included based on the equation. Neither are the interaction terms noted on page 41.

   **Reviewer 4:** No concerns

   **Reviewer 5:** The meaning of the distinction between being waitlisted and being waitlisted "in active status" is not clearly articulated. Presumably one must be "in active status" to receive a transplant, but why is this distinction so important and why are both measures necessary? Given that both measures meet reliability and validity criteria, would 3694 suffice? Also, it isn't entirely clear how "the percentage of prevalent patients waitlisted" is estimated at the practice level, given that the unit of observation is a patient-month, and each patient contributes up to 12 months during the measurement year.

**Reviewer 7:** No concerns. sp.06: adjust the max age for adults to 75

**Reviewer 8:** The following specifications are unclear / unstated: [1] In sp. 14 (which defines the denominator), it states "…assigned to a dialysis practitioner group practice according to each patient's treatment history during a given month during the reporting year." The method (if any) as to the selection of that "given month" is unstated. This is important for a variety of reasons, e.g. whether the opportunity for gaming is present here. [2] In 2b.20 (which defines the risk factors), it states "A set of prevalent comorbidities based on either Medicare inpatient or outpatient claims…" It's unstated whether these comorbidities are limited to claims prior to the measurement period. This is important so as to limit the risk factors to that were present at the start of care.

**Reviewer 9:** Obviously, a given patient may contribute to many subsequent months in the numerator which will help with the measure score, but not necessarily the individual patient.

**Reviewer 11:** NOTE: This measure does not address any patient condition that could be improved with a clinical intervention. I would classify the type of measure as "process: appropriate use" as the measure encourages practitioner groups to quickly place patients needing kidney or pancreas transplants on a waitlist.

**Reviewer 12:** See comments re: cancer and scleroderma (2yrs)

## RELIABILITY: TESTING

**Type of measure:**

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

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

**Data Source:**

**☒ Claims ☐ eCQM (HQMF) implemented in EHRs ☐ Abstracted from Electronic Health Records**
**☐ Abstracted from Paper Medical Records ☐ Instrument-Based Data ☒ Registry**
**☒ Enrollment Data ☐ Other (please specify)**
**Reviewer 5:** CROWNWeb, Nursing Home MDS, CMS Medical Evidence form
**Reviewer 12:** They claim this is an outcome measure. Is it?

**Level of Analysis:**

**☒ Group/Practice ☒ Individual Clinician ☐ Hospital/facility/agency ☐ Health Plan**
**☐ Population: Regional, State, Community, County or City ☐ Accountable Care Organization**
**☐ Integrated Delivery System ☐ Other (please specify)**

**Submission document:** Questions 2a.01-09

3. **Reliability testing level**

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

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

   **☒ Yes ☐ No**

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

   **☒ Yes ☐ No**

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

   **Submission document:** Question 2a.10

**Reviewer 1:** Use of IUR statistics for reliability is acceptable.

**Reviewer 2:** Appropriate methods used; ANOVA; calculated IUR.

**Reviewer 3:** The IUR method used is acceptable.

**Reviewer 4:** Method appears to be appropriate for testing measure scores between individual practices using inter-unit reliability to measure variance to indicate true differences

**Reviewer 5:** IUR was estimated using a bootstrapping method that appears to be appropriate.

**Reviewer 7:** No concerns

**Reviewer 8:** Signal to noise testing was conducted. This test is appropriate for reliability for this type of measure.

**Reviewer 9:** Over 2.5 million patients across over 2,00 practitioner groups with at least 11 eligible patients. Inter-unit reliability calculated for between-group variation for annual performance score. Bootstrap approach utilized to account for the fact that any given patient may have many different practitioner groups over the time frame.

**Reviewer 11:** The developer used any ANOVA approach comparing between and within variance across provider groups. They report an IUR (inter-unit reliability) for the between/total ratio.

**Reviewer 12:** SNR

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

   **Submission document:** Question 2a.11

   **Reviewer 1:** Reliability is relatively high using the IUR method.

   **Reviewer 2:** IUR was 0.93, indicating a high degree of reliability.

   **Reviewer 3:** The IUR is 0.93, which is quite high. However, it is not clear how this is affected by lack of accounting for non-independence among patient-months.

   **Reviewer 4:** IUR= 0.93

   **Reviewer 5:** IUR=0.93, which is high.

   **Reviewer 7:** No concerns

   **Reviewer 8:** Per the signal to noise test used, the results show an overall mean of 0.93. The finding reflects a high level of reliability.

   **Reviewer 9:** IUR was 0.93 which means that 93% of the variation observed is attributable to between-provider variation and 7% attributable to within-provider variation.

   **Reviewer 11:** The IUR value was 0.93 for groups with >10 patients and >1 expected events. This represents a high level of reliability.

   **Reviewer 12:** IUR 0.93

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

   **Submission document:** Question 2a.10-12

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable**

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

   **Submission document:** Question 2a.10-12

   ☒ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    **Reviewer 1:** The IUR value is high, but I am concerned with this measure, as with others in this set, that the inclusion of the same patient in consecutive months as if they were independent observations may be exaggerating the reliability of the measure.

    **Reviewer 2:** Appropriate methods used; IUR value was high.

    **Reviewer 3:** The developers should address the non-independence issue.

    **Reviewer 4:** No concerns

    **Reviewer 5:** No concerns

    **Reviewer 7:** Results are indicative of high reliability at the accountable entity level.

    **Reviewer 8:** Response to question #2: The following specifications are unclear / unstated: [1] In sp. 14 (which defines the denominator), it states "…assigned to a dialysis practitioner group practice according to each patient's treatment history during a given month during the reporting year." The method (if any) as to the selection of that "given month" is unstated. This is important for a variety of reasons, e.g. whether the opportunity for gaming is present here. [2] In 2b.20 (which defines the risk factors), it states "A set of prevalent comorbidities based on either Medicare inpatient or outpatient claims…" It's unstated whether these comorbidities are limited to claims prior to the measurement period. This is important so as to limit the risk factors to that were present at the start of care.

    **Reviewer 9:** IUR of 0.93 is indicative of high reliability.

    **Reviewer 11:** The methodologies describe to assess reliability were appropriate. The results for the item reliability were moderate. My concern is that this measure does not fit the definition of a patient outcome, but is better described as a practice process measure.

    **Reviewer 12:** Range.

## VALIDITY: TESTING

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

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

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

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

    ☒ **Yes**

    ☐ **No**

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

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

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

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

☐ **Face validity**

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

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

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

    **Submission document:** Question 2b.02

    ☒ **Yes**

    ☐ **No**

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

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

    **Submission document**: Question 2b.02

    **Reviewer 1:** Correlation with subsequent transplant and mortality rates is a reasonable way to establish validity on this particular measure.

    **Reviewer 2:** Compared performance on the measure to mortality rates and transplant rates

    **Reviewer 3:** The developers hypothesized that this proposed measure is conceptually related to two quality measures. One is mortality measure and another is overall transplant rates.

    **Reviewer 4:** Empirical validity testing was well described and appropriate

    **Reviewer 5:** Construct validity was assessed by group practice-level associations with two key outcome measures: mortality rates and overall transplant rates. The latter correlation seems self-evident, because listing is an essential prelude to transplantation, but it is still important to test for validation of the measure.

    **Reviewer 7:** It is recommended that other forms of validity testing are conducted to strengthen the measure's validity, e.g., face validity and data element validity.

    **Reviewer 8:** The tests to identify an association with two measures (i.e. transplant rates and mortality rates) appears to be appropriate given the information we have. However, we were not presented with information as to the methodology for calculating transplant rates and mortality rates.

    **Reviewer 9:** Validity was tested using the assumed association between the measure score and mortality and overall transplant rates for the attributable practitioner. Higher scores should be associated with higher transplant rates with a more modest association with mortality, due to many other impactful considerations. Tertiles were formed and tested against the above associations, with cutoffs of 9.6%, 14.1%, and 70.4%.

    **Reviewer 11:** Tertile comparison of measure scores is rather limited in its ability to demonstrate validity (i.e., stability vs movement among levels).

    **Reviewer 12:** The tertile groups based on the performance scores were defined as: T1 (best performance): 14.1% - 70.4% T2: 9.6% - 14.1% T3 (worst performance): 0% - 9.6%. The dialysis practitioner group level average mortality rates are 17.8, 18.3 and 19.2 deaths per 100 patient-years for T1, T2 and T3 respectively (trend test p=0.002). The Spearman correlation coefficient is -0.083 (p<0.0001). Average transplant rates are 5.0, 4.2 and 3.1 transplants per 100 patient-years for T1, T2 and T3 respectively (trend test p=0.002). The Spearman correlation coefficient is 0.279 (p<0.0001).

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

    **Submission document:** Questions 2b.03-04

    **Reviewer 1:** Validity is acceptable - both correlations were in the predicted direction.

**Reviewer 2:** Relationships with transplant rates and mortality were in the expected directions and statistically significant.

**Reviewer 3:** The developers showed that there is a weak negative association with the mortality measure as expected. The results based on the association with overall transplant rates provide empirical support of the validity of this measure.

**Reviewer 4:** Validity testing very thorough and support that differences in performance are valid

**Reviewer 5:** Both associations are in the expected direction, and clinically and statistically significant.

**Reviewer 7:** Results indicate low to moderate validity, driving the moderate rating for validity.

**Reviewer 8:** The test result regarding mortality rate was very weak: -0.08. The result regarding the transplant rate was modest: 0.28.

**Reviewer 9:** As assumed, higher performance scores did correlate with higher transplant rates and had a more modest correlation with mortality rate. 59 (2.6%) had an score better than expected, 2,104 (92.4%) had a score as expected, and 113 (5.0%) had a score that was worse than expected. The differences were interpreted and clinically meaningful.

**Reviewer 11:** Measure scores at highest level were highly skewed (T1=14.1%-70.4%; T2=9.6%-14.1%; T3=0%-9.6%). Statistical difference in mortality, but small range (T1=17.8; T2=18.3; T3=19.2); transplant rate showed similar relationship with measure score (T1=5.0; T2=4.2; T3=3.1). Practitioner group practice performance reported as “Better than—As—Worse than” expected. Median reported values for these groups are 19.9%, 11.9%, and 3.4%, respectively, with an overall average of 11.7%

**Reviewer 12:** Above

## VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

**Submission document:** Questions 2b.15-18.

**Reviewer 1:** None

**Reviewer 2:** None. All exclusions are clinically relevant. Exclusions did not change overall group performance scores much.

**Reviewer 3:** No concern

**Reviewer 4:** Exclusions well supported

**Reviewer 5:** Exclusions for age 75 or older, admission to SNF, hospice care, and dementia are all appropriate.

**Reviewer 7:** No concerns

**Reviewer 8:** No concerns.

**Reviewer 9:** Missing data, mostly the CMS-2728 was present in 3,125, or 1.11% of over 280,000 patients. Excluded patients were tabulated according the criteria in the measure specs and constituted 28.6% of the total population before exclusion. Scores were impacted moderately when the excluded population was included in the analysis.  as well as contributing to within-practitioner variation.

**Reviewer 11:** Measure exclusions identified but do make a slight difference in measure score before and after the exclusions are applied. That is, measure values before exclusions are lower than after exclusions are applied.

19. **Risk Adjustment**

**Submission Document:** Questions 2b.19-32

19a. **Risk-adjustment method**

☐ None ☒ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

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

☐ Yes ☒ No ☒ Not applicable

19c. **Social risk adjustment:**

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

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

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

19d.**Risk adjustment summary:**

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

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

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

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

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

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

**Reviewer 1:** Reasonable approach to model development using both clinical and social variables.

**Reviewer 2:** Appropriate methods. C-statistic of 0.763.

**Reviewer 3:** Four areas of concern regarding the risk adjustment approach: 1. Model equation on page 29 is not consistent with model specifications in the text. For example, on page 28, both transplant center fixed characteristics and random effect are listed. On page 41, it says "two-way interactions were examined and selected for the final model based on both the magnitude and statistical significance of the estimates." 2. Non-independence among patient-months is ignored, should it be ignored? 3. Shouldn't the model be validated using an external dataset? 4. Based on page 30, medical/clinical risk factors were selected based on their associations with one-year mortality. This seems odds given that one-year mortality is not the outcome.

**Reviewer 4:** Very thorough risk-adjustment strategy with rationale for which factors to include

**Reviewer 5:** This is a process measure. Accordingly, the selection of risk factors must be extremely well justified to avoid magnifying bias by adjusting for factors that are in the quality pathway. It may be appropriate to adjust for functional factors that obviously interfere with transplant eligibility, such as inability to transfer, inability to ambulate. Adjustment for major medical comorbidities that may be difficult to control, such a heart failure and COPD, is also appropriate. However, adjustment for social risk factors (such as ADI and dual eligibility) when severe disparities on these same factors are so well documented is shocking and unconscionable. Please refer to KDIGO 2020 guidelines as well as the ASPE reports on social risk factors. KDIGO does not recommend de-prioritizing patients for transplant based on area deprivation or dual eligibility. Only medically legitimate reasons for deferring or declining waitlisting belong in the risk adjustment model. Finally, adjustment for failure to submit a CMS-2728 form is inappropriate.

**Reviewer 7:** No concerns

**Reviewer 8:** The risk adjustment method appears reasonable from what information is shared. However, see response to #2. Specifically this excerpt: “In 2b.20 (which defines the risk factors), it states “A set of prevalent comorbidities based on either Medicare inpatient or outpatient claims…” It’s unstated whether these comorbidities are limited to claims prior to the measurement period. This is important so as to limit the risk factors to that were present at the start of care.”

**Reviewer 9:** Risk adjustment includes the factors relevant to the population.

**Reviewer 11:** Developer identifies the measure score as an outcome and makes a valient effort to describe how this "outcome" could be risk adjusted using a statistical process (see pg. 30 in their document). I think the measure score is a process measure and should not be risk adjusted.

**Reviewer 12:** c 0.76; graph

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

**Submission document:** Questions 2b.05-07

**Reviewer 1:** The measure seems to only identify high and low outliers, and should not be used to distinguish groups in the middle of the distribution.

**Reviewer 2:** None. 92.4% were "as expected"

**Reviewer 3:** No concern

**Reviewer 5:** No concerns.

**Reviewer 7:** No concerns

**Reviewer 8:** In response to 2b.06, 92.4% of facilities perform "as expected" per statistical testing. In turn, only 7.6% of facilities are "high" or "low" outliers. This low rate of identifying outlier faciltes means the measure is of low value in aiding consumers in their decision making based on quality (as this measure defines quality).

**Reviewer 9:** None

**Reviewer 11:** The measure score is compared with an expected score. The result is reporting provider group as "Better than—As—Worst than" Expected. The measure is not a patient outcome. Hence, the comparison among provider groups is based on dubious predictive standards. The developer/owner could simply establish "industry standard" strata and report provider group performance based on these strata.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

**Submission document:** Questions 2b.11-14.

**Reviewer 1:** N/A

**Reviewer 2:** Not applicable

**Reviewer 3:** No concern

**Reviewer 5:** Not applicable.

**Reviewer 7:** Not applicable

**Reviewer 8:** No concerns.

**Reviewer 9:** None

22. **Please describe any concerns you have regarding missing data.**

**Submission document:** Questions 2b.08-10.

**Reviewer 1:** None

**Reviewer 2:** None. Low rates of missing data.

**Reviewer 3:** No concern

**Reviewer 4:** No concerns

**Reviewer 5:** Missingness of the CMS-2728 medical evidence form is uncommon (1.1%), but it is inappropriate to adjust for missingness in this manner (because it incentivizes lower submission rates). A far better approach is to assume that these patients have no 2728 comorbidities, which would incentivize submission in the future.

**Reviewer 7:** No concerns

**Reviewer 8:** No concerns as the extent of missing data is small: 1.1%.

**Reviewer 9:** See above

**Reviewer 11:** Minimal missing data.

### For cost/resource use measures ONLY:

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

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

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

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

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

    ☒ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level **is required**; if not conducted, should rate as INSUFFICIENT.)

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

**Reviewer 1:** Empirical correlations were significant and in the predicted direction.

**Reviewer 2:** Appropriate methods. Hypothesized relationships were in the desired direction and statistically significant.

**Reviewer 3:** The developers should address the identified issues associated with the risk adjustment.

**Reviewer 4:** No concerns

**Reviewer 5:** Rating would be moderate except for the unconscionable addition of social factors to the risk-adjustment model, when serious disparities with respect to these factors have been repeatedly documented in the literature, leading to substantial harm to patients with ESRD.

**Reviewer 7:** See above

**Reviewer 8:** Response to question #17: The test result regarding mortality rate was very weak: -0.08. The result regarding the transplant rate was modest: 0.28. Response to question #20: In response to 2b.06, 92.4% of facilities perform "as expected" per statistical testing. In turn, only 7.6% of facilities are "high" or "low" outliers. This low rate of identifying outlier facilities means the measure is of low value in aiding consumers in their decision making based on quality (as this measure defines quality).

**Reviewer 9:** For the associations tested, which are deemed relevant to the population, the validity was high for subsequent transplantation and more modest for mortality, which was expected.

**Reviewer 11:** The measure is not a patient outcome and use of risk adjustment is not advised for process measures. Measure has some potential as a process measure, although method for reporting performance should be reconsidered.

**Reviewer 12:** Recc exclude cancer(s), connective tissue disease scleroderma rather than adjust. Nephrologist input?

### For composite measures ONLY

**Submission documents:** Questions 2c.01-08

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

## ADDITIONAL RECOMMENDATIONS

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

    **Reviewer 5:** This measure is a classic example of when NOT to adjust for social risk factors. It is process measure for which social factors are in the quality pathway.

## 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:

**2021 Submission:**

Updated evidence information here.

**2018 Submission:**

Evidence from the previous submission here.

**1a.01. Provide a logic model.**

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

**[Response Begins]**

This measure tracks the outcome of placement and maintenance on the kidney or kidney-pancreas transplantation waitlist in active status (meaning the patient is ready to be transplanted immediately if a suitable organ becomes available), with the intended objective of improving the overall health of patients on dialysis. Being waitlisted is an outcome as it represents a desirable change in health status for patients on dialysis, indicating achievement of a health condition conducive to kidney transplantation. This outcome results 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 and obesity, and as needed, optimizing mental health and social support systems. In addition, dialysis practitioners support the path for patients towards waitlisting or living donor transplantation through proper education about the transplantation option, referral to a transplant center and assistance with completion of the transplant evaluation process. The logic model for the steps involved is diagrammed below (with the outcome measure in bold):

Patients with ESRD are initiated on dialysis -> Patients not already on the wait list are 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 practitioner assists patient with completion of the transplant evaluation process and in 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 waitlist -> **Dialysis practitioner helps patient maintain active status on the**

**wait list through involvement in ongoing evaluation activities and by optimizing health and functional status, with possibility to receive a deceased donor kidney 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 include both patients with advanced chronic kidney disease prior to transplant, and recipients of transplants, and 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 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 kidney transplant, 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].

This measure focuses specifically on the prevalent dialysis population, examining waitlisting monthly in active status for each patient. As this measure assesses monthly waitlisting in active status of patients, it also evaluates and encourages maintenance of patients on the waitlist which is important given the long duration most patients have to wait to eventually access a deceased donor transplant (national median of roughly 4 years) [6]. In particular, maintenance of active status requires ongoing attention by dialysis practitioners to optimizing the health of patients, to ensure sustained suitability for transplant waitlisting. Maintenance of active status on the waitlist is additionally important given demonstrated disparities [7] and positive association with subsequent transplantation [8]. This is an important area to which dialysis practitioners can contribute through ensuring patients remain healthy, and complete any ongoing testing activities required to remain active on the wait list. In contrast to this measure, the First Year Standardized Waitlist Ratio focuses solely on new waitlistings and living donor kidney transplants to incentivize early action, rather than ongoing maintenance on the waitlist, as this measure does.

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. United States Renal Data System. 2020 USRDS Annual Data Report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2020.

7. Kulkarni S, Ladin K, Haakinson D, et al. Association of Racial Disparities With Access to Kidney Transplant After the Implementation of the New Kidney Allocation System. JAMA Surg 2019; 154(7):618-625.

8. Grams, M. E., Massie, A. B., Schold, J. D., Chen, B. P., & Segev, D. L. (2013). Trends in the inactive kidney transplant waitlist and implications for candidate survival. American Journal of Transplantation, 13(4), 1012-1018.

Abstract: In November 2003, OPTN policy was amended to allow kidney transplant candidates to accrue waiting time while registered as status 7, or inactive. We evaluated trends in inactive listings and the association of inactive status with transplantation and survival, studying 262,824 adult first-time KT candidates listed between 2000 and 2011. The proportion of waitlist candidates initially listed as inactive increased from 2.3% prepolicy change to 31.4% in 2011. Candidates initially listed as inactive were older, more often female, African American, and with higher body mass index. Postpolicy change, conversion from initially inactive to active status generally occurred early if at all: at 1 year after listing, 52.7% of initially inactive candidates had been activated; at 3 years, only 66.3% had been activated. Inactive status was associated with a substantially higher waitlist mortality (aHR 2.21, 95%CI:2.15-2.28, p<0.001) and lower rates of eventual transplantation (aRR 0.68, 95%CI:0.67-0.70, p<0.001). In summary, waitlist practice has changed significantly since November 2003, with a sharp increase in the number of inactive candidates. Using the full waitlist to estimate organ shortage or as a comparison group in transplant outcome studies is less appropriate in the current era.

**[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 aPPPW performance scores for all dialysis practitioner group practices that had at least 11 patients in 2019. The mean value of aPPPW was 12.3%. The interquartile range (Q3-Q1) is 7.3%, with the bottom quartile of practitioner group practices having 7.3% or less of prevalent patients waitlisted vs. the top quartile of practitioner group having 15.6% or more of their prevalent patients waitlisted.

N=dialysis practitioner groups=2276; N of patients=280,855; N of patient-months=2,541,229.

Table 1: Descriptive statistics of aPPPW (%), overall and by decile, 2019

| * | **Mean** | **Std Dev** | **Minimum** | **Maximum** | **Median** | **Lower Quartile** | **Upper Quartile** |
|---|---|---|---|---|---|---|---|
| Overall | * | * | * | * | * | * | * |
| * | 12.3 | 6.2 | 0.0 | 70.4 | 11.7 | 8.3 | 15.6 |
| Decile | * | * | * | * | * | * | * |
| 1 | 2.8 | 1.8 | 0.0 | 5.4 | 3.2 | 1.0 | 4.4 |
| 2 | 6.5 | 0.6 | 5.4 | 7.5 | 6.5 | 6.0 | 7.0 |
| 3 | 8.3 | 0.4 | 7.5 | 9.1 | 8.3 | 7.9 | 8.6 |
| 4 | 9.8 | 0.4 | 9.1 | 10.4 | 9.8 | 9.5 | 10.1 |
| 5 | 11.0 | 0.4 | 10.4 | 11.7 | 11.0 | 10.7 | 11.3 |
| 6 | 12.4 | 0.4 | 11.7 | 13.1 | 12.4 | 12.0 | 12.8 |
| 7 | 13.8 | 0.5 | 13.1 | 14.7 | 13.7 | 13.4 | 14.2 |
| 8 | 15.6 | 0.5 | 14.7 | 16.5 | 15.6 | 15.1 | 16.1 |
| 9 | 17.9 | 0.9 | 16.5 | 19.6 | 17.8 | 17.2 | 18.7 |
| 10 | 24.7 | 5.8 | 19.6 | 70.4 | 22.7 | 20.7 | 26.9 |

Table 1: Descriptive statistics of aPPPW (%), overall and by decile, 2019

*Cell intentionally left blank.

**[Response Ends]**

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

**[Response Begins]**

N/A

**[Response Ends]**

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

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

**[Response Begins]**

N= dialysis practitioner groups=2276; N of patients=280,855; N of patient-months=2,541,229.

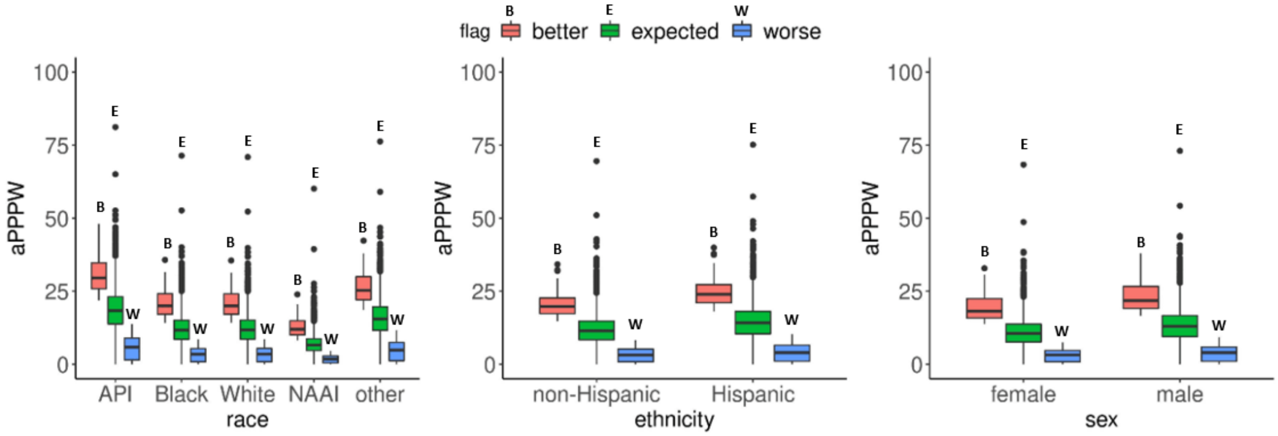
Table 2: Descriptive statistics of aPPPW (%), by race, ethnicity and sex, 2019

| * | Mean | Std Dev | Minimum | Maximum | Median | Lower Quartile | Upper Quartile |
|---|---|---|---|---|---|---|---|
| **Race** | * | * | * | * | * | * | * |
| Asian/Pacific Islander | 18.4 | 8.4 | 0.0 | 81.2 | 18.0 | 13.1 | 23.3 |
| Black | 12.0 | 6.1 | 0.0 | 71.4 | 11.4 | 8.1 | 15.2 |
| White | 12.0 | 6.1 | 0.0 | 70.9 | 11.5 | 8.1 | 15.3 |
| Native American/Alaskan Indian | 6.9 | 4.0 | 0.0 | 60.1 | 6.4 | 4.4 | 8.8 |
| “Other” race | 15.7 | 7.4 | 0.0 | 76.3 | 15.2 | 11.0 | 19.9 |
| **Ethnicity** | * | * | * | * | * | * | * |
| Non-Hispanic | 11.8 | 6.0 | 0.0 | 69.5 | 11.3 | 8.0 | 14.9 |
| Hispanic | 14.5 | 7.1 | 0.0 | 75.1 | 13.9 | 10.0 | 18.3 |
| **Sex** | * | * | * | * | * | * | * |
| Female | 10.9 | 5.7 | 0.0 | 68.3 | 10.4 | 7.3 | 13.8 |
| Male | 13.3 | 6.6 | 0.0 | 73.0 | 12.7 | 9.0 | 16.8 |

Table 2: Descriptive statistics of aPPPW (%), by race, ethnicity and sex, 2019

*Cell intentionally left blank.

Figure 1: Performance of aPPPW (%), by race, ethnicity and sex, 2019



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

**[Response Ends]**

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

**[Response Begins]**

N/A

**[Response Ends]**

## Criteria 2: Scientific Acceptability of Measure Properties

**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]**

Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW)

**[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]**

This measure tracks the percentage of patients in each dialysis practitioner group practice who were on the kidney or kidney-pancreas transplant waitlist in active status. Results are averaged across patients prevalent on the last day of each month during the reporting year.

The proposed measure is a directly standardized percentage, which is adjusted for covariates (e.g. age and risk factors).

**[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]**

N/A

**[Response Ends]**

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

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

**[Response Begins]**

Available in attached Excel or csv file

**[Response Ends]**

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

**sp.12. State the numerator.**

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

*DO NOT include the rationale for the measure.*

**[Response Begins]**

The numerator is the adjusted count of patient months in which the patient at the dialysis practitioner group practice is on the kidney or kidney-pancreas transplant waitlist in an active status as of the last day of each month during the reporting 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.13. Provide details needed to calculate the numerator.**

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

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

**[Response Begins]**

The adjusted count of patient months in which the patient at the dialysis practitioner group practice is on the kidney or kidney-pancreas transplant waitlist in an active status, adjusted for patient-mix. To be included in the numerator for a particular month, the patient must be on the kidney or kidney-pancreas transplant waitlist in active status as of the last day of the month during the reporting year.

**[Response Ends]**

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

**sp.14. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

All patient-months for patients who are under the age of 75 in the reporting month and who are assigned to a dialysis practitioner group practice according to each patient's treatment history during a given month during the reporting year.

**[Response Ends]**

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

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

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

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

**[Response Begins]**

During the target reporting months 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 month, 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.

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.16. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

- Patients who were at age 75 or older in the reporting month
- Patients who were admitted to a skilled nursing facility (SNF) during the month of evaluation were excluded from that month;
- Patients who were admitted to a skilled nursing facility (SNF) within one year of dialysis initiation according to form CMS-2728
- Patients determined to be in hospice were excluded from month of evaluation and remainder of reporting period
- 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 are not excluded from the measure. All patients who meet the denominator inclusion criteria are included and used to model a given dialysis practitioner group's expected waitlist rate. If a dialysis practitioner group has fewer than 11 patients, then the dialysis practitioner group is excluded from reporting outcomes.

**[Response Ends]**

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

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

**[Response Begins]**

The Nursing Home Minimum Dataset and the Questions 17u and 22 on CMS Medical Evidence Form were used to identify patients in skilled nursing facilities. For hospice patients, a separate CMS file that contains final action claims submitted by hospice providers was used to determine the hospice status. Nursing home status from the CMS-2728 form is only used for incident patients, i.e. patients in which the start of ESRD is within one year of the month of evaluation. Once a patient is determined to be on hospice, the patient is excluded from the measure in the month of evaluation and the remainder of the reporting period.

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.

**[Response Ends]**

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

*Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.*

**[Response Begins]**

N/A

**[Response Ends]**

**sp.19. Select the risk adjustment type.**

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

**[Response Begins]**

Statistical risk model

**[Response Ends]**

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

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Rate/proportion

**[Response Ends]**

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

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

**[Response Begins]**

Better quality = Higher score

**[Response Ends]**

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

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

**[Response Begins]**

See attached flowchart.

**[Response Ends]**

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

**[Response Begins]**

N/A

**[Response Ends]**

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

**[Response Begins]**

Claims

Registry Data

**[Response Ends]**

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

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

**[Response Begins]**

EQRS (formerly CROWNWeb), Medicare Claims, and the CMS Medical Evidence Form 2728 were used as the data sources for establishing the denominator. CROWNWeb 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 in active status). 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 and Questions 17u and 22 on the CMS Medical Evidence Form were 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 group practice. Area Deprivation Index (ADI) was obtained from Census data (2011-2015) based on patient zip code. In order to assess the transplant center characteristics, Scientific Registry of Transplant Recipients (SRTR) data was used.

**[Response Ends]**

**sp.30. Provide the data collection instrument.**

**[Response Begins]**

No data collection instrument provided

**[Response Ends]**

## 2a. Reliability

**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]**

2019 data derived from a combination of 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]**

01-01-2019 – 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]**

Using 2019 data, there were 2,276 dialysis practitioner groups included in these analyses, after restricting to practitioner group practices that had at least 11 eligible patients.

**[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 are 2,541,229 patient-months (280,855 patients) in total. Among all patients-months in 2019, the average age was 57.4 years old, 41.6% of patient-months were female, 54.8% were White, 37.9% were Black, 5.2% were Asian/Pacific Islander, 1.6% were American Indian/Alaskan Native, 0.4% were Other/Multi-racial/Unknown/missing and 18.0% were Hispanic.

At the patient-level, the mean age was 57.2 years old and 41.5% were female. Of these 56.2% were White, 36.5% were Black, 5.2% were Asian/Pacific Islander, 1.6% were American Indian/Alaskan Native, and 0.4% were other/Multi-racial/Unknown/missing and 17.6% 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 – Area Deprivation Index from 2015 Census data.

**[Response Ends]**

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

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

*Choose one or both levels.*

**[Response Begins]**

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

**[Response Ends]**

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

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

**[Response Begins]**

We used 2019 data to calculate dialysis practitioner group practice annual performance scores. Our approach for determining measure reliability aligns with one-way analysis of variance (ANOVA), in which the between dialysis practitioner group practice variation ($\sigma_b^2$) and the within- dialysis practitioner group practice variation ($\sigma_{t,w}^2$) in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of the measure (i.e., $\sigma_b^2 + \sigma_{t,w}^2$) that is attributed to the between – dialysis practitioner group practice variation, the true signal reflects the differences across dialysis practitioner group practices. We assessed reliability by calculating inter-unit reliability (IUR) for the annual performance scores. If the measure were an average of the individuals’ measurements under the care of one dialysis practitioner group practice, the usual ANOVA approach would be used. The yearly based measure, however, is not a simple average and we instead estimate the IUR using a bootstrap approach, which uses a resampling scheme to estimate the within dialysis practitioner group practice variation that cannot be directly estimated by ANOVA. A small IUR (near 0) reveals that most of the variation of the measures between dialysis practitioner group practices is driven by random noise, indicating the measure would not be a good characterization of the differences among dialysis practitioner group practices. A large IUR (near 1) indicates that most of the variation between dialysis practitioner groups practices is due to true differences between dialysis practitioner group practices.

Below is our approach to calculate IUR.

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

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

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

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

where ni is the number of subjects in the $i^{th}$ practitioner group, $\overline{T} = {}^{\sum n_i T_i}\big/_{\sum n_i}$, and

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

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

The reliability of aPPPW calculation only included dialysis practitioner group practices with at least 11 patients 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 0.93. Dialysis practitioner group practices with <11 eligible patients were excluded from this calculation.

**[Response Ends]**

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

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

**[Response Begins]**

The value of IUR indicates that about 93% of the variation in the aPPPW measure can be attributed to the between-dialysis practitioner group practice differences (signal) and about 7% of variation to within-dialysis practitioner group practice variation (noise). The value of IUR implies a high degree of reliability.

**[Response Ends]**

## 2b. Validity

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

**[Response Begins]**

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

Empirical validity testing

**[Response Ends]**

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

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

**[Response Begins]**

Validity of the measure was tested by evaluating the association between the dialysis practitioner group level measure performance, and mortality and overall transplant rates among all patients attributed to the practitioner groups. We hypothesized that practitioner groups with higher performance on the aPPPW measure would have 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 aPPPW measure would demonstrate lower 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 aPPPW (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 also computed the corresponding mortality rate and transplant rate for each practitioner group in the same year. We then tested the trend between the tertile grouping and these practitioner group-level outcomes. Finally, we examined the Spearman correlations between the practitioner group measure value and each of the outcomes respectively.

**[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:

T1 (best performance): 14.1% - 70.4%

T2: 9.6% - 14.1%

T3 (worst performance): 0% - 9.6%

The dialysis practitioner group level average mortality rates are 17.8, 18.3 and 19.2 deaths per 100 patient-years for T1, T2 and T3 respectively (trend test p=0.002). The Spearman correlation coefficient is -0.083 (p<0.0001). Average transplant rates are 5.0, 4.2 and 3.1 transplants per 100 patient-years for T1, T2 and T3 respectively (trend test p=0.002). The Spearman correlation coefficient is 0.279 (p<0.0001).

**[Response Ends]**

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

**[Response Begins]**

As expected, higher aPPPW performance correlated with higher transplant rate, with clear separation of transplant rates across dialysis practitioner group tertiles of performance. The direction of the relationship with mortality was also as expected, and statistically significant, with numerically lower mortality with higher performance on the measure although the magnitude of the association was smaller than for transplant rate.

**[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]**

To test the null hypothesis that the aPPPW for a given dialysis practitioner group is statistically different from the national average, we use a simulation method to calculate the nominal p-value as the probability that the observed number of events (a binary outcome of 0 indicates that the patient is not on the waitlist in active status in during that month and a binary outcome of 1 indicates that the patient is on the waitlist in active status during that month) should be at least as extreme as that expected. This calculation is based on the supposition that, having adjusted for case mix, this practitioner group has a true event rate corresponding to the average practitioner groups. We then converted the p-values to z-scores. 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. The 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.

**[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 aPPPW, stratified by classification category

| Classification category | N (%) | Median aPPPW |
|---|---|---|
| **Better than Expected** | 59 (2.6) | 19.9 |
| **As Expected** | 2104 (92.4) | 11.9 |
| **Worse than Expected** | 113 (5.0) | 3.4 |
| **Total** | 2276 (100) | 11.7 |

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

**[Response Ends]**

**2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.**

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

**[Response Begins]**

As shown in Table 2, most dialysis practitioner group practices (92.4%) had an aPPPW that was “As Expected”. Approximately 2.6% of dialysis practitioner group practices has a aPPPW that was “Better than Expected”, while approximately 5.0% were “Worse than Expected”. Across these categories, performance on waitlisting in active status varied widely (from 3.4% of patients waitlisted in the worse than expected category, to nearly 20% in the better than expected category), suggesting that differences are also clinically meaningful.

**[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 reporting month. 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 (i.e. at the time of dialysis initiation).

**[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 280,855 patients

| Data element | Missing (%) |
|---|---|
| Patients with missing CMS-2728 | 3,125 (1.11) |

**Table 4:** Distribution of missing data among 280,855 patients

**[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 9 in section 2b.24, patients with missing CMS-2728 form have a lower odds of waitlisting compared to those without a missing CMS-2728 form (OR = 0.510 ; 95% CI = 0.490, 0.531).

**[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 practitioner groups with fewer than 11 attributed 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 patient months excluded

| * | Before age, nursing home, hospice, and dementia exclusion | After age, nursing home, hospice, and dementia exclusion | Percentage excluded |
|---|---|---|---|
| Number of patient-months | 3,561,019 | 2,541,229 | 28.6% |

*This cell is intentionally left blank.

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

| **Variable excluded** | **Frequency (%)** |
|---|---|
| Age >= 75 | 766,648 (21.5) |
| Nursing home from CMS-2728 | 26,618 (0.8) |
| Nursing home from Nursing home history file | 302,227 (8.5) |
| Hospice | 14,581 (0.4) |
| Dementia | 152,951 (4.3) |

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

**Table 7:** Distribution of performance scores (aPPPW) before and after exclusions

| **aPPPW (%)** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** | **Q1** | **Median** | **Q3** |
|---|---|---|---|---|---|---|---|
| Before exclusion | 9.0 | 4.6 | 0.0 | 53.2 | 6.1 | 8.5 | 11.3 |
| After exclusion | 12.3 | 6.2 | 0.0 | 70.4 | 8.3 | 11.7 | 15.6 |

**Table 7:** Distribution of performance scores (aPPPW) before and after exclusions

**Figure 2:** Distribution of aPPPW (%) before exclusions

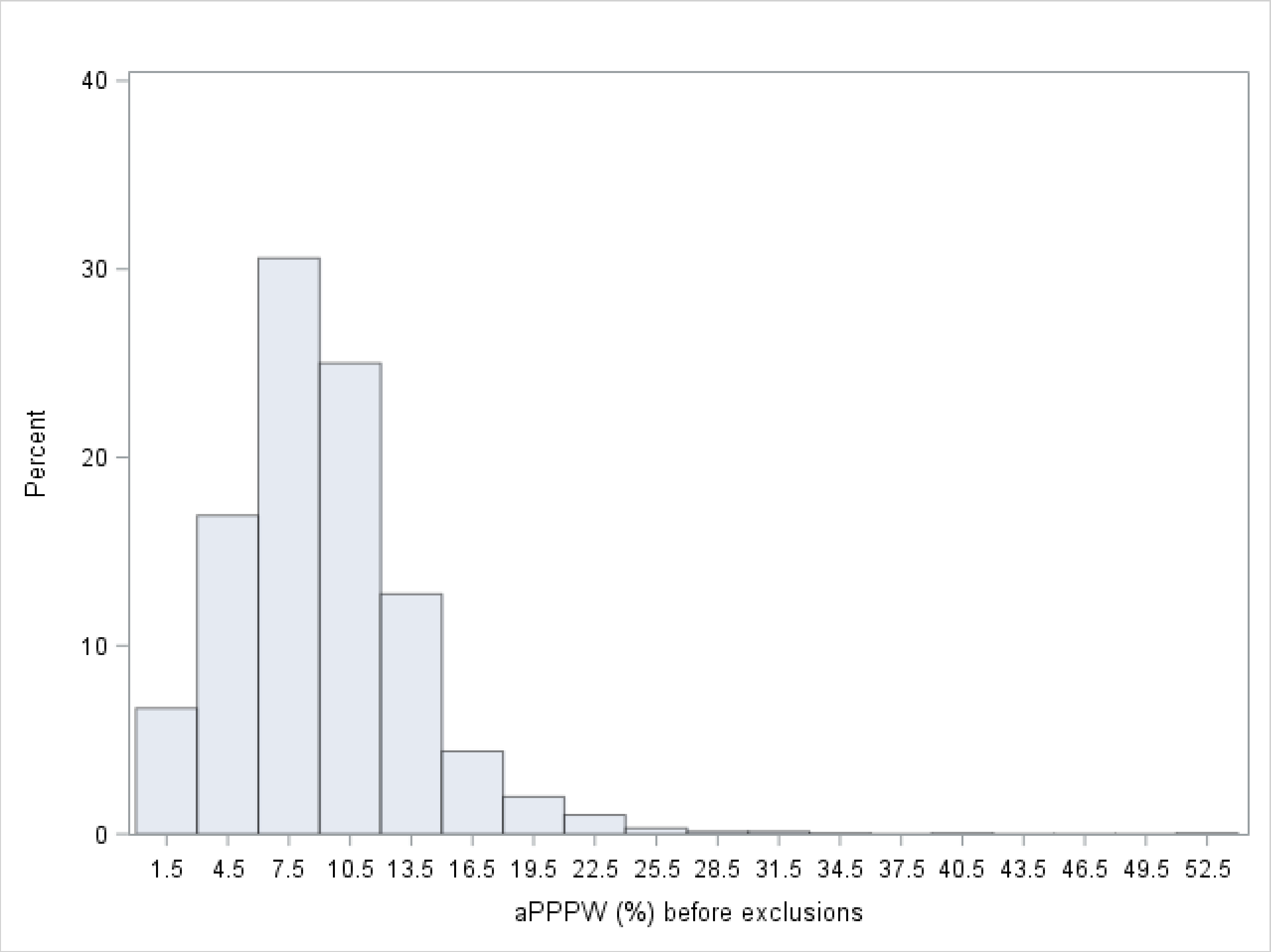


**Figure 3:** Distribution of aPPPW (%) after exclusions

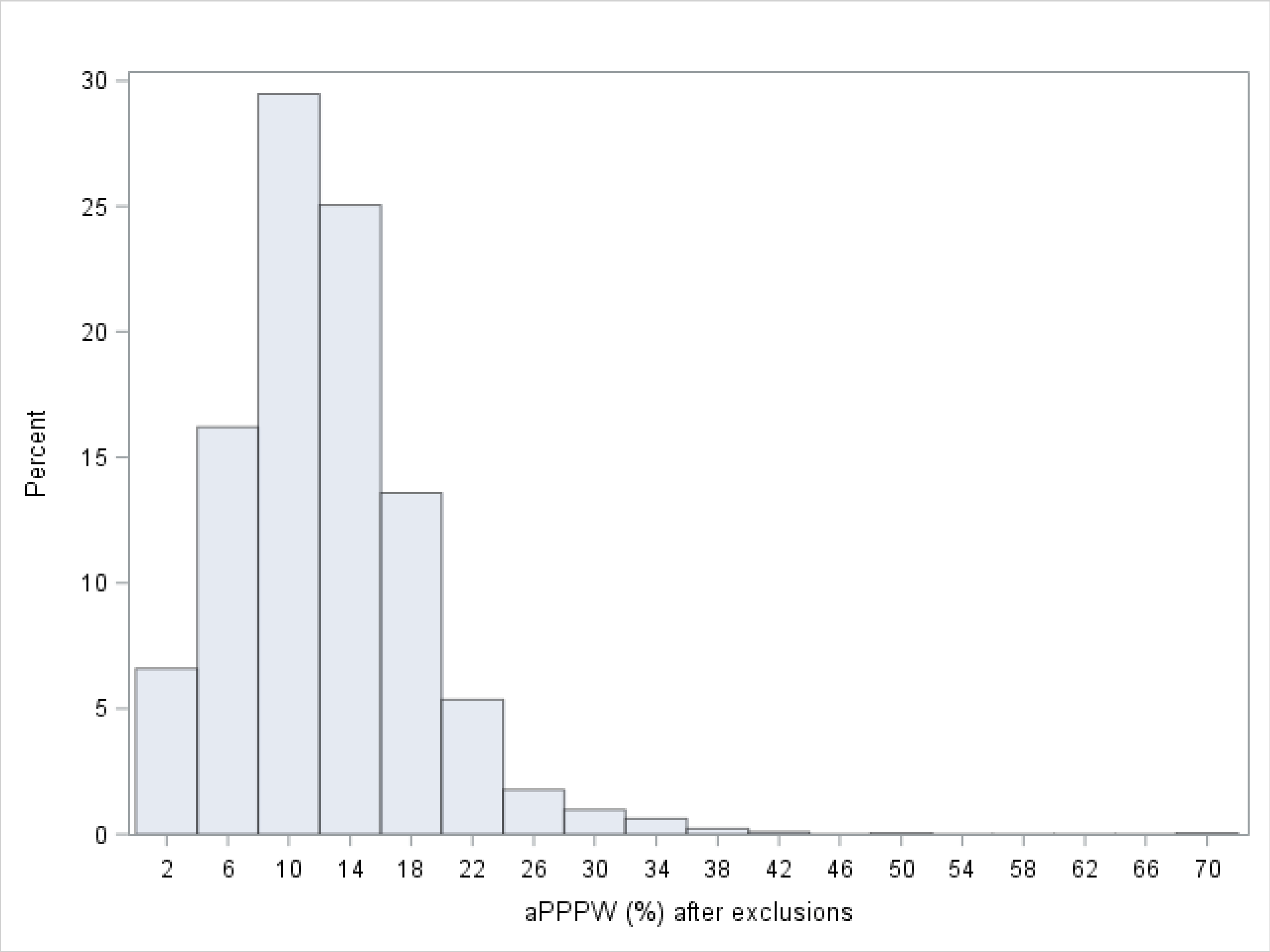
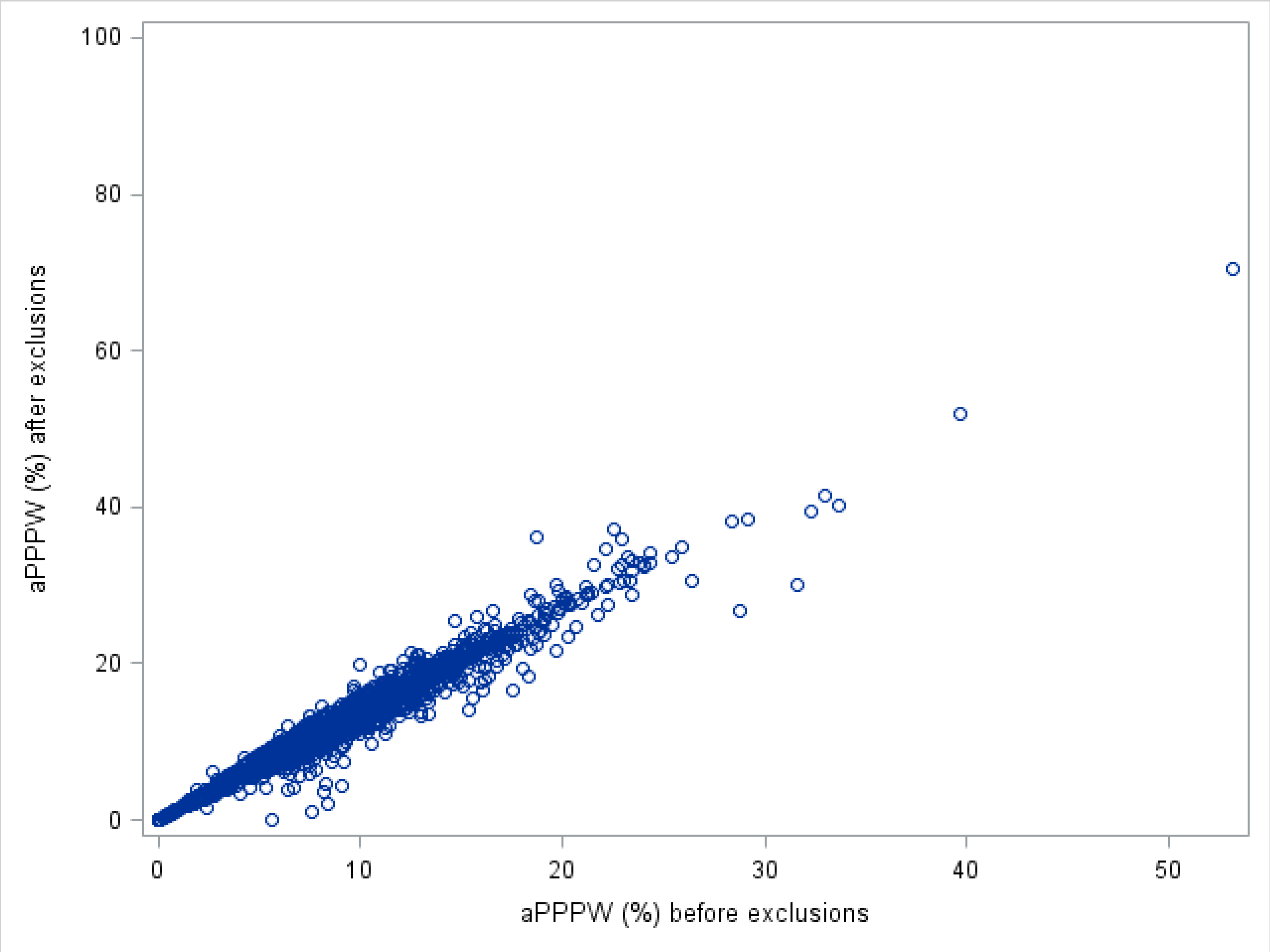


**Figure 4:** Scatterplot of aPPPW (%) with and without exclusions



The correlation coefficient is 0.977 (p<0.001).

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

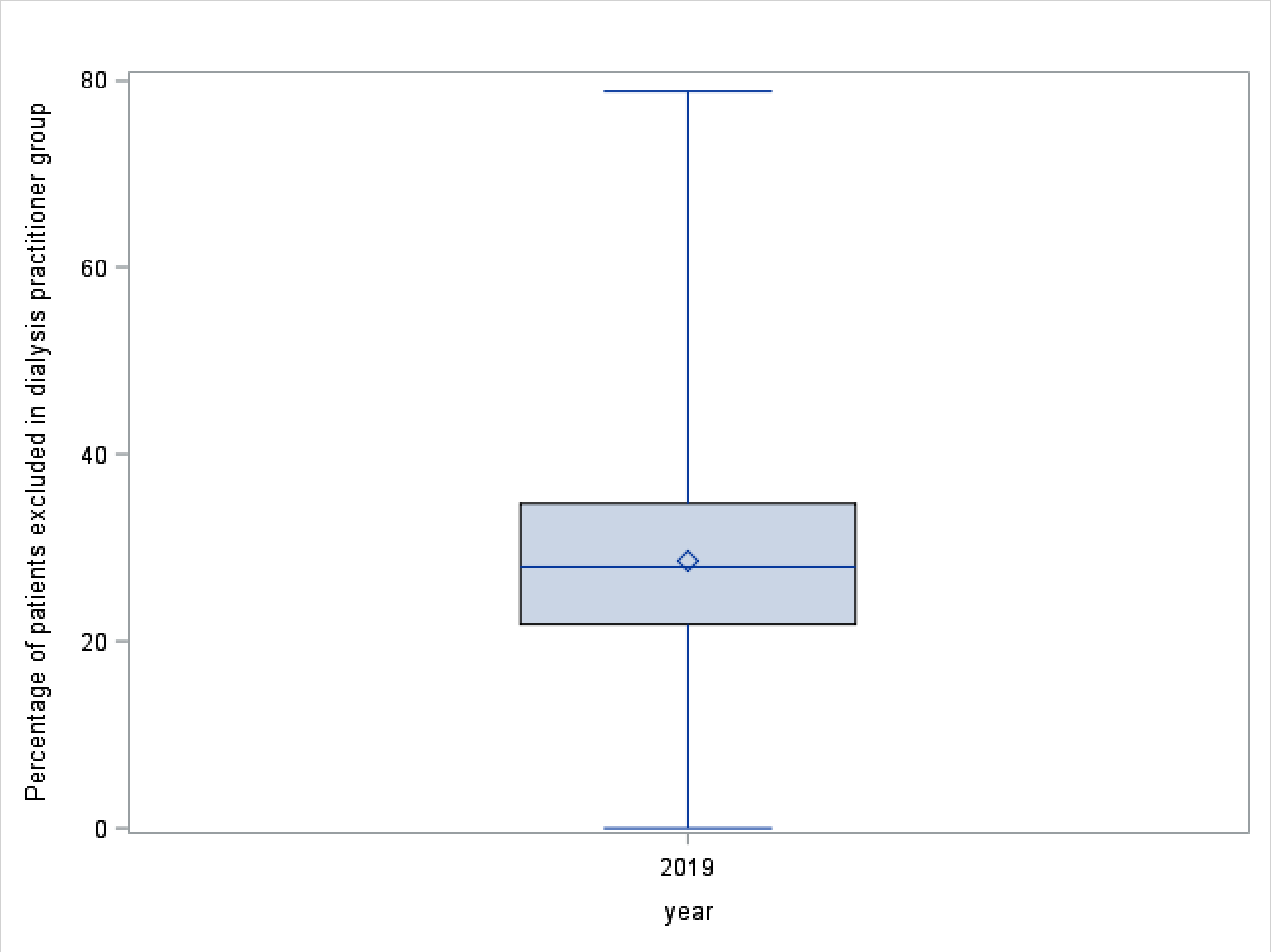
| * | * | aPPPW without patient-level exclusion | aPPPW without patient-level exclusion | aPPPW without patient-level exclusion | aPPPW without patient-level exclusion |
|---|---|---|---|---|---|
| * | * | **Better than Expected** | **As Expected** | **Worse than Expected** | **Total** |
| **PPPW with patient-level exclusion** | **Better than Expected** | 54 | 5 | 0 | 59 (2.6) |
| **PPPW with patient-level exclusion** | **As Expected** | 5 | 2069 | 30 | 2104 (92.4) |

| * | * | aPPPW without patient-level exclusion | aPPPW without patient-level exclusion | aPPPW without patient-level exclusion | aPPPW without patient-level exclusion |
|---|---|---|---|---|---|
| **PPPW with patient-level exclusion** | **Worse than Expected** | 0 | 12 | 101 | 113 (5.0) |
| **PPPW with patient-level exclusion** | **Total** | 59 (2.6) | 2086 (91.7) | 131 (5.8) | 2276 |

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

*This cell is intentionally left blank.

**Figure 5:** Distribution of Excluded patients at dialysis practitioner group practice



**[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 overall measure scores are changed moderately by the exclusions (see Table 7, figure 2-3), practitioner group performance rankings are minimally affected (Table 8). Nevertheless, 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 dialysis practitioner groups, as shown in Figure 5. Finally, as the data to determine the exclusions is readily available, there is minimal additional burden for analysis anticipated by using these exclusion criteria.

**[Response Ends]**

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

**[Response Begins]**

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

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

See 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]**

Covariates in the model are listed below:

- Age
    - Age is included as continuous variable as well as age spline knots at 15, 55, and 70
- ADI
- Dual eligibility
    - Dual Eligible
    - Not Dual Eligible
- Diabetes, primary cause of ESRD
- Comorbidities at ESRD incidence:
    - Congestive heart failure
    - Atherosclerotic heart disease and other cardiac disease
    - Cerebrovascular disease, CVA, TIA
    - Peripheral vascular disease
    - Diabetes other than as primary cause of ESRD (all types including diabetic retinopathy)
    - Chronic obstructive pulmonary disease
    - Inability to ambulate
    - Inability to transfer
    - Malignant neoplasm, cancer

  - Tobacco use (current smoker)
  - Drug dependence
  - No Medical Evidence (CMS-2728) Form
  - At least one of the comorbidities listed
- A set of prevalent comorbidities based on either Medicare inpatient or outpatient claims (individual comorbidities categorized into 64 categories – see below)
- Transplant center fixed characteristics and random effect

To estimate the probability that a prevalent patient is waitlisted in active status, we use a mixed-effects logistic regression model, in which dialysis practitioner groups are modeled as fixed effects and transplant centers are modeled as random effects. The expected number of prevalent patients waitlisted in active status for the dialysis practitioner group under evaluation is estimated as the sum of the probabilities of prevalent patients waitlisted across all dialysis practitioner groups and assuming their effects are the same as the dialysis practitioner group under evaluation.

Consider patient *k* at dialysis practitioner group practice *i* and transplant center *j* during calendar month *l*; we set the response variate to $Y_{ijkl}=1$ if the patient is on the wait list in active status and $Y_{ijkl}=0$ if not. The model and methods are described in some additional detail below:

- To estimate the probability that a prevalent patient is waitlisted, we use a mixed-effects logistic regression model:

  $$\log\left(\frac{p_{ijkl}}{1-p_{ijkl}}\right) = \gamma_i + \alpha_j + \beta^T Z_{ijkl}, \text{(1)}$$

  where $p_{ijkl}$ represents the probability that patient *k* at dialysis practitioner group practice *i* and transplant center *j* during calendar month *l* is waitlisted in active status, and $Z_{ijkl}$ represents the set of patient-level characteristics, including age (coded as a linear spline with empirically determined knots at ages 15, 55 and 70), incident comorbidities, prevalent comorbidities, ADI, and dual eligibility and i and the dialysis practitioner group practice indicators. In this mixed-effect model, $\gamma_i$ is the fixed effect for dialysis practitioner groups and $a_j$ is the random effect for transplant center *j* . It is assumed that the $a_j$s arise as independent normal variables (i.e., $a_j \sim N(0, \sigma^2)$).
- We then compute a$PPPW_m$ for each dialysis practitioner group practice *m* as follows: $PPPW_m = \sum_i\sum_j\sum_k\sum_l exp(\gamma_m + \alpha_j + \beta^T Z_{ijkl}) / \{1 + exp(\gamma_m + \alpha_j + \beta^T Z_{ijkl})\} / n$, where *n* = total number of patient-months included in the overall study sample.

**[Response Ends]**

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

Published literature

Internal data analysis

**[Response Ends]**

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

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

**[Response Begins]**

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

Social Risk Factors:

Under conceptual considerations, and as supported by the TEP, it was deemed important to adjust for social risk on the basis that it could affect suitability for transplant waitlisting. This could occur, for example, through difficulty with ability to pay for transplant immunosuppression medications, or lacking the resources to travel to a transplant center for care, which are considerations taken into account for suitability for transplant waitlisting. For this purpose, dual Medicare-Medicaid eligibility (at the patient level, representing socioeconomic disadvantage) and Area Deprivation Index (ADI) were investigated and included in our model. Dual eligibility was obtained from Medicare claims and could also be obtained from the CMS-2728 form for incident patients within the first year of ESRD. ADI was obtained based on patient zip code of residence and used as a proxy to adjust for potential differences in waitlisting for neighborhoods of different ranking of socioeconomic disadvantage (see Patzer et al reference below).

Functional Risk Factors:

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

Medical/clinical 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 practitioner group practices with a substantially older age composition than the average. A linear spline was used to model the effect of (continuous) age. The spline's knots were determined empirically using standard techniques.

Additionally, incident (at time of dialysis initiation) and prevalent comorbidities were included to account for adverse health conditions that could affect suitability for transplant waitlisting. Incident comorbidities identified on Form 2728 were selected for adjustment in the 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. First, 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). Moreover, to additionally account for transplant center effects, we also include adjustment for transplant centers using random effects. Our general aim is to adjust each practitioner group's measure for the potential effects of the transplant centers that are corresponding to its patients. With this approach, each practitioner group's measure is adjusted for our best estimate of the true effect of each transplant center, taking account of the distribution from which these effects arise. This has the advantage of circumventing problems with identifiability that would arise if transplant centers were included as fixed effects and also tends appropriately to dampen the effects of transplant centers with extreme outcomes. For transplant center adjustments in the model, patients were assigned to a transplant center based on historical waitlisting patterns in their zip code of residence.

References:

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

Abstract:

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

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

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

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

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

Abstract:

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

Design, setting, participants, & measurements: We examined 3029 consecutive adult patients referred fortransplantation 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 9:** Model statistics for risk factors in aPPPW model

| **Covariate** | **odds ratio** | **95% CI** |
|---|---:|---:|
| **Age** | * | * |
| Continuous (years) | 1.033 | 1.021, 1.046 |
| Spline at 15 | 0.950 | 0.939, 0.962 |
| Spline at 55 | 0.976 | 0.975, 0.978 |
| Spline at 70 | 0.796 | 0.789, 0.803 |
| Area Deprivation Index (ADI), per 10% increase on percentile scale | 0.916 | 0.991, 0.992 |
| Dual eligibility | 0.618 | 0.613, 0.624 |
| Diabetes, primary cause of ESRD | 0.643 | 0.632, 0.654 |
| **Comorbidities at incidence** | * | * |

| Covariate | odds ratio | 95% CI |
|---|---:|---:|
| Heart disease | 0.902 | 0.885, 0.920 |
| Other cardiac disease | 0.902 | 0.887, 0.916 |
| Congestive heart failure | 0.624 | 0.615, 0.632 |
| Chronic obstruction pulmonary disease | 0.608 | 0.588, 0.628 |
| Inability to ambulate | 0.362 | 0.342, 0.384 |
| Inability to transfer | 0.730 | 0.667, 0.798 |
| Cancer | 0.732 | 0.711, 0.753 |
| Peripheral vascular disease | 0.794 | 0.776, 0.812 |
| Cerebrovascular disease | 0.748 | 0.731, 0.765 |
| Tobacco use | 0.494 | 0.482, 0.505 |
| Drug use | 0.420 | 0.397, 0.444 |
| Diabetes, non-primary | 0.764 | 0.749, 0.779 |
| At least one incident comorbidity listed | 0.983 | 0.967, 1.000 |
| No Medical Evidence (CMS-2728 Form) | 0.510 | 0.490, 0.531 |
| At least 6 months of Medicare Coverage in prior year | 0.873 | 0.863, 0.884 |
| **Prevalent comorbidities** | * | * |
| Candidal esophagitis | 0.656 | 0.587, 0.733 |
| Sarcoidosis | 1.235 | 1.161, 1.314 |
| Cancer of Liver | 0.555 | 0.486, 0.635 |
| Cancer of Lung | 0.232 | 0.199, 0.270 |
| Cancer of Bladder | 0.569 | 0.501, 0.647 |
| Cancer of Bone | 0.175 | 0.139, 0.222 |
| Other Neoplasm | 0.867 | 0.805, 0.934 |
| Non-Hodgkins Lymphoma | 0.525 | 0.462, 0.596 |
| Multiple Myeloma | 0.299 | 0.274, 0.327 |
| Myelodysplastic Syndrome | 0.653 | 0.585, 0.727 |
| Diabetes without complications | 1.146 | 1.131, 1.161 |
| Diabetes with complications | 1.133 | 1.116, 1.149 |
| Glucocorticoid deficiency | 0.999 | 0.939, 1.064 |
| Malnutrition / Cachexia | 0.933 | 0.916, 0.951 |
| Disorders of urea cycle metabolism | 1.057 | 0.934, 1.195 |
| Other amyloidosis | 1.069 | 0.961, 1.188 |
| Other specified disorders of metabolism | 0.792 | 0.756, 0.831 |

| Covariate | odds ratio | 95% CI |
|---|---|---|
| Sickle-cell Anemia | 0.851 | 0.774, 0.936 |
| Pancytopenia | 0.837 | 0.806, 0.869 |
| Neutropenia | 0.920 | 0.856, 0.988 |
| Substance Related Disorders | 0.525 | 0.472, 0.583 |
| Opioid Dependance | 0.659 | 0.628, 0.692 |
| Schizophrenia | 0.303 | 0.272, 0.338 |
| Peripheral autonomic neuropathy in disorders classified elsewhere | 0.903 | 0.824, 0.991 |
| Epilepsy | 0.728 | 0.710, 0.746 |
| Bipolar Disorder | 0.745 | 0.710, 0.781 |
| Major depressive affective disorder | 0.711 | 0.699, 0.723 |
| Alcohol Related Disorders | 0.889 | 0.835, 0.947 |
| Coma | 0.899 | 0.825, 0.981 |
| Cerebral edema | 1.158 | 1.029, 1.303 |
| Myocardial Infarction | 0.651 | 0.633, 0.670 |
| Coronary Atherosclerosis | 0.874 | 0.857, 0.892 |
| Pulmonary embolism and infarction | 0.881 | 0.839, 0.926 |
| Primary pulmonary hypertension | 0.816 | 0.762, 0.873 |
| Pulmonary Heart Disease | 0.793 | 0.776, 0.810 |
| Cardiomyopathy | 0.780 | 0.766, 0.796 |
| Atrioventricular block, complete | 0.712 | 0.662, 0.765 |
| Paroxysmal Tachycardia | 0.836 | 0.802, 0.871 |
| Atrial fibrillation | 0.816 | 0.802, 0.831 |
| Atrial flutter | 0.893 | 0.860, 0.928 |
| Acute Cerebrovascular Disease | 0.788 | 0.766, 0.811 |
| Peripheral and Visceral Atherosclerosis | 0.876 | 0.863, 0.890 |
| Venous Thromboembolism | 0.769 | 0.744, 0.795 |
| Esophageal varices | 2.239 | 2.042, 2.454 |
| Chronic Obstructive Pulmonary Disease | 0.607 | 0.594, 0.619 |
| Aspiration Pneumonitis | 0.976 | 0.921, 1.035 |
| Other Lower Respiratory Diseases | 1.062 | 0.995, 1.134 |
| Respiratory Failure | 0.648 | 0.635, 0.661 |
| Cirrhosis of Liver | 0.820 | 0.792, 0.848 |

| Covariate | odds ratio | 95% CI |
|---|---|---|
| Other Liver Disease | 1.195 | 1.134, 1.259 |
| Pancreatitis | 0.724 | 0.683, 0.768 |
| Chronic Skin Ulcer | 0.604 | 0.592, 0.617 |
| Systemic lupus erythematosus and connective tissue disorders | 1.184 | 1.153, 1.215 |
| Rheumatoid Arthritis | 1.010 | 0.969, 1.053 |
| Pathologic Fracture | 0.798 | 0.722, 0.883 |
| Gangrene | 0.657 | 0.630, 0.685 |
| HIV | 0.522 | 0.497, 0.548 |
| Gastrostomy status | 0.920 | 0.842, 1.005 |
| Other artificial opening of urinary tract status | 0.536 | 0.468, 0.615 |
| Dependence on respirator, status | 1.024 | 0.944, 1.110 |
| Below knee amputation status | 0.544 | 0.523, 0.565 |
| Above knee amputation status | 0.483 | 0.444, 0.526 |
| Long-term (current) use of insulin | 1.056 | 1.042, 1.071 |
| Inflammatory polyarthropathy | 0.950 | 0.849, 1.062 |
| Weighted transplant center waitlist mortality ratio | 1.180 | 1.079, 1.291 |
| Weighted transplant center rate ratio | 0.705 | 0.676, 0.735 |

**Table 9:** Model statistics for risk factors in aPPPW model

*This cell is intentionally left blank.

**[Response Ends]**

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

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

**[Response Begins]**

As noted in section 2b.23, we included Medicare-Medicaid dual eligibility and ADI as social risk factors in the model on a clinical and conceptual basis, and as supported by an expert panel. Both factors were significantly associated with the outcome of waitlisting (see Table 9 in 2b.24).

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

**Table 10:** Odds Ratio and 95% Confidence Interval in the aPPPW model including race

| **Race** | **Odds Ratio** | **95%CI** |
|---|---|---|
| Asian/Pacific Islander | 1.190 | 1.119, 1.265 |
| Black | 0.885 | 0.876, 0.894 |
| White | Reference | Reference |
| Native American/Alaskan Indian | 0.602 | 0.558, 0.651 |
| “Other” race | 1.133 | 1.067, 1.202 |

**Table 10:** Odds Ratio and 95% Confidence Interval in the aPPPW model including race

**Table 11:** Odds Ratio and 95% Confidence Interval in the aPPPW model including ethnicity

| **Ethnicity** | **Odds Ratio** | **95%CI** |
|---|---|---|
| Non-Hispanic | 1.162 | 1.148, 1.176 |
| Hispanic | Reference | Reference |

**Table 11:** Odds Ratio and 95% Confidence Interval in the aPPPW model including ethnicity

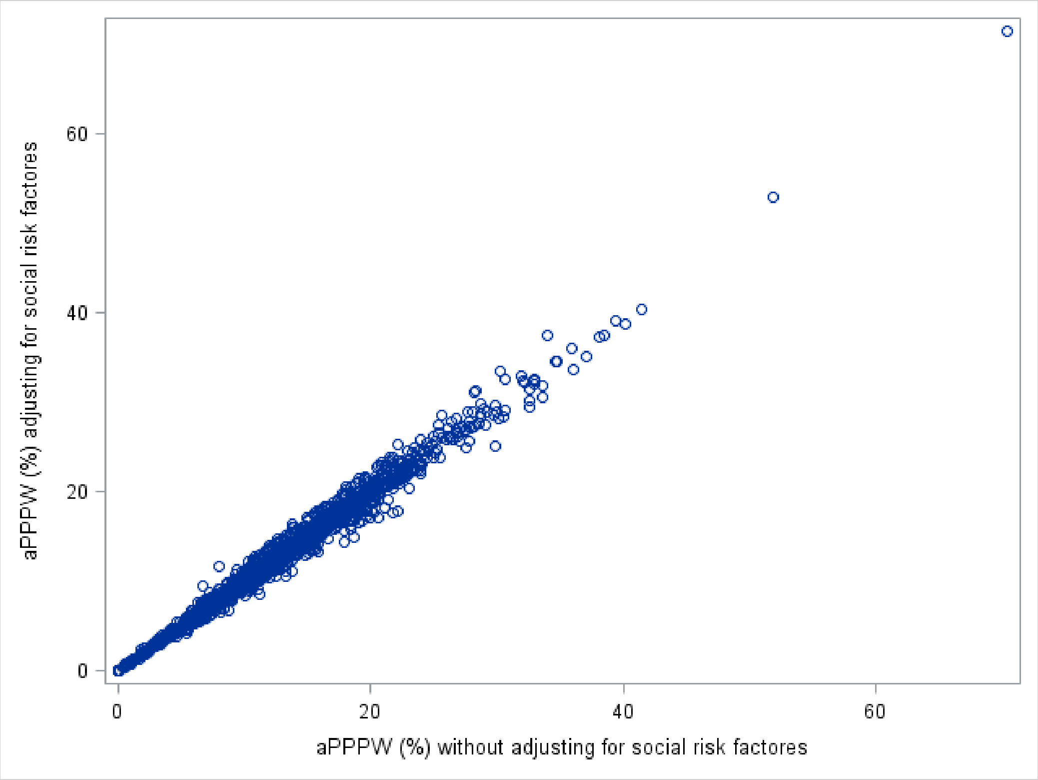
**Table 12:** Odds Ratio and 95% Confidence Interval in the aPPPW model including sex

| **Sex** | **Odds Ratio** | **95%CI** |
|---|---|---|
| Female | 0.836 | 0.828, 0.843 |
| Male | Reference | Reference |

**Table 12:** Odds Ratio and 95% Confidence Interval in the aPPPW model including sex

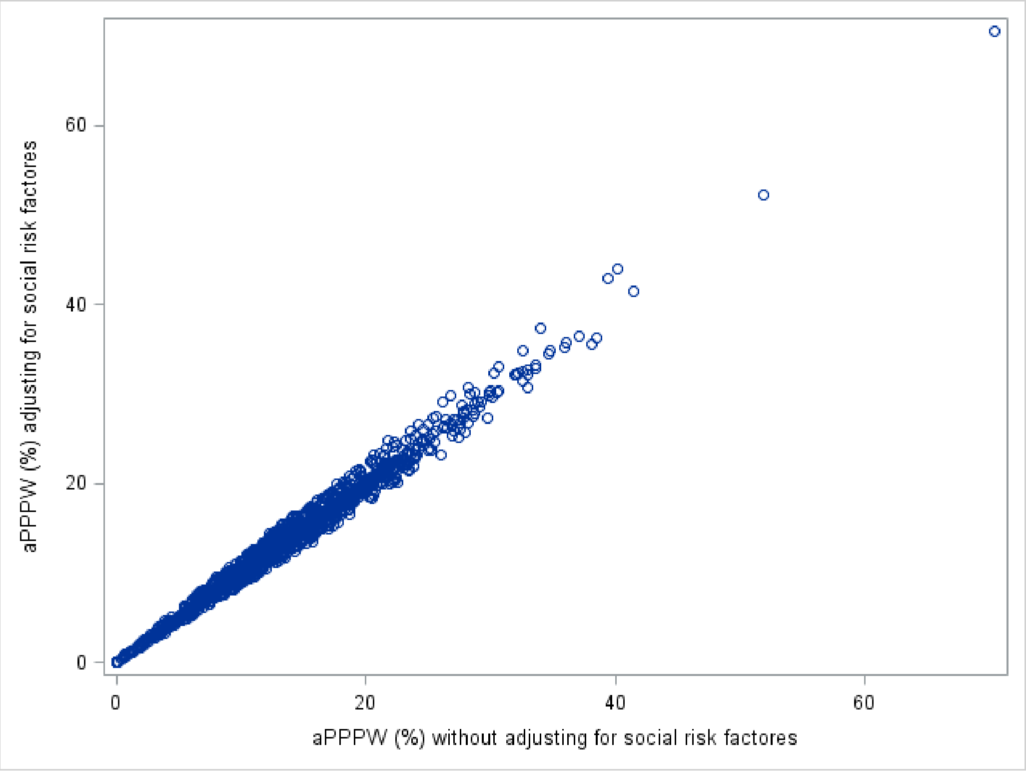
**Figure 6:** Correlation between aPPPW with and without risk factors

**Race**



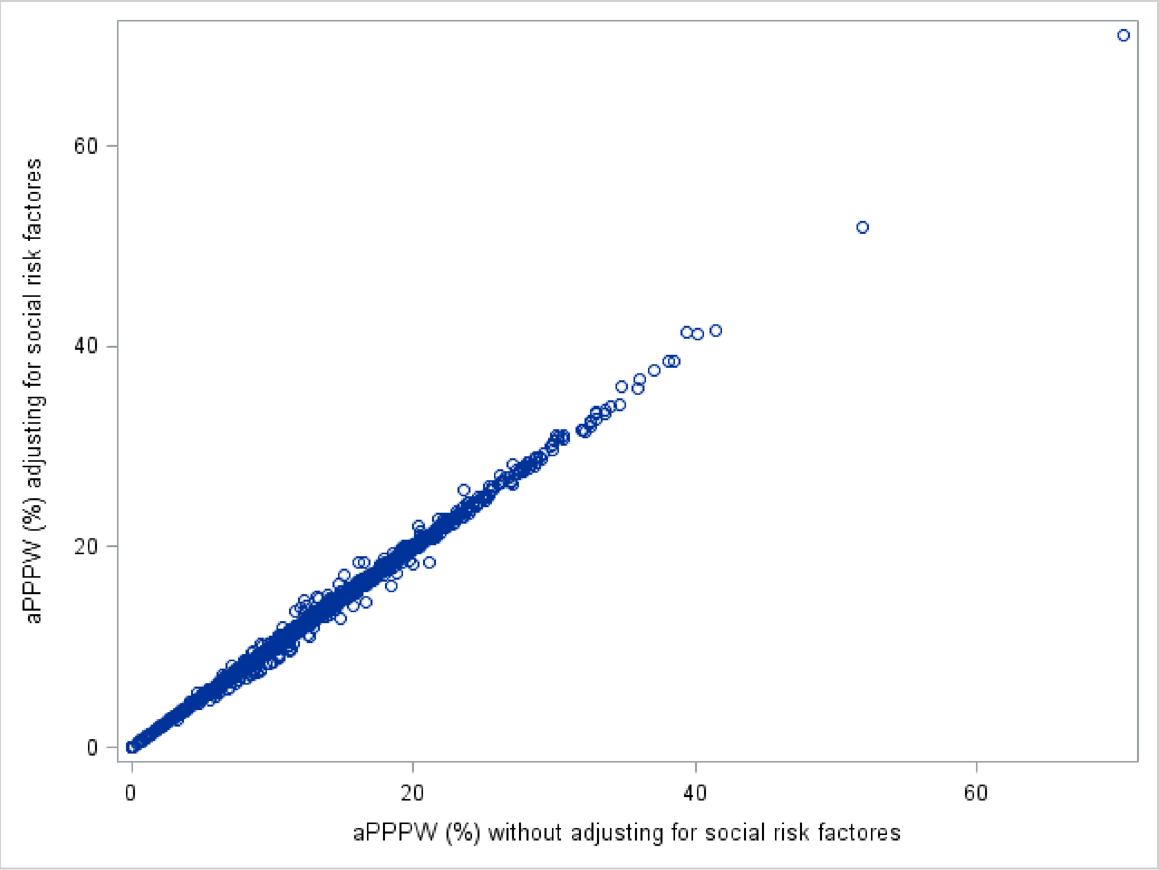
Correlation coefficient=0.992, p<0.0001

**Ethnicity**



Correlation coefficient=0.993, p<0.0001

**Sex**



Correlation coefficient=0.998, p<0.0001

**Table 13:** Comparison of performances with and without adjusting for risk factors

**Race**

| * | * | **aPPPW without race** | **aPPPW without race** | **aPPPW without race** | **aPPPW without race** |
|---|---|---|---|---|---|
| * | * | **Better than Expected** | **As Expected** | **Worse than Expected** | **Total** |
| **aPPPW with ethnicity** | **Better than Expected** | 53 | 4 | 0 | 57 (2.5) |
| **aPPPW with ethnicity** | **As Expected** | 6 | 2091 | 2 | 2099 (92.2) |
| **aPPPW with ethnicity** | **Worse than Expected** | 0 | 9 | 111 | 120 (5.3) |
| **aPPPW with ethnicity** | **Total** | 59 (2.6) | 2104 (92.4) | 113 (5.0) | 2276 |

**Table 13:** Comparison of performances with and without adjusting for risk factors: **Race**

*This cell is intentionally left blank.

**Ethnicity**

| * | * | **aPPPW without ethnicity** | **aPPPW without ethnicity** | **aPPPW without ethnicity** | **aPPPW without ethnicity** |
|---|---|---|---|---|---|
| * | * | **Better than Expected** | **As Expected** | **Worse than Expected** | **Total** |
| **aPPPW with ethnicity** | **Better than Expected** | 58 | 1 | 0 | 59 (2.6) |
| **aPPPW with ethnicity** | **As Expected** | 1 | 2100 | 3 | 2104 (92.4) |
| **aPPPW with ethnicity** | **Worse than Expected** | 0 | 3 | 110 | 113 (5.0) |
| **aPPPW with ethnicity** | **Total** | 59 (2.6) | 2104 (92.4) | 113 (5.0) | 2276 |

**Table 13:** Comparison of performances with and without adjusting for risk factors: **Ethnicity**

*This cell is intentionally left blank.

**Sex**

| * | * | **PPPW without sex** | **PPPW without sex** | **PPPW without sex** | **PPPW without sex** |
|---|---|---|---|---|---|
| * | * | **Better than Expected** | **As Expected** | **Worse than Expected** | **Total** |
| **aPPPW with sex** | **Better than Expected** | 56 | 0 | 0 | 56 (2.5) |
| **aPPPW with sex** | **As Expected** | 3 | 2100 | 6 | 2109 (92.7) |
| **aPPPW with sex** | **Worse than Expected** | 0 | 4 | 107 | 111 (4.9) |
| **aPPPW with sex** | **Total** | 59 (2.6) | 2104 (92.4) | 113 (5.0) | 2276 |

**Table 13:** Comparison of performances with and without adjusting for risk factors: **Sex**

*This cell is intentionally left blank.

Although there are differences in waitlisting by sex, ethnicity and race, it is unclear whether these associations are due to underlying biological or other patient factors, or represent disparities in care. Adjusting for these factors could have the unintended consequence of creating or reinforcing disparities. Furthermore, Tables 13 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. Two-way interactions were examined and selected for the final model based on both the magnitude and statistical significance of the estimates.

**[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 0.763, meaning that the model correctly ordered 76.3% of the pairs of patient-months that were discordant with respect to the response variate. Month-specific C-statistics were computed in order to identify any trends by month in the model's discriminatory ability.

**[Response Ends]**

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

**[Response Begins]**

The Hosmer-Lemshow (H-L) statistic is defined strictly for independent trials, and months within-patient are expected to be highly correlated. We therefore chose to perform the H-L statistic in a month-specific fashion, with the p values being low (p=0.0065 for January). However, in very large samples such as this even relatively small departures from the model will lead to significant results. While the p-value is significant, based on the decile plot in Figure 6 below, the observed and expected values by decile appear to be stable.

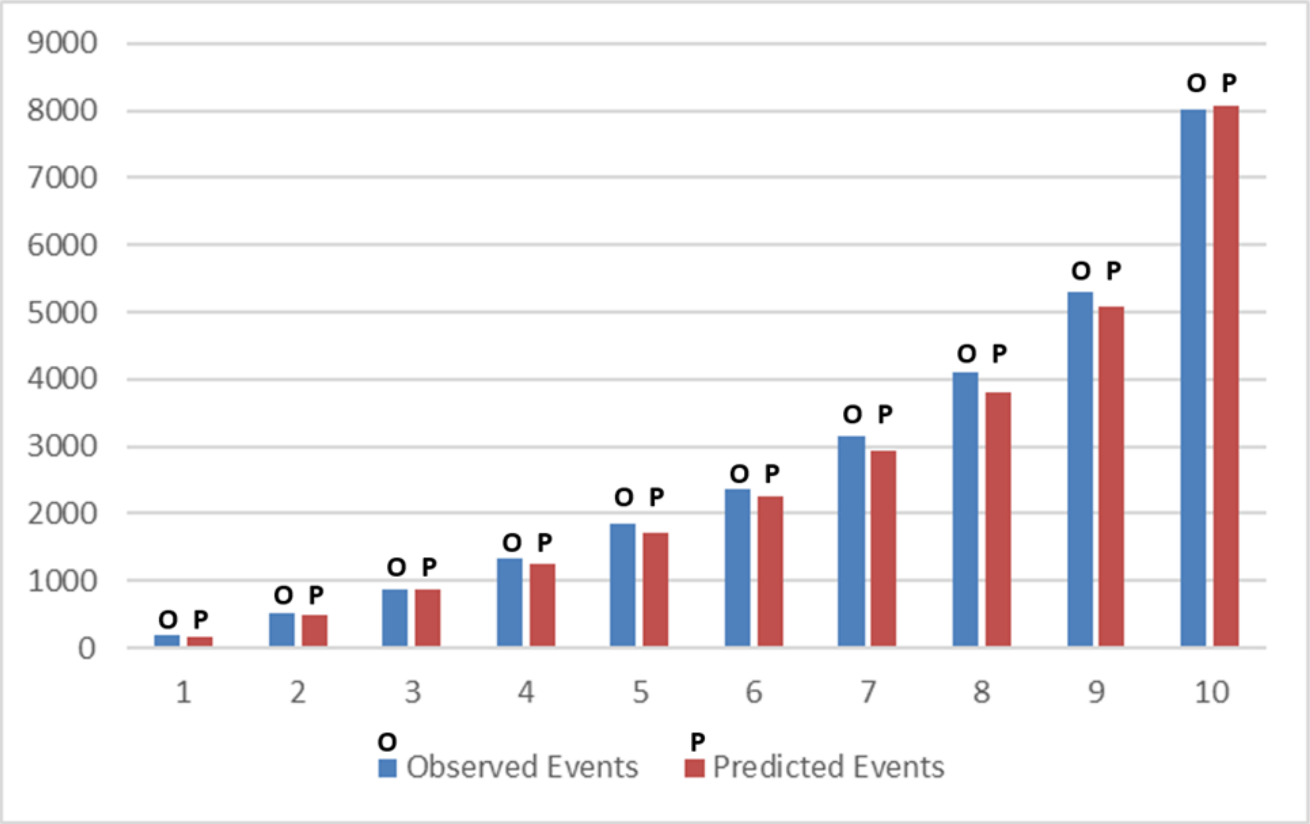
**[Response Ends]**

**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:** Observed and expected number of patients waitlisted by risk decile.



**[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, above in section 2b.29, shows that in no decile is there an important discrepancy between the observed number of waitlisted patients in a decile and that predicted by the model.

**[Response Ends]**

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

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

**[Response Begins]**

N/A

**[Response Ends]**

## Criterion 3. Feasibility

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

---

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

**[Response Begins]**

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

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

**[Response Ends]**

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

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

**[Response Begins]**

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

**[Response Ends]**

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

**[Response Begins]**

N/A

**[Response Ends]**

**3.04. Describe any efforts to develop an eCQM.**

**[Response Begins]**

N/A

**[Response Ends]**

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

**[Response Begins]**

None identified.

**[Response Ends]**

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

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

**Attach the fee schedule here, if applicable.**

**[Response Begins]**

N/A

**[Response Ends]**

## Criterion 4: Use and Usability

### 4a. Use

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

---

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

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

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

**Name of program and sponsor**

**URL**

**Purpose**

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

**Level of measurement and setting**

**[Response Begins]**

Not in use

**[Not in use Please Explain]**

The measure is undergoing initial endorsement review.

**[Response Ends]**

**4a.02. Check all planned uses.**

**[Response Begins]**

Public reporting

Payment Program

**[Response Ends]**

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

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

**[Response Begins]**

The measure is undergoing initial endorsement review.

**[Response Ends]**

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

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

**[Response Begins]**

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

**[Response Ends]**

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

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

**[Response Begins]**

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

**[Response Ends]**

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

**[Response Begins]**

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

**[Response Ends]**

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

**[Response Begins]**

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

**[Response Ends]**

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

**[Response Begins]**

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

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

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

**[Response Ends]**

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

**[Response Begins]**

See 4a.08.

**[Response Ends]**

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

**[Response Begins]**

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

**[Response Ends]**

## 4b. Usability

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

**[Response Begins]**

The measure is not yet implemented in a public reporting program, so improvement could not be evaluated.  CMS currently anticipates implementation of this 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]**

**[Response Ends]**

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

*(Can search and select measures.)*

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

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

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

**[Response Ends]**

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

**[Response Begins]**

Yes

**[Response Ends]**

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

**[Response Begins]**

N/A

**[Response Ends]**

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

*Provide analyses when possible.*

**[Response Begins]**
N/A
**[Response Ends]**

## Appendix

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

## Contact Information

**Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
**Measure Steward Point of Contact:** Dollar-Maples, Helen, helen.dollar-maples@cms.hhs.gov

**Measure Developer if different from Measure Steward:** University of Michigan Kidney Epidemiology and Cost Center
**Measure Developer Point(s) of Contact:** Parrotte, Casey, parrotte@med.umich.edu
Sardone, Jennifer, jmsto@med.umich.edu
George, Jaclyn, jaclynrg@med.umich.edu
Dollar-Maples, Helen, helen.dollar-maples@cms.hhs.gov

## Additional Information

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

**[Response Begins]**
Available in attached file
**[Response Ends]**

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

*Describe the members' role in measure development.*

**[Response Begins]**
David Axelrod, MD, MBA
Transplant Surgeon, University of Iowa

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

**[Response Ends]**

**3. Indicate the year the measure was first released.**

**[Response Begins]**
2022

**[Response Ends]**

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

**[Response Begins]**
01/2022
**[Response Ends]**

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

**[Response Begins]**
Annual
**[Response Ends]**

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

**[Response Begins]**
04/2023
**[Response Ends]**

**7. Provide a copyright statement, if applicable. Otherwise, indicate “N/A”.**

**[Response Begins]**
N/A
**[Response Ends]**

**8. State any disclaimers, if applicable. Otherwise, indicate “N/A”.**

**[Response Begins]**
N/A
**[Response Ends]**

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

**[Response Begins]**
N/A
**[Response Ends]**