Content

Brief Measure Information

CBE #: 3753

Corresponding Measures:

Measure Title: Delay in Progression of Chronic Kidney Disease (CKD) Measure

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: The Delay in Progression of CKD Measure is an outcome measure to assess how well providers delay progression from Stage 4 CKD to end-stage renal disease (ESRD) requiring chronic dialysis. The measure includes adult Medicare Fee-For-Service (FFS) beneficiaries with Stage 4 CKD. The measure outcome captures beneficiaries with Stage 4 CKD who progress to ESRD and require chronic dialysis. This measure is for nephrology practices (also referred to as "providers" in this submission) who care for patients with Stage 4 CKD.

1b.01. Developer Rationale: This measure will directly benefit patients by assessing the quality of care delivered by nephrology-related entities to Medicare beneficiaries with CKD and highlighting opportunities for nephrology practices to improve care.

sp.12. Numerator Statement: The measure outcome is progression from Stage 4 CKD to ESRD requiring chronic dialysis in the measurement year for patients aged 19 and older with stage 4 CKD. The outcome of interest is defined as enrollment in ESRD or ESRD-Dialysis Medicare coverage. Not all possible patient events will be counted in the numerator.

The following censoring events are not counted as outcome events in the numerator if they occur during the measurement year prior to ESRD enrollment:

- Kidney transplant (prior to or within one month of beginning ESRD enrollment); or
- Enrollment in hospice; or
- Death

sp.14. Denominator Statement: The cohort includes Medicare Fee-For-Service beneficiaries (patients) who are 19 years and older, with Stage 4 CKD, who are not enrolled in Medicare ESRD or ESRD-dialysis, who are not enrolled in Medicare hospice, who have not had a kidney transplant within the past 12 months, and who are being treated by a nephrology practice.

Content
The measure uses a time-to-event methodology in which the denominator represents the person-time in which eligible patients are at risk for the outcome. The observed patient-level denominator is the amount of eligible person-time for which the patient is at risk for the outcome. The observed provider-level denominator is the total person-time of eligible patients attributed to the practice. **sp.16. Denominator Exclusions:** The cohort excludes patients with advanced or metastatic cancer, defined as specific cancer-related ICD-10 codes from an inpatient encounter. A full list of codes is available in the attached data dictionary.
Measure Type: Outcome **sp.28. Data Source:** Claims; Other (Beneficiary Enrollment data including the hospice enrollment, ESRD or dialysis enrollment). **sp.07. Level of Analysis:** Clinician: Group Practice
IF Endorsement Maintenance—Original Endorsement Date: N/A New Measure **Most Recent Endorsement Date:** N/A New Measure
IF this measure is included in a composite, Composite#/title: N/A **IF this measure is paired/grouped, CBE#/title:** N/A **sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A
Staff Assessment: New Measure
Criterion 1: Importance to Measure and Report
1a. <u>Evidence</u>. The evidence requirements for a ***health outcome*** measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance can be used, assuming the data are from a robust number of providers and the results are not subject to systematic bias. For measures derived from a patient report, the evidence also should

Content

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 clinician level that assesses how well providers delay progression from Stage 4 CKD to end-stage renal disease (ESRD) requiring chronic dialysis.
- The developer provides a logic model that depicts how services from nephrology providers can achieve a delay in initiation of dialysis, which tends to provide improved patient-centered care and quality of life and reduction of comorbidities associated with dialysis.

Summary:

- The developer cited multiple studies supporting the positive outcomes including longer survival, greater quality of life, and greater patient engagement in treatment choices associated with delayed dialysis.
- The developer cited multiple studies associating the initiation of dialysis with a high-risk burden to patients, including risk of infection, pain from dialysis procedures, and high psychosocial impact.
  The developer cited a randomized controlled trial which found that a 6-month delay in dialysis resulted in savings of $18,000 (otherwise spent on the dialysis treatment plus transportation and hospitalizations) with no difference in quality of life or survival.

Question for the Standing Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance From the Evidence Algorithm

Outcome measure (Box 1) -> Empirical data on the relationship between the outcome and at least one health care action provided (Box 2) -> Pass

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

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

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

- The developer provided a distribution of performance results, risk-standardized ratio (RSR) for all providers and separately among nephrology practices caring for at least 25 patients.
  - The developer highlighted substantial variation in performance between measured entities after accounting for clinical risk.
  - The developer noted the range of 0.604 -1.629 (a 2.7-fold increase in progression hazard between the best- and worst-quality nephrologists after accounting for case mix) indicates a large gap that can be explained by a

Content

meaningful difference in performance.
    - The developer noted there are fewer progression events on a risk-adjusted basis at top-performing practices, which highlights that entities will not all receive the same score and can therefore be differentiated from one another based on their quality and performance.
- The developer notes that because this is a new measure and has not yet been fully implemented, data on improvement is not yet available.

Disparities

- For disparities data, the developer provided the distributions of clinical risk-adjusted measure score at the provider level, among quintiles of providers based on the prevalence of patients with each social risk factor in the Progression Development Dataset.
- The social risk factors include patients of Black race, dual eligible status, patients from low SES neighborhoods, and patients living in urban areas.
- Mean and standard deviation of clinical risk-adjusted measure scores among nephrology practices with 25+ patients:
    - Patients of Black race - Quintile 1: 0.99 (0.12), Quintile 2: 0.99 (0.14), Quintile 3: 1.02, (0.14) Quintile 4: 1.01 (0.14), Quintile 5: 1.02 (0.14)
    - Dual eligible patients - Quintile 1: 1.02 (0.14), Quintile 2: 1.00 (0.14), Quintile 3: 1.00 (0.14) Quintile 4: 1.00 (0.13), Quintile 5: 1.01 (0.12)
    - Patients from low SES neighborhood - Quintile 1: 1.01 (0.12), Quintile 2: 1.00 (0.14), Quintile 3: 1.01 (0.14) Quintile 4: 1.00 (0.14), Quintile 5: 1.01 (0.13)
    - Patients living in urban areas - Quintile 1: 1.00 (0.14), Quintile 2: 1.00 (0.14), Quintile 3: 1.00 (0.14) Quintile 4: 1.02 (0.15), Quintile 5: 1.01 (0.12)
- The developer noted the distribution is fairly consistent across each quintile, illustrating that after accounting for differences in clinical case mix, the risk of progression from stage 4 CKD to stage 5 or to ESRD does not depend substantially on the proportion of patients served who are Black, dual-eligible, low-SES, or urban residents.

Questions for the Standing Committee:

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

Preliminary rating for opportunity for improvement:

☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criteria 2: Scientific Acceptability of Measure Properties

Evaluators: Staff/William White

2a. Reliability: Specifications and Testing

Content
2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.
2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.
Specifications: • Measure specifications are clear and precise.
Reliability Testing: • Reliability testing conducted at the Accountable Entity Level: o The Progression Development dataset (Medicare FFS administrative claims and enrollment data) used in this analysis included 2,854 measure entities (clinician groups) which billed for nephrology services to Medicare FFS patients 18 years or older. These clinician groups were grouped by taxpayer identification number (TIN). o Data were obtained through CMS Chronic Conditions Data Warehouse and Virtual Research Data Center (CCW/VRDC) and the CMS integrated data repository (IDR). o The dataset for the analysis covered 1/1/2017 to 12/31/2018 and included 434,764 patients. o Social risk factors used in the analysis were dual eligible status, AHRQ-validated SES index score, race (black vs non-black) and urbanicity (2013 Rural-Urban Continuum Codes). o Signal-to-noise reliability was calculated all entities and statistics (mean, standard deviation, median, IQR, minimum, maximum) were calculated for all nephrology practices and those with at least 25 cases, a common threshold among other risk-adjusted claims-based measures in the CMS programs. o Signal-to-noise reliability based on Adams et al. approach. To estimate the overall signal and noise, the ICC was calculated for the Model Participant, j, using the estimates of between-entity variance τ2 and the formula for intraclass correlation coefficient (ICC) presented by Shrout and Fleiss. Specifically, the signal-to-noise reliability score for Model Participant, j, Rj is calculated as: $R_j = (n_j * ICC) / (1 + (n_j - 1) * ICC)$ $ICC = \tau^2 / (\tau^2 + \pi^2/6\gamma^2)$ o nj is the number of patients for the nephrologist j, τ2 is the between agency variance in a Weibull model with lognormal frailty that used to approximate the Cox model with lognormal frailty specified above and represent the signal, and π2⁄6γ2 represents the noise and γ is the shape parameter of the Weibull distribution. o Rj ranges from 0 to 1.0. The higher the score, the higher the reliability. o Among all nephrology practices, the mean reliability score was 0.614 and the median score was 0.696 which are considered moderate.

Content
○ Among the 1,970 nephrology practices with at least 25 cases, the mean reliability score was 0.787 and the median score was 0.821 which are considered moderate.
Questions for the Standing Committee regarding reliability: • *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*
Guidance From the Reliability Algorithm • Submitted specifications are precise, unambiguous and complete (Box 1) -> Empirical reliability testing conducted using statistical tests with the measure as specified (Box 2) -> Reliability testing conducted with computed performance measure scores for each measured entity (Box 4) -> Method described was appropriate for assessing the proportion of variability due to real difference among measured entities. Signal-to-noise analysis performed (Box 5) -> Moderate (Box 6b) • The highest possible rating is high.
Preliminary rating for reliability: ☐ High ☒ Moderate ☐ Low ☐ Insufficient
2b. Validity: Validity Testing; Exclusions; Risk Adjustment; Meaningful Differences; Comparability; Missing Data
2b2. Validity testing should demonstrate that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.
2b2-2b6. Potential threats to validity should be assessed/addressed. **Validity Testing** • Patient/encounter-level ○ Data element validation completed twice in 2 datasets (Progression EHR Datasets A and B) ○ Determined percent agreement between patients with at least one outpatient encounter in a CY for Stage 4 CKD and the presence of a confirmatory lab value ○ Match rate for beneficiary visits and confirmatory lab values (eGFR) 88.1% (Dataset A) and 83.5% (Dataset B) ○ Analyzed alignment between ESRD enrollment and documentation of dialysis facility billing codes • Empirical validity ○ Evaluated statistical performance of claims-only risk model by comparing with a model including eGFR data (clinical gold standard) abstracted from EHRs using 2 validation datasets ○ Inclusion of eGFR in risk model resulted in modest improvement to C-statistic: 0.794 to 0.865, Dataset A and

Content
0.779 to 0.843, Dataset B ○ Weighed against additional burden of reporting eGFR values, marginal improvement of adding eGFR data is not greatly meaningful for already strong performance of claims-only model • Face validity ○ Developer surveyed a technical expert panel of 15 members and asked if members believe the measure can be used to distinguish provider quality ○ 5/15 strongly agree, 6/15 somewhat agree, 3/15 somewhat disagree, 1/15 strongly disagree ○ Two members expressed concerns about the absence of eGFR data in the measure
Exclusions • Numerator exclusions: kidney transplant prior to or within one month of beginning ESRD enrollment, enrollment in hospice within one year prior to ESRD enrollment, death within one year prior to ESRD enrollment • Denominator exclusions: patients with advanced or metastatic cancer
Risk Adjustment • Measure is risk adjusted using an exponential function to determine probability of progression using a time-to-event outcome; risk-standardized ratio calculated by dividing number of predicted events by number of expected events. • Risk standardized rate may be calculated by multiplying ratio by national incidence rate per 100 patient-years. • Social risk factors were considered (dual eligibility, low AHRQ, race and urbanicity) but were not included in the final model, as there were no statistically significant relationships between any of the social risk factors tested and measure scores among nephrology practices with the highest proportion of patients with social risk factors.
Meaningful Differences • The developer examined summary statistics and compared top vs bottom quintiles: ○ Mean of top quintile (0.834) and top 4 quintiles (0.956) statistically different from bottom quintile ($p < 0.0001$). ○ Reflects a 1.45-fold increase in hazard attributable to differences in care; increase in risk considered meaningful by consulted nephrologists.
Missing Data • The developer used claims-based data used for development and testing; lack of a claim is treated as not having the corresponding diagnosis or procedure; all administratively coded diagnoses were available for measurement.
Comparability • The measure only uses one set of specifications for this measure
Questions for the Standing Committee regarding validity: • *Do you have any concerns with the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?*
Guidance From the Validity Algorithm All potential threats assessed (Box 1) -> Face validity performed (Box 2) -> Measure considered meaningful by experts (Box 3) ->

Content
Moderate (Box 4) The highest possible rating is moderate.
Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 3. Feasibility
3. <u>Feasibility</u> 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 reported that the data elements needed to compute the performance scores are coded by someone other than the person obtaining original information. • The developer noted that all data elements are in defined fields in a combination of electronic sources. • The developer stated that there are no fees or licensing requirements to use this measure as specified. • The developer also noted there are currently no efforts underway to develop an eCQM.
Questions for the Standing Committee: • *Is the data collection strategy ready to be put into operational use?*
Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 4: Use and Usability
4a. Use (4a1. <u>Accountability and Transparency</u>; 4a2. <u>Feedback on measure</u>)
4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.
4a1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.
Current uses of the measure Publicly reported? ☐ Yes ☒ No

Content			
Current use in an accountability program?	☐ Yes	☒ No	☐ UNCLEAR
Planned use in an accountability program?	☒ Yes	☐ No	☐ N/A

Accountability program details

- The developer noted that this measure is not yet in use because it is a new measure.
- The developer indicated planned use in payment programs and quality improvement with benchmarking (external benchmarking to multiple organizations).
- The developer highlights the measure is planned to be implemented in the voluntary Kidney Care Choices model to help reduce cost and improve quality of care for patients with late-stage CKD and ESRD while delaying the need for dialysis and encouraging kidney transplantation.

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

Feedback on the measure provided by those being measured or others

- The developer reported that CORE included several practicing nephrologists in the technical expert panel (TEP) and they offered the following feedback:
  - Delaying the progression of CKD is an important outcome for both providers and patients, and noted the lack of care coordination between providers, which could be ameliorated by the implementation of such a measure.
  - The measure construct is useful and a valid aspect of quality to measure.
  - There are no current measures that focus on the prevention of CKD progression.
  - The measure is valid as specified and appropriate to use in the context of a voluntary payment model.
- The developer noted that they obtained additional feedback from the CMS Innovation Center Kidney Care Choices model team to ensure the measure specifications align with the goals and requirements of the model.
- The developer noted that input from nephrologists in the TEP, clinical subject matter experts, and from the CMS Innovation Center Kidney Care Choices model team, played a key role in shaping development of the measure.

Questions for the Standing Committee:

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

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

Content
4b. Usability (4b1. Improvement; 4b2. Benefits of measure)
4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.
4b1 Improvement. Progress toward achieving the goal of high quality, efficient healthcare for individuals or populations is demonstrated. **Improvement results** • The developer stated the measure has not yet been implemented; there are no measured entities among which to assess performance or improvement.
4b2. Benefits versus harms. The benefits of the performance measure in facilitating progress toward achieving high quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists). **Unexpected findings (positive or negative) during implementation** • Not applicable; the measure has not yet been implemented. **Potential harms** • None identified.
Questions for the Standing Committee: • *How can the performance results be used to further the goal of high quality, efficient healthcare?* • *Do the benefits of the measure outweigh any potential unintended consequences?*
Preliminary rating for Usability and Use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 5: Related and Competing Measures
Related Measures • CBE #1662 Adult Kidney Disease Angiotensin Converting Enzyme (ACE) or Angiotensin Receptor Blocker (ARB) Therapy measure

Content
Harmonization • The developer stated that while the ACE/ARB Therapy process measure is conceptually related, it bears very little overlap in specifications to the Progression Delay outcome measure and any need for harmonization is minimal. • The developer indicated there is some overlap in target population, but the ACE/ARB Therapy measure is designed for a broad cohort of patients across a range of clinical settings while the Progression Delay measure is focused specifically on patients with stage 4 CKD being cared for by nephrology practices.

QUALITY MEASURE SUBMISSION FORM

Version: 1.0; Generated: 13 April 2023

Introduction

Thank you for your interest in submitting a measure to Battelle for possible endorsement.

What criteria are used to evaluate measures? Measures are evaluated on standardized criteria: importance to measure and report, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures. For your measure to be evaluated against these measure evaluation criteria, you must complete the measure submission form.

Why do I have to complete a form? Due to the volume and/or complexity of proposed measures, Battelle provides measure information to committee reviewers in a standardized format to facilitate their evaluation of whether the measure meets the measure evaluation criteria. This form allows the measure steward to present information demonstrating that the proposed measure meets endorsement criteria.

What is on the form? The information requested in this form is directly related to the measure evaluation criteria.

Can't I just submit our files for consideration? No. Measures must be submitted through the online form to be considered for the Spring 2023 cycle. Requested information should be entered directly into this form and as well as any necessary or required attachments.

Can I submit additional details and materials? Additional materials will be considered only as supplemental. Do NOT rely on material provided in an appendix to provide measure specifications or to demonstrate meeting the criteria. The core information needed to evaluate the measure should be provided in the appropriate submission form fields and required attachments. Please contact PQMsupport@battelle.org regarding questions about submitting supplemental materials.

What do I do first? If you have started a new submission by answering five qualifying questions, you may proceed to the “Previous Submission Information" tab to continue with your submission. The “Conditions” tab will list the conditions that must be met before your proposed measures may be considered and evaluated for suitability as endorsed voluntary consensus standards. You are asked to acknowledge reading and accepting the conditions.

Can I make changes to a form once I have submitted it? No. Once you submit your measure, you will NOT be able to return to this submission form to make further revisions. You will need to contact project staff.

What if I need additional help? Please contact the project staff at PQMsupport@battelle.org if you have questions regarding the information requested or submitting supplemental materials.

NOTE: All measure submissions should be 508-compliant. Refer to the Checklist for Developer 508 Guidelines (PDF) to ensure all guidelines apply to all parts of your submission, including all fields and attachments used within the measure submission form.

Please email us at PQMsupport@battelle.org if you experience technical difficulties using the online submission form.

Thank you for your interest in submitting measures to Battelle.

Previous Submission Information (1 – 4)

1) Select whether this measure was previously submitted to the prior consensus-based entity (the National Quality Forum [NQF]) and given an identifying number.

☐ Previously submitted to NQF
☒ x New measure, never submitted.

2) Provide the measure number of the previously submitted measure.

Not applicable; this is a new measure.

3) If the measure has an electronic clinical quality measure (eCQM) version, provide the measure number of the previously submitted measure.

Not applicable; this is a new measure.

4) If this eCQM has a registry version, provide the measure numbers of the previously submitted measure.

Not applicable; this is a new measure.

Conditions (1 - 2)

Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. If any of the conditions are not met, the measure will not be accepted for consideration.

A. A Measure Steward Agreement is signed or the steward is a government organization. (All non-government organizations must sign a Measure Steward Agreement.) For more information about completing a Measure Steward Agreement, please go to: Endorsement | Partnership for Quality Measurement (p4qm.org) and follow the instructions.

B. The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.

C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.

D. The measure is fully specified and tested for reliability and validity.

E. The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.

F. The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.

1) Check if either of the following apply.

☐ Proprietary measure or components (e.g., risk model, codes)

☐ Proprietary measure or components with fees

☒ x None of the above

2) Check the box below to agree to the conditions listed above.

☒ x I have read and accept the conditions as specified above

Specifications: Maintenance Update (spma.01 - spma.02)

Not applicable; this is a new measure.

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

☐ *No*

☐ *Yes*

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

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

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

Measure Specifications (sp.01 - sp.32)

sp.01) Provide the measure title.

Measure titles should be concise yet convey who and what is being measured.

Delay in Progression of Chronic Kidney Disease (CKD) Measure

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).

The Delay in Progression of CKD Measure is an outcome measure to assess how well providers delay progression from Stage 4 CKD to end-stage renal disease (ESRD) requiring chronic dialysis. The measure includes adult Medicare Fee-For-Service (FFS) beneficiaries with Stage 4 CKD. The measure outcome captures beneficiaries with Stage 4 CKD who progress to ESRD and require chronic dialysis. This measure is for nephrology practices (also referred to as "providers" in this submission) who care for patients with Stage 4 CKD.

sp.03) Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

Not applicable – not a paired measure.

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

☐ Behavioral Health
☐ Behavioral Health: Alcohol, Substance Use/Abuse
☐ Behavioral Health: Anxiety
☐ Behavioral Health: Attention Deficit Hyperactivity Disorder (ADHD)
☐ Behavioral Health: Bipolar Disorder
☐ Behavioral Health: Depression
☐ Behavioral Health: Domestic Violence
☐ Behavioral Health: Other Serious Mental Illness
☐ Behavioral Health: Post-Traumatic Stress Disorder (PTSD)
☐ Behavioral Health: Schizophrenia
☐ Behavioral Health: Suicide
☐ Cancer
☐ Cancer: Bladder
☐ Cancer: Breast
☐ Cancer: Colorectal
☐ Cancer: Gynecologic

- ☐ Cancer: Hematologic
- ☐ Cancer: Liver
- ☐ Cancer: Lung, Esophageal
- ☐ Cancer: Prostate
- ☐ Cancer: Renal
- ☐ Cancer: Skin
- ☐ Cancer: Thyroid
- ☐ Cardiovascular
- ☐ Cardiovascular: Arrythmia
- ☐ Cardiovascular: Congestive Heart Failure
- ☐ Cardiovascular: Coronary Artery Disease
- ☐ Cardiovascular: Coronary Artery Disease (AMI)
- ☐ Cardiovascular: Coronary Artery Disease (PCI)
- ☐ Cardiovascular: Hyperlipidemia
- ☐ Cardiovascular: Hypertension
- ☐ Cardiovascular: Secondary Prevention
- ☐ Critical Care
- ☐ Critical Care: Assisted Ventilation
- ☐ Critical Care: Intensive Monitoring
- ☐ Dental
- ☐ Dental: Caries
- ☐ Dental: Tooth Loss
- ☐ Ears, Nose, Throat (ENT)
- ☐ Ears, Nose, Throat (ENT): Ear Infection
- ☐ Ears, Nose, Throat (ENT): Hearing
- ☐ Ears, Nose, Throat (ENT): Pharyngitis
- ☐ Ears, Nose, Throat (ENT): Tonsilitis
- ☐ Endocrine
- ☐ Endocrine: Calcium and Metabolic Bone Disorders
- ☐ Endocrine: Diabetes
- ☐ Endocrine: Female and Male Endocrine Disorders
- ☐ Endocrine: Hypothalamic-Pituitary Disorders
- ☐ Endocrine: Thyroid Disorders
- ☐ Eye Care
- ☐ Eye Care: Age-related macular degeneration (AMD)
- ☐ Eye Care: Cataracts
- ☐ Eye Care: Diabetic retinopathy
- ☐ Eye Care: Glaucoma
- ☐ Gastrointestinal (GI)
- ☐ Gastrointestinal (GI): Constipation

☐ Gastrointestinal (GI): Gall Bladder Disease
☐ Gastrointestinal (GI): Gastroenteritis
☐ Gastrointestinal (GI): Gastro-Esophageal Reflux Disease (GERD)
☐ Gastrointestinal (GI): Hemorrhoids
☐ Gastrointestinal (GI): Hernia
☐ Gastrointestinal (GI): Inflammatory Bowel Disease
☐ Gastrointestinal (GI): Irritable Bowel Syndrome
☐ Gastrointestinal (GI): Peptic Ulcer
☐ Genitourinary (GU)
☐ Genitourinary (GU): Benign Prostatic Hyperplasia
☐ Genitourinary (GU): Erectile Dysfunction/Premature Ejaculation
☐ Genitourinary (GU): Incontinence/pelvic floor disorders
☐ Genitourinary (GU): Prostatitis
☐ Genitourinary (GU): Urinary Tract Injection (UTI)
☐ Gynecology (GYN)
☐ Gynecology (GYN): Abnormal bleeding
☐ Gynecology (GYN): Endometriosis
☐ Gynecology (GYN): Infections
☐ Gynecology (GYN): Menopause
☐ Gynecology (GYN): Pelvic Pain
☐ Gynecology (GYN): Uterine fibroids
☐ Infectious Diseases (ID)
☐ Infectious Diseases (ID): HIV/AIDS
☐ Infectious Diseases (ID): Influenza
☐ Infectious Diseases (ID): Lyme Disease
☐ Infectious Diseases (ID): Meningococcal Disease
☐ Infectious Diseases (ID): Pneumonia and respiratory infections
☐ Infectious Diseases (ID): Sepsis
☐ Infectious Diseases (ID): Sexually Transmitted
☐ Infectious Diseases (ID): Tuberculosis
☐ Liver
☐ Liver: Viral Hepatitis
☐ Musculoskeletal
☐ Musculoskeletal: Falls and Traumatic Injury
☐ Musculoskeletal: Gout
☐ Musculoskeletal: Joint Surgery
☐ Musculoskeletal: Low Back Pain
☐ Musculoskeletal: Osteoarthritis
☐ Musculoskeletal: Osteoporosis
☐ Musculoskeletal: Rheumatoid Arthritis

☐ Neurology
☐ Neurology: Alzheimer's Disease
☐ Neurology: Autism
☐ Neurology: Brain Injury
☐ Neurology: Epilepsy
☐ Neurology: Migraine
☐ Neurology: Parkinson's Disease
☐ Neurology: Spinal Cord Injury
☐ Neurology: Stroke/Transient Ischemic Attack (TIA)
☐ Other (please specify here:)
☐ Palliative Care and End-of-Life Care
☐ Palliative Care and End-of-Life Care: Advanced Directives
☐ Palliative Care and End-of-Life Care: Amyotrophic Lateral Sclerosis (ALS)
☐ Palliative Care and End-of-Life Care: Hospice Management
☐ Palliative Care and End-of-Life Care: Inappropriate use of acute care services
☐ Palliative Care and End-of-Life Care: Pain Management
☐ Perinatal Health
☐ Perinatal Health: Labor and Delivery
☐ Perinatal Health: Newborn Care
☐ Perinatal Health: Post-Partum Care
☐ Perinatal Health: Preconception Care
☐ Perinatal Health: Prenatal Care
☐ Renal
☐ Renal: Acute Kidney Injury
☒ x Renal: Chronic Kidney Disease (CKD)
☒ x Renal: End Stage Renal Disease (ESRD)
☐ Renal: Infections
☐ Reproductive Health
☐ Reproductive Health: Family planning and contraception
☐ Reproductive Health: Infertility
☐ Reproductive Health: Male reproductive health
☐ Respiratory
☐ Respiratory: Acute Bronchitis
☐ Respiratory: Allergy
☐ Respiratory: Asthma
☐ Respiratory: Chronic Obstructive Pulmonary Disease (COPD)
☐ Respiratory: Dyspnea
☐ Respiratory: Pneumonia
☐ Respiratory: Sleep Apnea
☐ Surgery

☐ Surgery: Cardiac Surgery
☐ Surgery: Colorectal
☐ Surgery: Neurosurgery / Spinal
☐ Surgery: Orthopedic
☐ Surgery: Orthopedic Hip/Pelvic Fractures
☐ Surgery: Pediatric
☐ Surgery: Perioperative and Anesthesia
☐ Surgery: Plastic
☐ Surgery: Thoracic Surgery
☐ Surgery: Trauma
☐ Surgery: Vascular Surgery

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

☒ x Access to Care
☒ x Care Coordination
☐ Care Coordination: Readmissions
☐ Care Coordination: Transitions of Care
☐ Disparities Sensitive
☐ Health and Functional Status
☒ x Health and Functional Status: Change
☐ Health and Functional Status: Nutrition
☐ Health and Functional Status: Obesity
☐ Health and Functional Status: Physical Activity
☐ Health and Functional Status: Quality of Life
☐ Health and Functional Status: Total Health
☐ Immunization
☐ Other (please specify here:)
☐ Person-and Family-Centered Care: Person-and Family-Centered Care
☐ Person-and Family-Centered Care: Workforce
☐ Primary Prevention
☐ Primary Prevention: Nutrition
☐ Primary Prevention: Tobacco Use
☐ Safety
☐ Safety: Complications
☐ Safety: Healthcare Associated Infections
☐ Safety: Medication
☐ Safety: Overuse
☐ Screening

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.

☒ x Adults (Age >= 18)
☐ Children (Age < 18)
☐ Elderly (Age >= 65)
☐ Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
☐ Populations at Risk: Individuals with multiple chronic conditions
☐ Populations at Risk: Veterans
☐ Women

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.

☐ Accountable Care Organization
☒ x Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☐ Integrated Delivery System
☐ Other (please specify here:)
☐ Population: Community, County or City
☐ Population: Regional and State

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

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

☒ x Ambulatory Care
☐ Behavioral Health
☐ Home Care
☐ Inpatient/Hospital
☐ Other (please specify here:)
☐ Outpatient Services
☐ Post-Acute Care

sp.09) Provide a Uniform Resource Locator (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".

None available.

sp.10) Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the measure authoring tool (MAT) for eCQMs - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications). HQMF specifications are attached.

☒ x HQMF specifications are NOT attached (Please explain).

Not applicable; measure is not an eCQM.

sp.11) Attach the simulated testing attachment.

All eCQMs require a simulated testing attachment to confirm that the HTML output from Bonnie testing (or testing of some other simulated data set) includes 100% coverage of measured patient population testing, with pass/fail test cases for each sub-population. This can be submitted in the form of a screenshot.

☐ Testing is attached
☒ x Testing is NOT attached (please explain)

Not applicable; measure is not an eCQM.

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

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

☒ x Available in attached Excel or csv file
☐ No data dictionary/code table – all information provided in the submission form

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

sp.13) State the numerator.

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

DO NOT include the rationale for the measure.

The measure outcome is progression from Stage 4 CKD to ESRD requiring chronic dialysis in the measurement year for patients aged 19 and older with stage 4 CKD. The outcome of interest is defined as enrollment in ESRD or ESRD-Dialysis Medicare coverage. Not all possible patient events will be counted in the numerator.

The following censoring events are not counted as outcome events in the numerator if they occur during the measurement year prior to ESRD enrollment:

Kidney transplant (prior to or within one month of beginning ESRD enrollment); or
Enrollment in hospice; or
Death.

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

sp.14) Provide details needed to calculate the numerator.

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

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

The observed outcome at the patient level is progression from stage 4 CKD to ESRD during the measurement period (provided the patient is eligible for the cohort and considered "at risk" for the outcome as discussed below in Sp.15 "State the denominator"). The observed outcome at the provider level is the total number of progression events for each provider among "at risk" patients included in the measure who are attributed to that provider.

"Progression to ESRD" is identified as the date on which a patient with stage 4 CKD enrolls in in ESRD or ESRD-Dialysis Medicare coverage, based on the Medicare Enrollment Database (EDB).

There are three censoring events that are not counted toward the progression outcome. Patients who die or who enroll in Medicare hospice, without requiring chronic dialysis during the measurement year, are eligible for the outcome until the point of death or hospice enrollment, at which point they are no longer "at risk" for the outcome of progression. Patients who have a kidney transplant prior to or within one month of ESRD enrollment are similarly censored from the date of the transplant. These censoring events are discussed in greater detail in sp.16 (denominator details).

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

sp.15) State the denominator.

Brief, narrative description of the target population being measured.

The cohort includes Medicare Fee-For-Service beneficiaries (patients) who are 19 years and older, with Stage 4 CKD, who are not enrolled in Medicare ESRD or ESRD-dialysis, who are not enrolled in Medicare hospice, who have not had a kidney transplant within the past 12 months, and who are being treated by a nephrology practice.
The measure uses a time-to-event methodology in which the denominator represents the person-time in which eligible patients are at risk for the outcome. The observed patient-level denominator is the amount of eligible person-time for which the patient is at risk for the outcome. The observed provider-level denominator is the total person-time of eligible patients attributed to the practice.

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

sp.16) Provide details needed to calculate the denominator.

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

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

This measure includes Medicare FFS patients:

1. With stage 4 CKD during the measurement period (defined as at least one occurrence of International Classification of Diseases, 10th Revision (ICD-10) code N18.4 "Chronic kidney disease, Stage 4 (Severe)" in at least one claim during the performance year), and
2. Who are age 18 or older in the year prior to the measurement period (that is, age 19 or older in the measurement period), and
3. With continuous enrollment in Medicare FFS Parts A and B for one full year prior to the performance year as well as the full performance year, or until the date at which they are censored (due to death, kidney transplant, or hospice enrollment) or enroll in Medicare ESRD in the performance year, and
4. Who are attributed to a nephrology practice.

This measure does not include:

1. Patients already enrolled in Medicare ESRD or ESRD-Dialysis coverage, or
2. Patients already enrolled in the Medicare hospice program, or
3. Patients who have had a kidney transplant within the 12 months prior to becoming otherwise eligible for the cohort.

ESRD and Hospice patients are identified from the Medicare Enrollment Database. Kidney transplants are defined as a patient with one of the codes in Table sp.16:1 below.

Table sp.16:1. Transplant Codes Not Counted in Measure Outcome, from Agency for Healthcare Research and Quality Clinical Classifications Software (AHRQ CCS) 105

Code	Description of Code
0TY00Z0	Transplantation of Right Kidney, Allogeneic, Open Approach

Code	Description of Code
0TY00Z1	Transplantation of Right Kidney, Syngeneic, Open Approach
0TY00Z2	Transplantation of Right Kidney, Zooplastic, Open Approach
0TY10Z0	Transplantation of Left Kidney, Allogeneic, Open Approach
0TY10Z1	Transplantation of Left Kidney, Syngeneic, Open Approach
0TY10Z2	Transplantation of Left Kidney, Zooplastic, Open Approach
50360-50365	Kidney transplant
50380	Kidney transplant
S2065	Kidney transplant

Patient attribution: Nephrology practices responsible for a patient's care are defined as those having at least two encounters with the patient, providing nephrology specialty services (with specialty code 39) during the performance year. Eligible patient visits were defined as those with the Healthcare Common Procedure Coding System (HCPCS) Evaluation/Management (E&M) service codes listed in Table sp.16:2. If a patient visited multiple practices that provide specialty care, the patient is attributed, 1) to the practice that provided highest number of E&M claims to the patient; or if there is a tie, 2) to the practice that billed the most for those services; or there is still a tie, 3) the practice who provided the most recent service; or if there is still a tie, 4) a randomly selected practice. This approach aligns with the CMS Kidney Care Choices Model strategy to identify nephrology practices [1].

Table sp.16:2. E&M HCPCS Codes Identifying Clinician Groups Who Delivered Nephrology Specialty Services

HCPCS Codes	Service
99201-99205, 99211-99215	Office/Outpatient Visit E/M
99354-99355	Prolonged E/M
99495-99496	Transitional Care Management Services
99497-99498	Advance Care Planning
G0402, G0438, G0439	Welcome to Medicare and Annual Wellness Visits
99490	Chronic Care Management Services

Denominator calculation (time at risk for included cohort): The raw denominator is the sum of at-risk time for the cohort. The at-risk time from each patient is calculated from when the patient becomes eligible for the cohort in the measurement period (detailed below) until the earliest time of either: death; enrollment in ESRD or ESRD-Dialysis; a kidney transplant; enrollment in Medicare hospice; or end of the measurement period.

The measure considers the length of time a patient is eligible (at risk) for the outcome. As noted above, patients must have 12 months of claims data prior to the performance period and at least one Stage 4 CKD claim during the measurement year to be eligible. An eligible patient will enter the cohort and begin contributing at-risk time once they are both attributed to a nephrology practice and have a confirmed diagnosis of Stage 4 CKD, either:

- The beginning of the performance year (January 1) if the patient had both Stage 4 CKD and had two encounters with their attributed nephrology practice in the previous year, or
- During the performance year on the date once the patient is both 1) attributed to a nephrology practice (as described above) *and* 2) has a Stage 4 CKD diagnosis

The patient will leave the cohort (that is stop contributing at-risk time) on the first date of any of the following:

- Patient enrolls in ESRD or ESRD-Dialysis (outcome event)
- Patient receives a kidney transplant before or within one month of ESRD enrollment
    - Censoring event: not counted in outcome; no longer eligible for the outcome for the remainder of the measurement year.
- Patient dies or enrolls in Medicare hospice
    - Censoring event: not counted in outcome; no longer eligible for the outcome for the remainder of the measurement year.
- End of the measurement year

The total person-time contributed by a patient is the span between the date of entering the cohort and the date of leaving the cohort.

References:

1. Request for Applications (RFA): Kidney Care Choices (KCC) Model. 2019.Centers for Medicare & Medicaid Services (CMS), Center for Medicare and Medicaid Innovation (CMMI). https://innovation.cms.gov/files/x/kcc-rfa.pdf

sp.17) Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

The cohort excludes patients with advanced or metastatic cancer, defined as specific cancer-related ICD-10 codes from an inpatient encounter. A full list of codes is available in the attached data dictionary.

sp.18) Provide details needed to calculate the denominator exclusions.

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

Patients are excluded if coded in an inpatient setting with advanced or metastatic cancer in the year prior to the measurement year. Specific ICD-10 codes are from the following Condition Categories: CC8, CC10, CC12, CC177, CC178. A list of codes is in the data dictionary.

Rationale: The outcome for these patients is likely more influenced by cancer treatment than care associated with their chronic kidney disease, or nephrologist. Additionally, many patients in this population may be too ill for dialysis and have a high risk of mortality; thus, we find it inappropriate to attribute outcomes for these patients to their nephrologists' quality of care.

sp.19) Provide all information required to stratify the measure results, if necessary.

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

This measure is not currently stratified.

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

☐ Yes
☒ x No

sp.21) Select the risk adjustment type.

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

☐ No risk adjustment or risk stratification
☒ x Statistical risk model
☐ Stratification by risk category/subgroup (specify number of risk factors)
☐ Other approach to address risk factors (please specify here:)

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

Attachment: If available, please provide a sample report.

☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g. average
☐ Count
☐ Frequency Distribution
☐ Non-weighted score/composite/scale
☐ Other (please specify here:)
☐ Rate/proportion
☒ x Ratio
☐ Weighted score/composite scale

sp.23) Select the appropriate interpretation of the measure score.

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

☐ Better quality = Higher score
☒ x Better quality = Lower score
☐ Better quality = Score within a defined interval
☐ Passing score defines better quality

sp.24) Diagram or describe the calculation of the measure score as an ordered sequence of steps.

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

Calculation Algorithm

First, identify the cohort of patients meeting all inclusion criteria:

- Patient is enrolled in Medicare FFS Parts A and B for one full year prior to the performance year as well as the full performance year, or until the date of outcome (ESRD enrollment) or censoring (due to death, hospice enrollment, or kidney transplant) in the performance year,
- Patient is at least 18 years old at the start of the year prior to the performance year (that is, at least 19 years old at the start of the performance year),
- Patient has at least one occurrence of ICD-10 code N18.4: "CKD, Stage 4 (Severe)" in at least one claim during the performance year,
- Patient is not already enrolled in Medicare ESRD or hospice, and
- If patient had a prior kidney transplant, at least one year has passed post-transplant.

Second, apply exclusions to the cohort:

- Metastatic and advanced cancers, defined as specific cancer-related ICD-10 codes from an inpatient encounter. Patients are excluded if coded with advanced or metastatic cancer within one year prior to the earlier date of either being attributed to a nephrology practice or being diagnosed with Stage 4 CKD in the measurement year.

Third, attribute patients to nephrology practices as detailed above in section sp.16) Provide details needed to calculate the denominator.

Fourth, calculate the event of interest, which is the development of ESRD or ESRD-Dialysis. The measure uses a time-to-event outcome, which incorporates not only whether progression to ESRD with initiation of chronic dialysis occurred, but also the elapsed time from Stage 4 CKD to ESRD with initiation of chronic dialysis. The start time is the beginning of the performance year if there is at least one Stage 4 CKD claim observed in the prior year. The end time is the date of the first observed enrollment dates of development of either ESRD or ESRD-Dialysis, date of receiving a kidney transplant, date of death, date of hospice enrollment, or the end of the performance year. The precise methodology for cohort eligibility detailed in sp.16 Details needed to calculate the denominator.

Measure score calculation method for time-to-event outcome:

Assume that the hazard function of an event for patient *i* serviced by provider *j*, with a vector of risk factors X_{ij} is defined as a frailty model under the proportional hazard framework:

$$h_{ij}(t_{ij}) = w_j h_0(t_{ij}) \exp(\boldsymbol{X_{ij}\beta}),$$

where the w_j is the frailty for each provider *j* (that is, the provider-level hazard effect).

So, for the patient *ij*, define the predicted probability of progression (that is, predicted for a patient with the same clinical risk factors with that specific nephrologist) at time t as cumulative hazard at the time t_{ij}

$$P_{ij} = H_{ij}(t_{ij}) = \int_0^{t_{ij}} w_j h_0(t) \exp(X_{ij}\beta)\, dt = w_j \exp(X_{ij}\beta) \int_0^{t_{ij}} h_0(t) dt = w_j \exp(X_{ij}\beta) H_0(t_{ij})$$

Correspondingly, we define the expected probability of progression (that is, expected for a patient with the same clinical risk factors with a nephrologist of median quality) by setting w_j = 1 as:

$$E_{ij} = \exp(X_{ij}\beta) H_0(t_{ij})$$

The risk-standardized ratio (RSR) in a frailty model for provider j will simply be the frailty estimate w_j for provider *j* since

$$RSR_j = \frac{predicted\ number\ of events}{expected\ number\ of\ events} = \frac{\sum_{i=1}^{n_j} P_{ij}}{\sum_{i=1}^{n_j} E_{ij}} = \frac{w_j \sum_{i=1}^{n_j} \exp(X_{ij}\beta) H_0(t_{ij})}{\sum_{i=1}^{n_j} \exp(X_{ij}\beta) H_0(t_{ij})} = w_j$$

Where n_j is the number of patients seeing provider *j*.

The frailty estimate (that is, the ratio of predicted to expected progression hazard) is distributed according to a lognormal distribution, $\log(w_j) \sim N(0,\theta)$, where median($w_j$) = 1. The 95% confidence interval for RSR_j (the frailty) will be a direct output from estimation software. For a given nephrology practice, an RSR of precisely 1 indicates median performance; an RSR greater than 1 indicates a higher risk of progression (and therefore worse performance) than expected while an RSR less than 1 indicates lower risk (better performance).

There is also an option to convert the measure score from a ratio to a rate. A Risk Standardized Incidence Rate (RSIR) may be calculated as $RSIR_j = RSR_j * IR$, where the constant IR is the national incidence rate per 100 patient-years (calculated as 100 times the total number of progression events divided by the total patient-years). A nephrology practice with an RSR of 1 (median performance) would have an RSIR equal to IR (the overall national rate); similarly an RSIR

greater than IR indicates a higher rate of progression (worse performance) and RSIR less than IR indicates a lower rate of progression (better performance). As RSR and RSIR are directly proportional, the choice of score does not affect providers' relative performance; only the interpretation of the numeric scores is affected. We have elected to report the RSR (ratio) here as the centering of the distribution around 1 more clearly highlights providers' relative performance.

sp.25) Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

☐ Copy of instrument is attached.

☒ x Copy of instrument is NOT attached (please explain). Not applicable; measure does not use an instrument.

sp.26) Indicate the responder for your instrument.

☐ Patient

☐ Family or other caregiver

☐ Clinician

☒ x Other (specify) Not applicable; measure does not use an instrument.

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

Examples of samples used for testing:
• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
• The sample should represent the variety of entities whose performance will be measured. The samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.
• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.
• When possible, units of measurement and patients within units should be randomly selected.

Not applicable. The measure is not based on a sample or survey.

sp.28) Identify whether and how proxy responses are allowed.

Not applicable. The measure is not based on a sample or survey.

sp.29) Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

Not applicable. The measure is not based on a sample or survey.

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

☐ Assessment Data
☒ x Claims
☐ Electronic Health Data
☐ Electronic Health Records
☐ Instrument-Based Data
☐ Management Data
☒ x Other (please specify here: Beneficiary Enrollment data including the hospice enrollment, ESRD or dialysis enrollment)
☐ Paper Medical Records
☐ Registry Data

sp.31) Identify the specific data source or data collection instrument.
For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

Data sources for the measure:

- Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician claims for the 12 months prior to eligibility for the cohort.
- Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare enrollment in ESRD and ESRD-Dialysis, hospice, and vital status. These data have previously been shown to accurately reflect patient vital status[1].

Reference:

1. Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. Medical Care. 1992; 30(5): 377-91.

sp.32) Provide the data collection instrument.

☐ Available at measure-specific web page URL identified in sp.09
☐ Available in attached appendix in Question 1 of the Additional Section
☒ x No data collection instrument provided

Importance to Measure and Report: Maintenance of Endorsement (1ma.01)

Not applicable; this is a new measure.

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

☐ *Yes*

☐ *No*

Importance to Measure and Report: Evidence (Complete for Outcome Measures) (1a.01 - 1a.03)

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

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

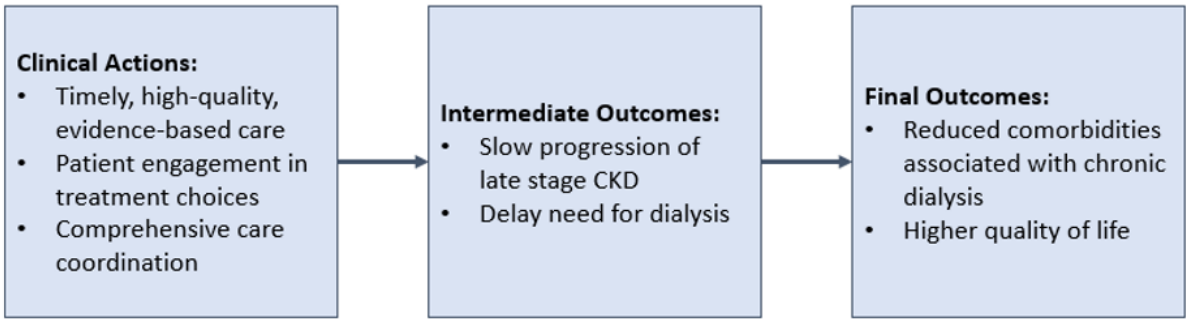
1a.01) Provide a logic model.

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

This is a patient-centered measure that will incentivize nephrology practices and other related entities to improve the care of patients with Stage 4 CKD (who are not yet dependent on dialysis) to slow the progression of their disease. This measure also seeks to encourage stronger communication between clinical entities treating patients with CKD.

This logic model describes how services from nephrology providers can achieve a delay in initiation of dialysis (which tends to provide improved patient-centered care and quality of life) and reduction of comorbidities associated with dialysis. As evidenced in 1a.03 below, services nephrology practices provide (including delivery of timely high-quality evidence-based clinical interventions, fostering patient engagement in treatment choices, and comprehensive care coordination for patients with late-stage CKD) can slow the progression of late-stage CKD and delay the need for dialysis. As evidenced in 1a.02 below, this delayed dialysis results in improved patient-centered care and greater quality of life for patients with CKD while mitigating undesirable outcomes associated with chronic dialysis.

Figure 1a.01. Logic Model



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.

While dialysis can be life-prolonging for patients with renal failure, the decision to begin dialysis is not benign and delaying dialysis initiation can be beneficial for patients. Several studies show that an early start to dialysis does not lead to better patient outcomes in terms of mortality or quality of life [1-6]. In fact, there is some data to suggest that early initiation of dialysis may also be harmful [1,7,8]. Beginning dialysis is associated with a high-risk burden to patients, including risk of infection, pain from dialysis procedures, and high psychosocial impact [2,5,9-11]. Adults suffer from a loss of executive cognitive function once dialysis is initiated, possibly limiting their abilities to make decisions and maintain normal function [12]. On the other hand, delayed dialysis has been associated with positive outcomes including longer survival and greater quality of life [8,13,14], particularly among older patients with other comorbidities for whom conservative (non-dialysis) CKD care may better align with their priorities [8,15].

In addition to the health benefits of delayed dialysis initiation, there may be substantial associated cost savings for patients; a randomized controlled trial found that a 6-month delay in dialysis resulted in savings of $18,000 (otherwise spent on the dialysis treatment plus transportation and hospitalizations) with no difference in quality of life or survival [16].

As part of measure development, we convened a technical expert panel (TEP) comprising a variety of experts and stakeholders; the roster of the TEP is reported in the Appendix (A2). This TEP included four individuals in the capacity of patients and patient advocates, to gain feedback throughout development from a patient-centered perspective. These individuals expressed support for the measure construct, noting that even for patients with more advanced stages of CKD, slowing progression is important and valuable. They stated that a measure like this would promote optimal care processes for these patients, for example through improved care coordination and interventions.

References:

1. Ku E, McCulloch CE, Johansen KL. Starting Renal Replacement Therapy: Is It About Time? *American Journal of Nephrology.* 2019;50(2):144-151.
2. Chen T, Lee VW, Harris DC. When to initiate dialysis for end-stage kidney disease: evidence and challenges. *The Medical Journal of Australia.* 2018;209(6):275-279.
3. O'Hare AM, Wong SP, Yu MK, et al. Trends in the Timing and Clinical Context of Maintenance Dialysis Initiation. *Journal of the American Society of Nephrology.* 2015;26(8):1975-1981.
4. Crews DC, Scialla JJ, Boulware LE, et al. Comparative Effectiveness of Early Versus Conventional Timing of Dialysis Initiation in Advanced CKD. 2014;63(5):806-815.
5. Rosansky SJ, Cancarini G, Clark WF, et al. Dialysis Initiation: What's the Rush? *Seminars in Dialysis.* 2013;26(6):650-657.
6. Cooper BA, Branley P, Bulfone L, et al. A Randomized, Controlled Trial of Early versus Late Initiation of Dialysis. *New England Journal of Medicine.* 2010;363(7):609-619.
7. Susantitaphong P, Altamimi S, Ashkar M, et al. GFR at Initiation of Dialysis and Mortality in CKD: A Meta-analysis. *American Journal of Kidney Diseases.* 2012;59(6):829-840.
8. Rosansky SJ, Eggers P, Jackson K, Glassock R, Clark WF. Early Start of Hemodialysis May Be Harmful. *Archives of Internal Medicine.* 2011;171(5).
9. Dąbrowska-Bender M, Dykowska G, Żuk W, Milewska M, Staniszewska A. The impact on quality of life of dialysis patients with renal insufficiency. *Patient Preference and Adherence.* 2018;Volume 12:577-583.
10. Miskulin DC, Abebe KZ, Chapman AB, et al. Health-Related Quality of Life in Patients With Autosomal Dominant Polycystic Kidney Disease and CKD Stages 1-4: A Cross-sectional Study. *American Journal of Kidney Diseases.* 2014;63(2):214-226.
11. Davison SN. End-of-Life Care Preferences and Needs: Perceptions of Patients with Chronic Kidney Disease. *Clinical Journal of the American Society of Nephrology.* 2010;5(2):195-204.
12. Kurella Tamura M, Vittinghoff E, Hsu CY, et al. Loss of executive function after dialysis initiation in adults with chronic kidney disease. *Kidney Int.* 2017;91(4):948-953.
13. Trivedi HS, Pang MM, Campbell A, Saab P. Slowing the progression of chronic renal failure: economic benefits and patients' perspectives. Am J Kidney Dis. 2002;39(4):721-729.7
14. Wright S, Klausner D, Baird B, et al. Timing of dialysis initiation and survival in ESRD. Clin J Am Soc Nephrol. 2010;5(10):1828-1835.
15. Da Silva-Gane M, Wellsted D, Greenshields H, Norton S, Chandna SM, Farrington K. Quality of life and survival in patients with advanced kidney failure managed conservatively or by dialysis. *Clin J Am Soc Nephrol.* 2012;7(12):2002-2009.
16. Harris A, Cooper BA, Li JJ, et al. Cost-Effectiveness of Initiating Dialysis Early: A Randomized Controlled Trial. *American Journal of Kidney Diseases.* 2011;57(5):707-715.

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

Multiple studies have described interventions that successfully slowed progression of late-stage CKD, contrary to the historical view that progression to ESRD was inevitable and rapid below a certain level of kidney function; while historically patients with advanced CKD were excluded from many studies or trials, more recent research that does include these patients presents promising results [1]. One randomized trial found administration of angiotensin-converting enzyme (ACE) inhibitors to patients with advanced renal insufficiency resulted in a reduced risk of primary outcome (defined as doubling of serum creatinine, progression to ESRD, or death) as well as several secondary outcomes (such as proteinuria levels and rate of renal function decline) [2]. A retrospective cohort study found that the renoprotective effect of ACE inhibition is independent of renal failure severity and reduced the risk of ESRD for patients with severe renal dysfunction [3]. Another study found that stabilization of GFR and delay of progression events through erythropoietin therapy to manage anemia among stage 4 and 5 CKD patients is achievable in a CKD clinic setting [4].

Other studies have demonstrated that coordinated and comprehensive care in CKD clinics support greater patient engagement in treatment choices [5,6].

References:

1. Batlle D, Ramadugu P, Soler M.J. Progress in retarding the progression of advanced chronic kidney disease: Grounds for optimism. *International Society of Nephrology*. 2006;70(104):S40–S44.
2. Hou FF, Zhang X, Zhang GH et al. Efficacy and safety of benazepril for advanced chronic renal insufficiency. *N Engl J Med.* 2006; 354: 131–140.
3. Ruggenenti P, Perna A, Remuzzi G. ACE inhibitors to prevent end-stage renal disease: when to start and why possibly never to stop: a post hoc analysis of the REIN trial results. Ramipril Efficacy in Nephropathy. *J Am Soc Nephrol.* 2001; 12: 2832–2837.
4. Serrano AHJ, Ghossein C, Nishi L et al. Stabilization of GFR in advanced chronic kidney disease: A two-year follow up of a cohort of CKD patients stages 4 and 5. *Adv Chronic kidney Dis*. 2006 in press.
5. Levin A, Lewis M, Mortiboy P, et al. Multidisciplinary Predialysis Programs: Quantification And Limitations Of Their Impact On Patient Outcomes In Two Canadian Settings. *Am J Kidney Dis.* 1997; 29: 533–540.
6. Singh AK. The Emergence of the CKD Clinic: Modeling Practices for Excellence CME. avaible at: http://medscape.com/viewprogram/193

Importance to Measure and Report: Evidence (Complete for Process Measures) (1a.03 - 1a.16)

Not applicable; this is an outcome measure.

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

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01) Provide a logic model.

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

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other (please specify here:)

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, you may add additional tables to the relevant sections. Please follow the 508 Checklist for tables.

Evidence - Systematic Reviews Table (Repeatable)

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

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

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

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

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

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

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

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

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

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

Evidence

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

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

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

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

Importance to Measure and Report: Gap in Care/Disparities (1b.01 - 1b.05)

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.

This measure will directly benefit patients by assessing the quality of care delivered by nephrology-related entities to Medicare beneficiaries with CKD and highlighting opportunities for nephrology practices to improve care.

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.

Table 1b.02 below summarizes the distribution of performance results (RSR) for all providers and separately among those caring for at least 25 patients. As discussed in Section sp.24 (calculation of measure score), RSR is a ratio measure with a score of 1 indicating median performance for a given case mix, a score less than 1 indicating lower chance of progression (better performance) than expected, and a score greater than 1 indicating greater chance of progression than expected.

Summary statistics include the mean (with standard deviation), median (with interquartile range), and the range (minimum and maximum values). Full histograms for RSR distribution are shown in Figures 1b.01:1-2 for all practices and those with at least 25 patients. As shown by these distributions, there was substantial variation in performance between measured entities after accounting for clinical risk. The range of 0.604-1.629 (a 2.7-fold increase in progression hazard between the best- and worst-quality nephrologists after accounting for case mix) indicates a large gap that can be explained by a meaningful difference in performance. There is also a fairly broad distribution between those extreme values as illustrated in the histogram. At top-performing practices, there are fewer progression events on a risk-adjusted basis. This shows that not all entities will score the same and that entities can be distinguished from each other in terms of quality.

Table 1b.02. Measure Performance (RSR) Statistics for All Providers and Providers with 25 or More Patients, Progression Development Dataset

Statistics	All Practices (N=2,854)	Practices with 25 + Patients (N=1,970)
RSR: Mean (SD)	1.007 (0.117)	1.007 (0.136)
RSR: Median (IQR)	0.996 (0.948 – 1.056)	0.993 (0.922, 1.083)
RSR: Range (min-max)	0.604 – 1.629	0.604 - 1.629

Figure 1b.02:1. Distribution of RSR, all nephrology practices (n=2,854)

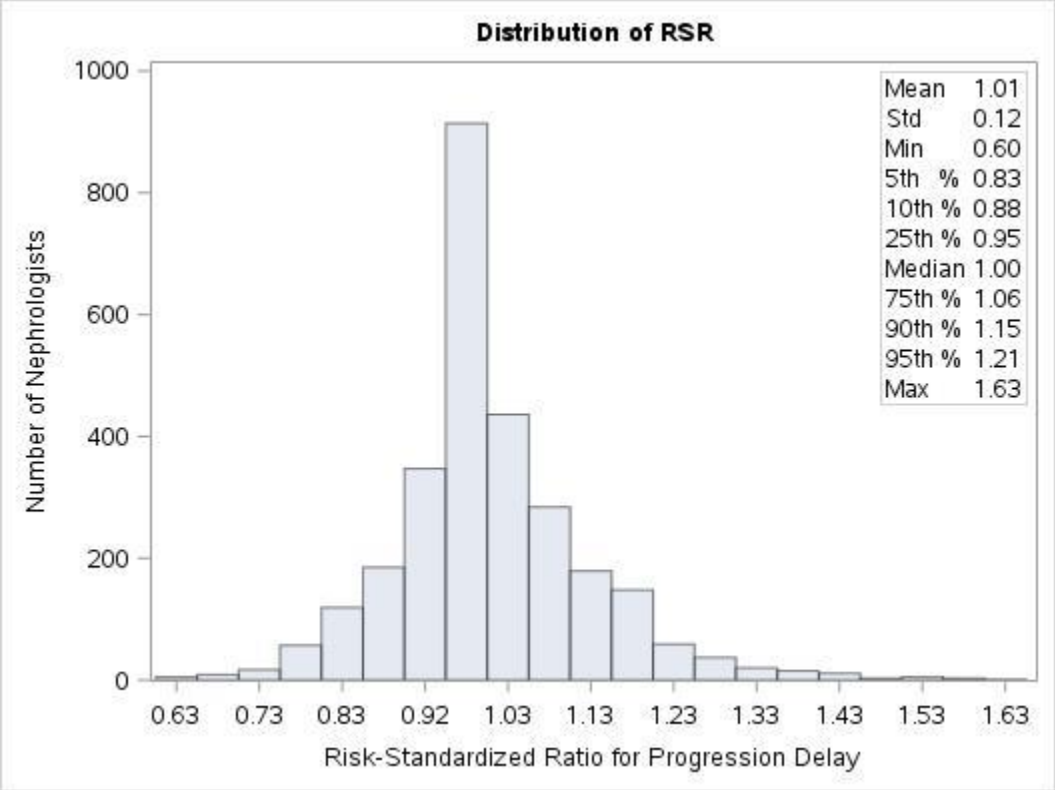
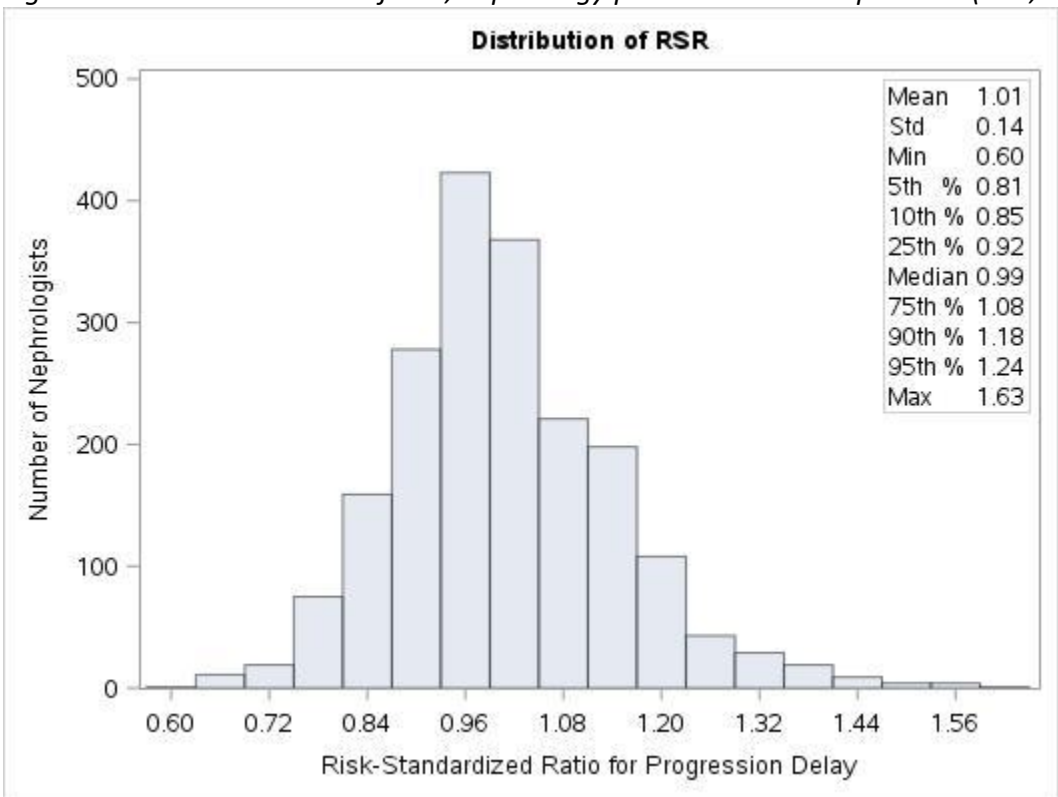


Figure 1b.02:2. Distribution of RSR, nephrology practices with 25+ patients (n=1,970)



As the measure has not yet been fully implemented, there is no data yet available on improvement.

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.

Not applicable; data are reported above.

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.

For more discussion of empiric disparity results, see Section 2b.25.

Tables 1b.04:1-4 below include the distributions of clinical risk-adjusted measure score at the provider level, among quintiles of providers based on the prevalence of patients with each social risk factor in the Progression Development Dataset. The distribution is fairly consistent across each quintile, illustrating that after accounting for differences in clinical case mix, the risk of progression from stage 4 CKD to stage 5 or to ESRD does not depend substantially on the proportion of patients served who are Black, dual-eligible, low-SES, or urban residents. Notably, the variation in outcomes within each quintile is much greater than any variation between quintiles.

Table 1b.04:1. RSR distribution by quintiles of % patients of Black race among nephrology practices with 25+ patients

Quintile	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
% patients of Black race	0.0 - 0.0	0.2 - 5.0	5.1 - 12.7	12.7 - 28.0	28.0 - 100.0
N (practices)	397	391	394	394	394
Mean (SD)	0.99 (0.12)	0.99 (0.14)	1.02 (0.14)	1.01 (0.14)	1.02 (0.14)
Minimum	0.64	0.60	0.67	0.71	0.64
10th percentile	0.85	0.83	0.86	0.84	0.85
Q1	0.92	0.91	0.92	0.92	0.93
Median	0.97	0.99	1.01	0.99	1.01
Q3	1.06	1.07	1.11	1.09	1.09
90th percentile	1.16	1.16	1.19	1.18	1.18
Maximum	1.56	1.63	1.55	1.52	1.52

Table 1b.04:2. RSR distribution by quintiles of % dual eligible patients among nephrology practices with 25+ patients

Quintile	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
% dual eligible patients	0.0 - 6.3	6.3 - 11.7	11.8 - 17.5	17.5 - 30.0	30.1 - 100.0
N (practices)	395	391	397	393	394
Mean (SD)	1.02 (0.14)	1.00 (0.14)	1.00 (0.14)	1.00 (0.13)	1.01 (0.12)
Minimum	0.64	0.64	0.67	0.60	0.66
10th percentile	0.87	0.82	0.84	0.85	0.87
Q1	0.93	0.91	0.91	0.92	0.93
Median	0.99	0.99	0.98	0.99	1.00
Q3	1.10	1.09	1.08	1.08	1.08
90th percentile	1.20	1.18	1.17	1.16	1.17
Maximum	1.52	1.56	1.63	1.51	1.57

Table 1b.04:3. RSR distribution by quintiles of % patients from low SES neighborhood (bottom quintile AHRQ SES score) among nephrology practices with 25+ patients

Quintile	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
% low SES patients	0.0 - 2.4	2.4 - 12.2	12.3 - 25.0	25.1 - 43.5	43.6 - 100.0
N (practices)	394	393	395	394	394
Mean (SD)	1.01 (0.12)	1.00 (0.14)	1.01 (0.14)	1.00 (0.14)	1.01 (0.13)
Minimum	0.67	0.64	0.69	0.60	0.66
10th percentile	0.88	0.83	0.85	0.84	0.86
Q1	0.93	0.90	0.92	0.91	0.93
Median	0.99	0.98	0.99	0.99	1.00
Q3	1.08	1.08	1.08	1.08	1.08
90th percentile	1.17	1.18	1.20	1.17	1.19
Maximum	1.57	1.55	1.63	1.52	1.49

Table 1b.04:4. RSR distribution by quintiles of % urban patients among nephrology practices with 25+ patients (note that top quintile with 100% urban patient population includes more than 20% of providers)

Quintile	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
% urban patients	0.0 - 65.0	65.0 - 92.1	92.1 - 99.3	99.3 - 99.9	100.0 - 100.0
N (practices)	394	394	395	370	417
Mean (SD)	1.00 (0.14)	1.00 (0.14)	1.00 (0.14)	1.02 (0.15)	1.01 (0.12)
Minimum	0.66	0.60	0.64	0.69	0.66
10th percentile	0.83	0.84	0.83	0.85	0.89
Q1	0.91	0.92	0.92	0.91	0.94
Median	0.99	0.99	1.00	1.00	0.99
Q3	1.08	1.09	1.08	1.10	1.07
90th percentile	1.16	1.18	1.18	1.19	1.16
Maximum	1.51	1.52	1.52	1.63	1.53

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.

Not applicable; data are reported above.

Scientific Acceptability: Maintenance (2ma.01 - 2ma.04)

Not applicable; this is a new measure.

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

☐ *Yes*
☐ *No*

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

☐ *Yes*
☐ *No*

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

☐ *Yes*
☐ *No*

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

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

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

☐ *Yes - Additional risk adjustment analysis is included*

☐ *No additional risk adjustment analysis included*

Scientific Acceptability: Reliability - Testing (2a.01 - 2a.12)

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

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

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

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

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

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

AND

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

2b3. For outcome measures and other measures when indicated (e.g., resource use):

• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

• rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

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

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

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

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

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

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

Definitions

Reliability testing applies to both the data elements and computed measure score.

Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

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

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

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

☐ Assessment Data
☒ x Claims
☐ Electronic Health Data
☐ Electronic Health Records
☐ Instrument-Based Data
☐ Management Data
☒ x Other (please specify here: Beneficiary Enrollment data including hospice enrollment, ESRD/dialysis enrollment)
☐ Paper Medical Records
☐ Registry Data

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).

Multiple datasets were used for these analyses. See section 2a.07 for additional details.

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

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

Dates of data vary by dataset. See section 2a.07 for additional details.

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.

☐ Accountable Care Organization
☒ x Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☐ Integrated Delivery System
☐ Other (specify)
☐ Population: Community, County or City
☐ Population: Regional and State

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.

In testing, measured entities are any clinician groups billing for nephrology services to Medicare FFS patients 18 years or older, grouped by taxpayer identification number (TIN), and identified through Medicare FFS administrative claims data. There were 2,854 measured entities.

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.

The number of patients varied by dataset; see 2a.07 for details.

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.

Table 2a.07:1. Dataset Descriptions

Dataset	Applicable Section	Description of Dataset
Progression Development Dataset (Medicare Fee-For-Service Administrative Claims & Enrollment Data)	Section 2a.09 Reliability Testing Section 2b.01 Validity Testing Section 2b.05 Meaningful Differences Section 2b.30 Risk Adjustment/Stratification Section 2b.16 Testing of Measure Exclusions 2b.20 Statistical Risk Model Discrimination Statistics 2b.28 Statistical Risk Model Calibration Statistics	Dates of data: January 1, 2017 – December 31, 2018 Number of patients in the dataset: 434,764 See Table 2a.07:2 below for patient characteristics. Number of measured entities (nephrology practices): 2,854 Data obtained through CMS Chronic Conditions Data Warehouse and Virtual Research Data Center (CCW/VRDC) and the CMS integrated data repository (IDR).
Progression EHR Dataset A	Section 2b.01 Validity Testing	Dates of data: 2013-2019 Number of patients in the data set: 7,599 Number of patient visits: Data source: Single health system

Dataset	Applicable Section	Description of Dataset
Progression EHR Dataset B	Section 2b.01 Validity Testing	Dates of data: 2018-2021 Number of patients in the data set: 10,198 Number of patient visits: 14,070 Data source: Nation-wide non-profit healthcare system
Master Beneficiary Summary File (MBSF)	Section 2b.30: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: July 2016 – June 2019 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the measure outcome and dual-eligible status.
The American Community Survey (ACS)	Section 2b.30: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: 2013-2017 We used the AHRQ SES index score derived from the American Community Survey (2013-2017) to study the association between the outcome and social risk factors. The AHRQ SES index score is based on patient 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
US Department of Agriculture Economic Research Service: 2013 Rural-Urban Continuum Codes	Section 2b.30: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: Collected in 2010 census and 2006-2010 American Community Survey; used for urban/rural social risk factor analysis

Table 2a.07:2. Patient Characteristics, Progression Development Dataset (N=434,764)

Characteristic	Number of Patients	Percentage of Patients
Total Patients	434,764	100.00
Age in year prior to measure year (2017)	*	*
Mean (SD)	75.9	10.2
Minimum, Maximum	18	109
Q2 (IQR)	76	13
Gender	*	*
Male	208,029	47.85
Female	226,735	52.15
Race	*	*
Black	57,179	13.15
Non-Black	374,384	86.85

Characteristic	Number of Patients	Percentage of Patients
Dual eligible in 2018	*	*
No	362,861	83.46
Yes	71,903	16.54

*Intentionally left blank

Progression EHR Dataset A and Progression EHR Dataset B:

Progression EHR Dataset A was derived from a single health system and included patients with any outpatient encounter from 2013-2019 with Stage 4 or 5 CKD diagnosis code or eGFR lab value under 30. Progression EHR Dataset B was derived from a large multihospital system and included patients with any outpatient visit from July 2018-December 2021 with Stage 4 CKD diagnosis or eGFR between 15-29. Data for both EHR datasets included deidentified, retrospective demographics (age, sex, gender, race), creatinine and eGFR values, and claims history (comorbidities). Minor data cleaning was applied, including: encounters on the same day were combined as one; patients who only had one encounter were removed; and for encounters where patients only had creatinine, eGFR was calculated using the CKD-EPI 2009 equation[1].

Reference:

1. Equation can be accessed at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2763564/

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.

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

We selected SES variables to analyze after reviewing the literature, developing our conceptual model, and examining available national data sources. The causal pathways for SES variable selection are described below in Section 2b.23.The SES variables used for analysis were:

- **Dual eligible status**: Dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data is obtained from the CMS Master Beneficiary Summary File (MBSF).
    - Following guidance from ASPE[1,2], NQF[3], and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable. We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous outcome.

- **AHRQ-validated SES index score** (range from 0 indicating lowest SES to 100 indicating highest SES) summarizing the information from the following seven variables): percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room.
    - We analyzed the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas, namely ZIP codes[4]. We analyzed the lowest quartile of the AHRQ SES Index versus the other three quartiles. In our data, the lowest quartile equated to an AHRQ SES index score equal to or below 43.
- **Race (Black compared to non-Black)**. Data source: Medicare enrollment database.
    - We used the Medicare enrollment database to identify the patient-level race variable (Black) that we used in these analyses. The Black variable has been shown to be reliable for use in this dataset[5].
- **Urbanicity.** Data Source: US Department of Agriculture Economic Research Service: 2013 Rural-Urban Continuum Codes
    - The Rural-Urban Continuum Codes assign each county in the US to one of nine subgroups. They distinguish metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. The most recent Rural-Urban Continuum Codes are based on data from the 2010 decennial census and the 2006-10 American Community Survey[6].

References:

1. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2019.
2. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/reports/second-report-congress-social-risk-medicares-value-based-purchasing-programs Accessed July 2, 2020.
3. National Quality Forum. Driving Measurable Health Improvements Together Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk within Healthcare Performance Measurement August 2021. Accessed at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=96087
4. Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;2.
5. Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. Health Care Financing Review. 2004;26(2)

6. USDA ERS. Rural-Urban Continuum Codes. 2020; https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx. Accessed December 29, 2022

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

Choose one or both levels.

☐ Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
☒ x Accountable Entity Level (e.g., signal-to-noise analysis)

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.

We provide the signal-to-noise reliability statistic among all providers and those with 25 or more cases during the measurement year, showing the mean, standard deviation, and median, quartiles, minimum and maximum.

Measures reported publicly or used in a payment system typically include a minimum volume or case count to ensure reported results are reliable. The precise threshold is typically determined in implementation and is not prescribed here as a measure specification. We have reported select results only among providers with at least 25 eligible cases (a common threshold among other risk-adjusted claims-based measures in CMS programs) as an example value, to ensure results are not distorted by very low-volume outliers. However, we note that these results are still based on scores estimated among all eligible practices with at least one case, and accordingly is free to be re-set in implementation if appropriate with no impact to the measure methodology or numerator/denominator calculations.

We used the formula for signal-to-noise reliability presented by Adams et al. to calculate individual clinician-level and TIN-level reliability scores[1]. To estimate the overall signal and noise, we first calculated the ICC for the Model Participant, j, using the estimates of between-entity variance τ2 and the formula for intraclass correlation coefficient (ICC) presented by Shrout and Fleiss[2]. Specifically, the signal-to-noise reliability score for Model Participant, j, Rj is calculated as:

$$R_j = \frac{n_j ICC}{1 + (n_j - 1)ICC}$$

while

$$ICC = \frac{\tau^2}{\tau^2 + \pi^2/6\gamma^2}$$

nj is the number of patients for the nephrologist j, τ2 is the between agency variance in a Weibull model with lognormal frailty that used to approximate the Cox model with lognormal

frailty specified above and represent the signal, and $\pi 2/6\gamma 2$ represents the noise and γ is the shape parameter of the Weibull distribution.

Rj ranges from 0 to 1.0. The higher the score, the higher the reliability. Also, we can see that the reliability of agency measure score will vary depending on the number of patient encounters. Entities with higher volume will tend to have more reliable scores, while those with lower volume will tend to have less reliable scores.

References:

1. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.
2. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin. 1979;86(2):420. 35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977:159-174.

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, Measure Evaluation Criteria).

Table 2a.11. Signal-To-Noise Reliability Statistics Among All Nephrology Practices and Those With At Least 25 Cases, Progression Development Dataset

Description	Number of Providers	Mean (SD)	Median (IQR)	Minimum – Maximum
Among All Nephrologists	2,854	0.614 (0.293)	0.696 (0.375 – 0.872)	0.036 – 0.991
Among Nephrologists with at least 25 cases	1,970	0.787 (0.141)	0.821 (0.681 -0.909)	0.484 – 0.991

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?)

As shown in Table 2a.11, among all nephrology practices (including those with small case counts), the median signal-to-noise reliability was 0.696, indicating at least half of providers have reliability above 0.7. Among those with at least 25 cases, the median signal-to-noise reliability was higher at 0.821, indicating that more than half have at least reliability of 0.8. These results demonstrate reliability sufficient for a publicly reported quality measure.

Scientific Acceptability: Validity - Testing (2b.01 - 2b.04)

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

☒ x Patient or Encounter-Level (data element validity must address ALL critical data elements)
☐ Accountable Entity Level (e.g., hospitals, clinicians)
☒ x Empirical validity testing of the measure score
☒ x Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

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.

Data Element Validity

We validated the accuracy of those patients with Stage 4 CKD (cohort) and the outcome of ESRD requiring chronic dialysis. All critical data elements were validated.

Data element validation was completed for the variable ICD-10 code N18.4 twice, in two different datasets, to ensure validity. The two datasets used were the **Progression EHR Dataset A and Progression EHR Dataset B**, further explained in Section 2a.07 above. To establish data element validity, we sought to determine the percent agreement between patients with at least one outpatient encounter in a calendar year with a diagnosis code for Stage 4 CKD (N18.4) and the presence of a confirmatory lab value. Clinically, CKD is defined by the Estimated Glomerular Filtration Rate (eGFR) kidney function biomarker, which measures how efficiently the kidneys filter waste from blood. Stage 4 CKD is defined as eGFR between 15-29 mL per minute per 1.72 square meters. For encounters that did not have an eGFR in the EHR, a creatinine result was converted into an eGFR.

Agreement was defined as: beginning with an outpatient encounter with Stage 4 CKD (ICD-10 code N18.4) (the definition used in the measure to define the denominator), there was either a) within the same encounter a lab value for eGFR between 15-19; or b) an encounter within 180 days prior (or 30-days forward) with a lab value for eGFR between 15-29.

We additionally assessed the validity of using ESRD or ESRD-Dialysis enrollment as the indicator of outcome (progression to ESRD) by analyzing alignment between ESRD enrollment and documentation of dialysis facility billing codes. Among beneficiaries in 2018, we compared the overlap in ESRD enrollment (which requires completion of CMS Form 2728 documenting evidence of ESRD and start date) to 1) occurrence of dialysis facility claim (Type of Bill [TOB] 72)

and 2) TOB 72 claim plus Monthly Capitation Payment (MCP) HCPCS billing codes for ESRD (table 2b.02).

Table 2b.02. HCPCS MCP codes for ESRD and related services

HCPCS MCP code	Service
90957-90959	ESRD related services monthly, for patients 12-19
90960-90962	ESRD related services monthly, for patients 20 years of age and older
90965, 90966	End-Stage Renal Disease Services

Empiric Validity –Model Validation

As reported in 2b.32, we further compared the claims-only risk model to a model additionally including eGFR data abstracted from EHRs using Progression EHR Datasets A and B. eGFR is a clinical "gold standard" for CKD staging and should be a relevant predictor of risk. In terms of validity, the purpose of this testing was to evaluate statistical performance for the claims-only model compared to a model including that gold standard.

Face Validity

A Technical Expert Panel (TEP) of 15 members comprised of diverse stakeholders completed a survey assessing the face validity of the measure after three rigorous meeting sessions in 2020, 2021, and 2022. Each participant was integrated into the measurement development process and participated in each TEP meeting or submitted detailed feedback post-meeting. Participants were asked the following question: "Do you believe this measure (claims-based) can be used to distinguish provider quality among nephrologists caring for patients with stage 4 CKD?" Responses were limited to one of the following: strongly agree, somewhat agree, somewhat disagree, strongly disagree. For the composition and affiliation of the TEP members, please refer to Additional Item 2.

Reference:

1. Equation can be accessed at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2763564/

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

Examples may include correlations or t-test results.

Data Element Validity Results

The observed match rate was high among non-missing eGFRs in both datasets; a majority of patients with a Stage 4 CKD diagnosis had documented lab values supporting that diagnosis:

- Progression EHR Dataset A: The match rate for beneficiary visits with a Stage 4 CKD diagnosis and a confirmatory eGFR laboratory value was 88.1%.
- Progression EHR Dataset B: The match rate for patient visits with a Stage 4 CKD diagnosis and a confirmatory eGFR laboratory value the was 83.5%.

In 2018, we found 367,637 total patients enrolled in Medicare ESRD. We found 312,324 total patients with TOB72 claims, of whom 310,264 were also enrolled in ESRD. Examining the

overlap shows that 57,373 patients (15.6%) with ESRD enrollment did not have TOB72, while only 2,060 (0.67%) of patients with TOB72 claims were not enrolled in ESRD.

Among patients with TOB72 claims we found that 302,740 also had MCP ESRD service codes, of whom 301,298 were also ESRD-enrolled. Examining the overlap shows that 66,339 patients (18.0%) with ESRD enrollment did not have TOB72 ESRD claims, while only 1,442 (0.48%) of patients with TOB72 ESRD claims were not enrolled in ESRD.

Empiric Validity –Model Validation

Results of testing model performance with claims-only risk factors vs. including lab eGFR values are reported in 2b.32, Table 2b.32 and figures 2b.32:1-4.

Face Validity

73% of TEP members agreed somewhat or strongly that the Progression of CKD Measure can be used to distinguish provider quality (11 out of 15).

The breakdown of responses is shown in Table 2b.03:2 below:

Table 2b.03:2. TEP Responses

Response	Number of Responses
Strongly Agree	5
Somewhat Agree	6
Somewhat Disagree	3
Strongly Disagree	1

A majority of panelists agreed that the measure is a valid measure of provider quality. These panelists noted that currently there are no incentives in place focusing on progression in kidney disease and this measure would provide an incentive for nephrologists to improve care with respect to preventing the progression of disease. Panelists also noted that the measure could help to improve standardization of processes related to kidney care. One TEP member emphasized that the training and validation data sets provide high confidence for the ability of the measure to distinguish provider quality.

A few TEP members disagreed (three "somewhat," one "strongly") that the measure is able to differentiate provider quality. Two of these individuals were concerned about the absence of eGFR data (the clinical gold standard to define stages of CKD and progression to ESRD) in the measure, as claims are potentially less granular and less consistently coded. Two other individuals cited patient-level factors outside of providers' control and potential unintended consequences that may result.

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?)

Data Element Validity

The results of our analyses showed a strong match rate between instances of Stage 4 CKD from claims with confirmatory eGFR laboratory values. Therefore, we conclude that using Stage 4 CKD claims to define the denominator achieves the intended measure cohort.

Because ESRD enrollment requires documented evidence of ESRD (submitted in Form 2728), and ESRD enrollment is required in order to bill Medicare for regular dialysis treatment, the face validity of ESRD enrollment as an indicator of progression to ESRD (with a clear start date) is high. This is corroborated by comparison to two other potential claims-based means of identifying ESRD progression; ESRD enrollment includes nearly all beneficiaries with ESRD billing claims. We conclude that ESRD is the most comprehensive and reliable record for progression to ESRD requiring dialysis available using administrative data sources.

A TEP reviewed all results and agreed with the validity of the data elements used in the measure; the affirmative TEP vote for measure face validity is expanded upon directly below.

Empiric Validity –Model Validation

As reported in 2b.32, inclusion of eGFR values as predictors in the risk model modestly improved some performance statistics while offering no clear improvement in others above the claims-only model. In terms of validity, this suggest that while eGFR values are a gold standard for CKD staging and would offer some predictive value, the benefit of this is empirically modest to small and using only the claims-based risk factors satisfactorily captures differences in patients' risk. Particularly when weighed against the additional burden that would be required of providers to report eGFR values and the already strong performance of the claims-only model, the marginal improvement of adding eGFR data to the model is not greatly meaningful.

Face Validity

The vast majority of the TEP members agreed with the measure's ability to differentiate provider quality, offering several points in favor of the measure. Members broadly agreed that claims data provides sufficient information to determine the overall quality of care provided to patients and noted high confidence in the measure's ability to do so given CORE's testing results. Panelists noted that the measure would fill a gap in incentives, as there is currently no measure of nephrologists' quality of care in preventing CKD progression. They noted another potential benefit in the measure improving standardization of processes related to kidney care.

The TEP members who did not agree with the validity of the measure raised two main themes in their rationales: the lack of lab eGFR values in the cohort, outcome, or risk adjustment; and concern about unintended consequences of not adjusting for certain patient-level factors outside of providers' control (such as poor disease self-management).

While CORE acknowledges that including eGFR values may be preferable in an ideal world, there are substantial tradeoffs that would be required to include at the present time (namely the additional provider burden if reporting eGFR lab values was required). Multiple TEP members who did support the measure validity similarly believed that despite the added value of eGFR its inclusion would not be worth the additional burden. The current risk model includes

43 relevant claims-based clinical variables (including age, history of CKD, proteinuria, cardiovascular disease, diabetes, and hypertension among others) and CORE's testing indicate these variables produce a model that sufficiently categorizes patients' risk for the outcome. Concerns about the accuracy of relevant billing codes in claims are also mitigated by the measure's limitation to nephrology providers, who may tend to have more accurate nephrology coding practices than other clinicians.

While there is patient-level variability that the measure does not capture, such r disease self-management, this is true of all measures and is ideally mitigated to some extent by provider care, education and support.

Scientific Acceptability: Validity - Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) (2b.05 - 2b.14)

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.

Examination of provider-level results include measure scores for all nephrologists and those with at least 25 patients, along with statistics summarizing their distribution. We are using 25 as an example minimum case count, which aligns with several CMS publicly reported outcome measures (as discussed in 2a.10).

We also examined the top (worst) quintile of scores and tested for significant difference from the bottom (best) quintile, and from the first four quintiles combined among providers with at least 25 patients.

We assessed the clinical value of difference with input from two nephrologists consulted throughout the development process.

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.

As discussed in Section sp.24 (calculation of measure score), RSR is a ratio measure with a score of 1 indicating median performance for a given case mix, a score less than 1 indicating lower mortality (better performance) than expected, and a score greater than 1 indicating greater mortality than expected.

Table 2b.06:1 below shows the distribution of RSR and volume for all providers and for providers with at least 25 patients. The full distributions of RSR are shown in Figure 2b.06:1 (all providers) and 2b.06:2 (providers with at least 25 patients)

Table 2b.06:1. Measure Performance Statistics (RSR) for All Providers and Providers with 25+ patients, Progression Development Dataset

Statistics	N	RSMR: Mean (SD)	RSMR: Median (IQR)	RSMR: Range (min-max)	Volume: Mean (SD)	Volume: Median (IQR)	Volume: min - max
All practices	2,854	1.007 (0.117)	0.996 (0.948-1.056)	0.604-1.629	152.3 (245.7)	61 (16-182)	1-2,924
Practices with 25+ patients	1,970	1.007 (0.136)	0.993 (0.922-1.083)	0.604-1.629	216.6 (272.3)	122 (57-266)	25-2924

The nephrologists consulted by CORE suggested a difference from mean of 0.5 standard deviations or greater to be one they would find meaningful in practice. In the mortality development dataset among providers with 25+ patients, this would include providers with RSR less than 0.939 or greater than 1.075.

Table 2b.06:2 below shows the distribution of RSR among the fifth/top quintile of scores (worst performance), the first/bottom quintile (best), and the first four quintiles combined among providers with at least 25 patients.

Table 2b.06:2. Measure Performance Statistics (RSR) by quintiles of RSR, Mortality Development Dataset (all among providers with 25+ patients)

Statistics	N	RSMR: Mean	RSMR: SD	RSMR: Range (min-max)	P-value: Difference from 5th quintile
Fifth quintile (worst performance)	394	1.209	0.094	1.114 – 1.629	n/a
First quintile (best)	394	0.834	0.057	0.604 – 0.901	<0.0001
Quintiles 1-4 (not worst)	1576	0.956	0.090	0.604 – 1.114	<0.0001

Figure 2b.06:1. Distribution of RSR, all nephrology practices (n=2,854)

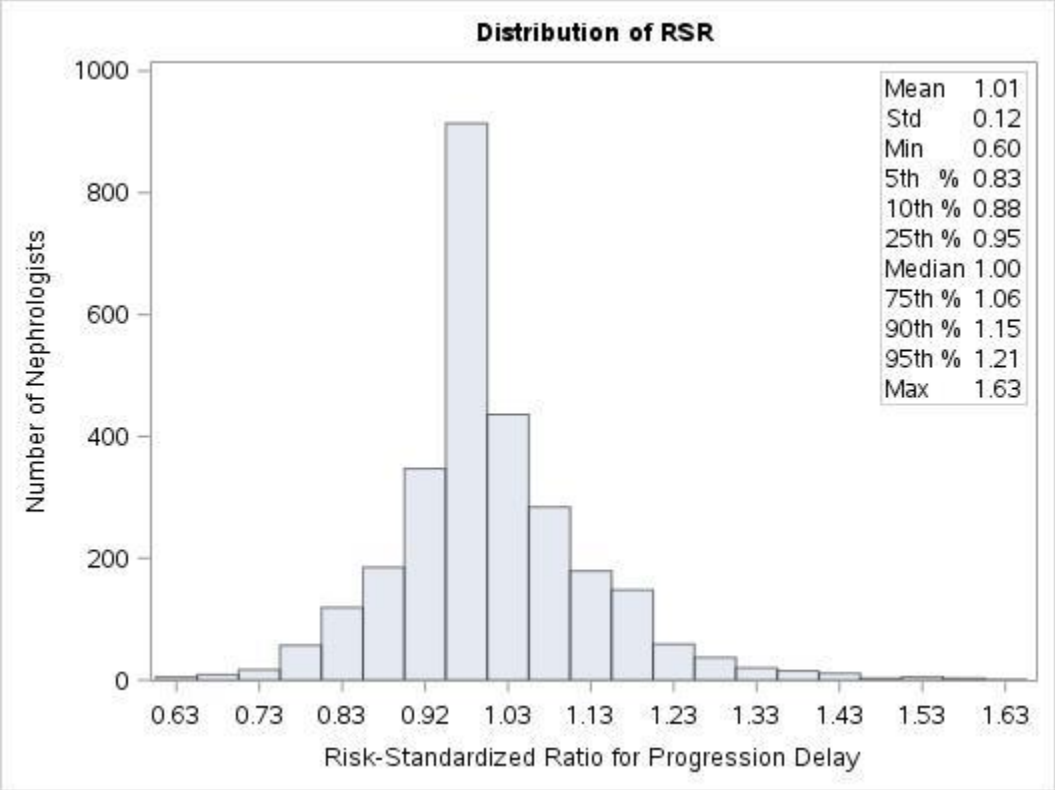
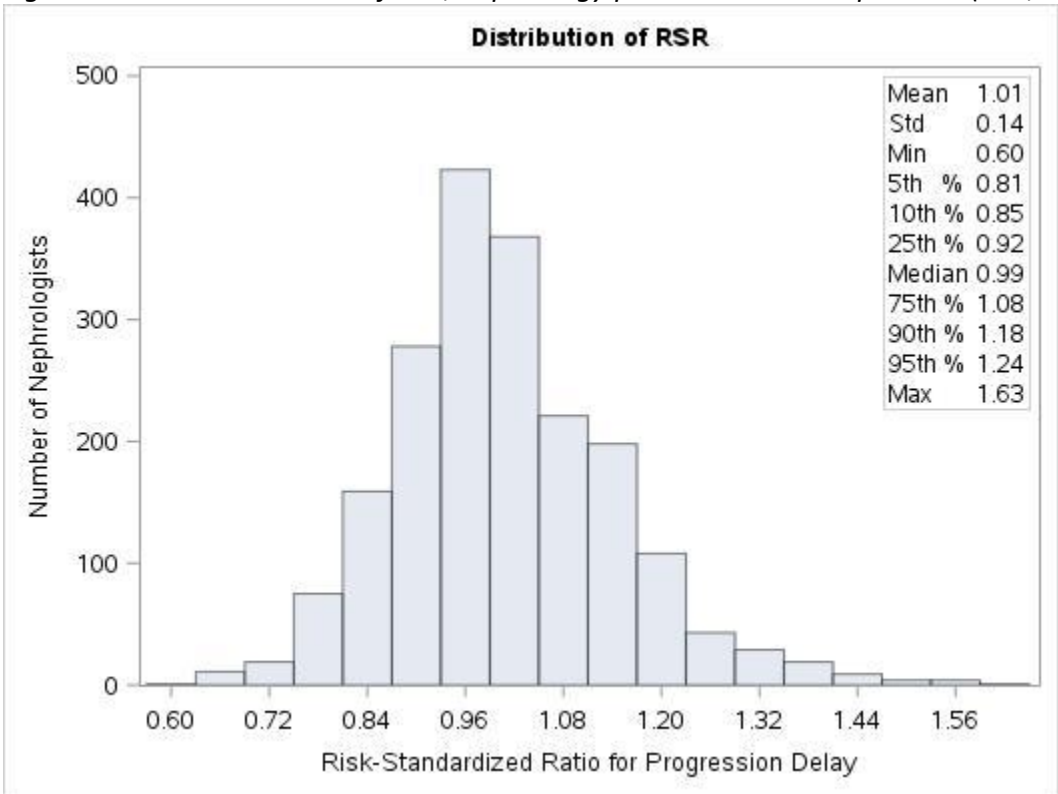


Figure 2b.06:2. Distribution of RSR, nephrology practices with 25+ patients (n=1,970)



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?

As shown by the distributions of the RSR performance score in Section 2b.06, there was substantial variation in performance between measured entities after accounting for clinical risk. The range of 0.604 – 1.629 (a 2.7-fold increase in progression hazard between the best- and worst-quality nephrologists after accounting for case mix) indicates a large gap that can be explained by a meaningful difference in performance.

The significant difference between the means of the bottom (best) and top (worst) quintile providers of 0.834 – 1.209 reflects a still-substantial 1.45-fold increase in hazard attributable to differences in care; this 45% increase in risk was considered meaningful by CORE's consulting nephrologists. Of note, both means fall more than 1.0 standard deviations away from the overall mean of 1.007 (SD 0.136). Even the maximum score in the bottom quintile (0.901) and the minimum score in the top quintile (1.114) surpass the thresholds of 0.5*SD from mean score suggested by the nephrologists, illustrating a great deal of meaningful performance differentiation. This broad distribution is illustrated in the histograms in 2b.06. Scores in the top quintile were also significantly different than those in the other four combined, indicating this measure is capable of differentiating poor performers.

Overall, there are significantly and meaningfully more progression events on a risk-adjusted basis at worst-performing practices. This shows that not all entities will score the same and that entities can be distinguished from each other in terms of quality.

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.

The measure used claims-based data for development and testing. All administratively coded diagnoses were available for measurement. Lack of a claim is treated as not having the corresponding diagnosis or procedure.

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).

The measure used claims-based data for development and testing. All administratively coded diagnoses were available for measurement.

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.

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.

The measure used claims-based data for development and testing. All administratively coded diagnoses were available for measurement. The data on patient deaths were obtained from the Medicare Enrollment Database; these data have previously been shown to accurately reflect patient vital status[1]. As discussed in section 2b (Validity testing) the validation of the claims-based methodology against EHR data demonstrates that the measure satisfactorily identifies the appropriate patients.

Reference:

1. Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. Medical Care. 1992; 30(5): 377-91.

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

☐ Yes, there is more than one set of specifications for this measure
☒ x No, there is only one set of specifications for this measure

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.

Not applicable; there is only one set of specifications for this measure.

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.

Not applicable; there is only one set of specifications for this measure.

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.

Not applicable; there is only one set of specifications for this measure.

Scientific Acceptability: Validity - Other Threats to Validity (Exclusions, Risk Adjustment) (2b.15 - 2b.32)

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

☐ N/A or no exclusions
☒ x Yes, the measure uses exclusions.

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?

All exclusions were determined by careful clinical review and were been made based on clinically relevant decisions. The prevalence of exclusions in the Progression Development Dataset are reported below.

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.

There were 4,919 patients excluded due to having metastatic cancer within one year prior to their Stage 4 CKD diagnosis. The final study cohort was 434,764, so this represents 1.1% of the final cohort.

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.

The measure excludes patients with metastatic and advanced cancers since the outcome is not a reliable signal of care quality among these patients. Many patients in this population may be too ill for dialysis and have a high risk of mortality independent of CKD progression; thus, we find it inappropriate to attribute outcomes for these patients to their nephrologists' quality of care. These exclusions are minimal in practice but are important for measure validity given the outcome.

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

☒ x Statistical risk model with risk factors (specify number of risk factors: 43)
☐ Stratification by risk category (specify number of categories)
☐ Other (please specify here:)

☐ No risk adjustment or stratification

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.

The goal of risk adjustment is to account for differences among nephrologists in patient demographic and clinical characteristics. The measure incorporates risk adjustment to account for factors that are associated with the outcome, vary across providers, and are unrelated to quality of care, so that measure scores reflect true differences in quality of care. Accounting for case-mix differences is important because it recognizes that some providers care for sicker patients who may have higher anticipated progression rates. Through the risk-adjustment modeling, a higher expected outcome rate is set for providers who care for patients with certain risk factors. We identified potential candidate risk factors using a focused literature search, clinical experts' input, and empirical analysis. We used logistic regression with a binary outcome to select risk variables for final models.

We considered age and medical history (comorbidities/frailty) as candidate variables.

- Comorbidities for inclusion in risk adjustment were identified through inpatient and outpatient administrative claims during the twelve months prior to entering the cohort.
- We used Yale-Modified FY20 v24 CC Map that contains 197 CMS condition categories (CMS-CCs), based on publicly available CMS-CCs, to group ICD-10 diagnosis codes into CMS-CCs as candidate clinical risk factors.

Next,

- We examined all condition categories (CMS-CCs).
- Examined frequencies and bivariate associations with outcome (including odds ratios) of all CMS-CCs.
- CMS-CCs that were not statistically significant were removed, unless deemed clinically relevant to the outcome by expert nephrologists (ex: cancer-related CC). Statistical significance was defined by having a p-value less than 0.05 (23 CC removed).
- CMS-CCs with low frequency (<1% of cohort) were grouped into one variable, except for CC1 HIV/AIDS (35 CC grouped).
- CC132 Kidney Transplant Status was split into two: CC132Z ICD-10-CM codes beginning with 'Z' 'codes indicating general aftercare or status); and CC132T ICD-10-CM codes beginning with 'T' 'codes indicative of a kidney failure or complication).

This resulted in 135 candidate risk variables.

Final Risk Variable Selection

We selected the final set of risk variables using bootstrap methods using logistic regression from the candidate variables:

- 500 random samples were generated with replacement.

- For each of the 500 samples, a logistic regression model (binary outcome) was selected by using backward selection approach.
- All variables significant at p<0.0001 were retained in each final bootstrap risk model. For each variable, we note its % retained in the 500 bootstrap models.
- We then selected all variables that were retained in the model which are above 70% threshold (cut-off). The threshold was based on clinical and statistical evaluation to have a clinically meaningful, statistically robust, and parsimonious risk model.
    - Low frequency CC variable was removed. This group was very heterogeneous; removing aligns with many other measures that excluded prior to bootstrap results.
    - Three additional CCs were included that were below the 70% cutoff, for face validity per our expert nephrologists (Dialysis Status [CC134]; Diabetes without Complication [CC19]; and Cirrhosis of Liver [CC28].
    - Proteinuria identified by ICD-10 code (R80.9) was included as a risk variable; Proteinuria is within the CC 179 Minor Symptoms, Signs, Findings, which fell below the 70% cutoff. Adding the whole condition category is not as predictive as adding a specific variable, therefore, only the ICD-10 code for proteinuria has been added. We included the Proteinuria code as a separate variable based on input from nephrologists regarding its clinical relevance and importance for face validity.

There are 43 final risk variables. We evaluated the performance of the model in Cox model with the selected risk factors. Frequencies, parameter estimates, and adjusted hazard ratios are reported in Table 2b.24.

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.

Not applicable; the measure is risk-adjusted.

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

☒ x Published literature
☒ x Internal data analysis
☐ Other (please specify here:)

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

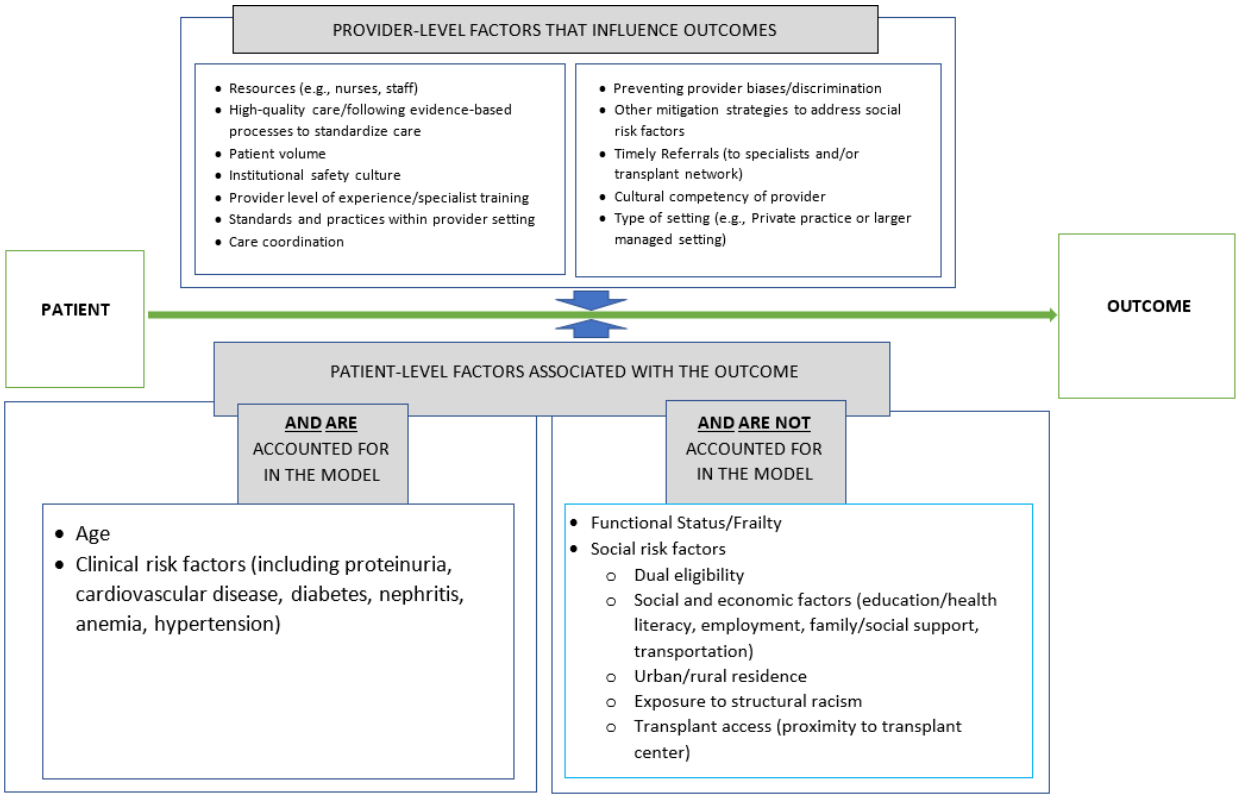
and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added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

Methods for identifying clinical risk factors are detailed in Section 2b.20.
A patient's progression to dialysis is likely also influenced by social risk factors (SRFs). Kidney care providers have the ability to partially or fully address these SRFs and mitigate the impact on progression. We considered whether to adjust for SRF using a comprehensive approach that evaluates the following:

1. Conceptual influence of SRFs on measure outcome (and provider role)
2. Feasibility of utilizing meaningful SRFs in available data
3. Empiric testing of SRFs for inclusion in the measure risk models

Updated NQF guidance emphasizes that developers should share the conceptual model that was used to guide empiric testing and decisions around inclusion of social risk factors within the measure's risk model. [11] Conceptual models should ilustrate the pathway between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured healthcare outcome. Our conceptual model for CKD progression is shown in Figure 2b.23.The conceptual relationship, or potential causal pathways by which these social risk factors could influence the risk of progression to dialysis, are varied and complex. Some social risk factors may, for instance, influence the patient's ability to manage self-care, such as following dietary recommendations. However, the best quality care should slow progression for all patient groups, especially if tailored to a particular patient's situation and preferences. Therefore, the conceptual rationale for risk-adjustment is limited.

Figure 2b.23. Conceptual Model



A review of literature highlights the well-established disparities in CKD outcomes including progression from CKD to ERSD, for patients with social risk factors, including race/ethnicity as well as socioeconomic status [1-5]. The most recent (2022) United States Renal Data Systems (USRDS) report underscores findings from earlier studies; for example, there are race/ethnic disparities in the one-year risk for ESRD among patients with stage 4 or 5 CKD; and while the rate of ERSD were higher among patients living in lower-income neighborhoods for all race/ethnicity groups, disparities by race/ethnicity were still apparent within income categories [2].

Multiple studies have shown that both lower SES as well as Black race are associated with comorbid conditions that predispose beneficiaries toward CKD, such as hypertension and diabetes, but that inadequate access to or delivery of primary and specialty care also drive outcomes. As compared with white individuals, Black individuals are less likely to have pre-ESRD nephrology care, receive adequate dialysis treatment, have an arteriovenous fistula placed for dialysis access, and have access to kidney transplantation [6-10]. A recent study underscores these findings; among patients with diabetes enrolled in a clinical trial, there were no meaningful differences in outcomes between Black and white patients (outcomes included change in eGFR, incident albuminuria, strictly defined incident CKD, and kidney failure, over a median follow-up time of 4–5 years) [9]. Patients in this clinical trial were receiving standardized type 2 diabetes care, suggesting that disparities can be addressed by delivering standardized, evidence-based, high-quality care [11].

While our literature search for functional status/frailty, Medicaid dual-eligibility, urban/rural residence, and proximity/access to transplant centers did not yield clear evidence of association to progression, we have included these as potential factors in our conceptual model as well based on expert consideration. Dual-eligible status is an indicator of low income and a proxy indicator of various socioeconomic factors that may affect patients' options for CKD treatment. Similarly, urban vs. rural residence may affect the proximity to effective CKD care that may delay progression. Patient's functional status and frailty may also play a role in patients' ability to access care for CKD. Finally, patients' ability to receive kidney transplants (a censoring event for this measure) may be affected by their proximity to transplant centers.

To define a list of SRF indicators that would be feasible to test, we first compiled an initial list of SRFs to consider, using the National Academies of Sciences, Engineering, and Medicine (NASEM) report framework, which categorized social risk factors into four domains:

- Socioeconomic position;
- Race, ethnicity (not biological factors but proxy for the social risk factor of exposure to systemic racism), and cultural factors;
- Social relationships; and
- Residential and community context

Second, we identified candidate SRFs for analyses, based on:

- Internal hypotheses regarding the relationships between the SRF to progression for patients with CKD;
- Potential / perceived ability of a kidney care provider to mitigate the SRF; and
- Data availability and feasibility, including level of analysis (availability of patient-level or area-level data).

Among candidate SRFs, we identified the corresponding variable from different data sources and linked them to the test dataset based on the related patient information. The candidate social risk variables considered are listed in Section 2a.08, and below in Table 2b.23. Candidate social risk factors from Medicare FFS claims including Medicaid dual-eligibility and Black race, while Agency for Healthcare Research and Quality (ARHQ) socioeconomic status (SES) index, and urban residence could be linked to patients at the ZIP code and county level respectively. We did not identify a suitable and feasible SRF indicator for functional status/frailty or proximity to transplant centers.

Table 2b.23. Candidate Social Risk Factors

Variable	Description	Data level
Dual-eligible	Dual-eligible for Medicare and Medicaid vs. Medicare-only (reference variable)	Patient

Variable	Description	Data level
Race	Black race variable vs. non-Black race variables (reference variable). Note: Medicare administrative claims data are not a reliable source for accurate race information except for Black race, as noted in the literature. Included here as above to explore general impact using available data.	Patient
AHRQ SES index	Lowest AHRQ quartile for socioeconomic status indicator (higher score = less social risk) vs. other quartiles (reference variable)	Zip code
Urban resident	Residence in metro area county vs. non-metro county (suburban and rural are considered non-urban) (reference variable)	County

Methods for testing each social risk factor included examining the prevalence and distribution of SRFs, bivariate (unadjusted) associations of SRFs with progression, risk adjusted associations, and risk model performance when incorporating SRFs, including impact on provider performance scores.

Some patient level factors potentially associated with the outcome but not accounted for in the model were ultimately not tested for the following reasons. Social and economic factors such as health literacy and transportation do not have enough available data to accurately capture and adjust for their influence. Transplant is considered a censoring event in our measure and does not count as an adverse event rather it is encouraged as a positive event. There are a large number of external and individual factors that can impact transplant access other than geographic location, including patient willingness to receive care and the availability of kidneys for transplant once on the waitlist. Frailty is potentially associated with progression of CKD but is not accounted for in this measure due to a lack of suitable data; however, aspects of frailty will be captured in the clinical factors that are adjusted for in the model.

References

1. Nicholas SB, Kalantar-Zadeh K, Norris KC: Socioeconomic disparities in chronic kidney disease. *Adv Chronic Kidney Dis* 22: 6–15 (2015).
2. United States Renal Data System: USRDS annual data report: Epidemiology of kidney disease in the United States (2022). Available at https://usrds-adr.niddk.nih.gov/2022/supplements-covid-19-disparities/14-racial-and-ethnic-disparities; accessed January 5, 2023.
3. Patzer RE, McClellan WM: Influence of race, ethnicity and socioeconomic status on kidney disease. Nat Rev Nephrol 8: 533–541 (2012).
4. Udler MS, Nadkarni GN, Belbin G, Lotay V, Wyatt C, Gottesman O, Bottinger EP, Kenny EE, Peter I: Effect of genetic African ancestry on eGFR and kidney disease. J Am Soc Nephrol 26: 1682–1692, 2015 Fedewa SA, McClellan WM, Judd S, Gutiérrez OM, Crews DC: The association between race and income on risk of mortality in patients with moderate chronic kidney disease. BMC Nephrol 15: 136 (2014).

5. Owen, W. F. Jr, Chertow, G. M., Lazarus, J. M. & Lowrie, E. G. Dose of hemodialysis and survival: differences by race and sex. JAMA 280, 1764–1768 (1998).
6. Leonard, M. B., Stablein, D. M., Ho, M., Jabs, K. & Feldman, H. I. Racial and center differences in hemodialysis adequacy in children treated at pediatric centers: a North American Pediatric Renal Transplant Cooperative Study (NAPRTCS) report. J. Am. Soc. Nephrol. 15, 2923–2932 (2004).
7. Wasse, H., Hopson, S. D. & McClellan, W. Racial and gender differences in arteriovenous fistula use among incident hemodialysis beneficiaries. Am. J. Nephrol. 32, 234–241 (2010).
8. Hall, Y. N., Choi, A. I., Xu, P., O'Hare, A. M. & Chertow, G. M. Racial ethnic differences in rates and determinants of deceased donor kidney transplantation. J. Am. Soc Nephrol. 22, 743–751 (2011).
9. Prakash, S. et al. Racial composition of residential areas associates with access to pre-ESRD nephrology care. J. Am. Soc. Nephrol. 21, 1192–1199 (2010).
10. Gerber C, Cai X, Lee J, et al. Incidence and Progression of Chronic Kidney Disease in Black and White Individuals with Type 2 Diabetes. Clin J Am Soc Nephrol. 2018;13(6):884-892. doi:10.2215/CJN.11871017
11. Assistant Secretary for Planning and Education Office of Health Policy Report to Congress: Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program; accessed at: https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//195046/Social-Risk-in-Medicare's-VBP-2nd-Report-Executive-Summary.pdf
12. National Quality Forum: Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk Within Healthcare Performance Measurement (2022). Available at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=97639; accessed March 16, 2023.

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.

The final 43 risk variables with frequencies, estimates, and hazard ratios (HR) with 95% confidence interval using Cox Proportional Hazard Model with Frailty Regression Model are listed in Table 2b.24 below.

Table 2b.24. Parameter Estimates for Final Risk Variables Using Cox Proportional Hazard Model with Frailty Regression Model, Progression Development Dataset (N= 434,764 Patients)

Risk Variable	Percentage	Parameter Estimates (Standard Error)	Hazard Ratio (95% Confidence Interval)
Age: Mean (SD)	75.86 (10.23)	-0.032 (0.001)	0.969 (0.967-0.970)
Proteinuria (ICD-10 DX Code R80.9)	22.58	0.353 (0.015)	1.423 (1.381-1.466)
Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock (CC2)	10.08	-0.256 (0.024)	0.774 (0.738-0.811)

Risk Variable	Percentage	Parameter Estimates (Standard Error)	Hazard Ratio (95% Confidence Interval)
Diabetes with Chronic Complications (CC18)	55.77	0.312 (0.024)	1.365 (1.302-1.433)
Diabetes without Complication (CC19)	56.55	0.088 (0.024)	1.092 (1.042-1.145)
Morbid Obesity (CC22)	13.78	-0.088 (0.019)	0.916 (0.882-0.952)
Other Significant Endocrine and Metabolic Disorders (CC23)	35.94	0.147 (0.015)	1.158 (1.125-1.192)
Disorders of Fluid/Electrolyte/Acid-Base Balance (CC24)	51.23	0.179 (0.017)	1.196 (1.157-1.235)
Cirrhosis of Liver (CC28)	2.69	0.101 (0.038)	1.106 (1.026-1.191)
Chronic Hepatitis (CC29)	1.40	0.264 (0.044)	1.302 (1.196-1.418)
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (CC40)	11.24	-0.130 (0.023)	0.878 (0.839-0.919)
Osteoarthritis of Hip or Knee (CC42)	17.99	-0.114 (0.020)	0.892 (0.858-0.928)
Osteoporosis and Other Bone/Cartilage Disorders (CC43)	24.81	-0.129 (0.017)	0.879 (0.849-0.909)
Iron Deficiency and Other/Unspecified Anemias and Blood Disease (CC49)	71.63	0.438 (0.021)	1.55 (1.487-1.615)
Dementia Without Complication (CC52)	12.10	-0.276 (0.027)	0.759 (0.719-0.800)
Major Depressive, Bipolar, and Paranoid Disorders (CC59)	11.15	-0.142 (0.024)	0.867 (0.828-0.909)
Other Psychiatric Disorders (CC63)	19.07	-0.135 (0.019)	0.873 (0.841-0.907)
Congestive Heart Failure (CC85)	48.08	0.173 (0.017)	1.189 (1.150-1.230)
Angina Pectoris (CC88)	11.20	-0.119 (0.023)	0.888 (0.849-0.928)
Coronary Atherosclerosis/Other Chronic Ischemic Heart Disease (CC89)	49.35	0.047 (0.016)	1.048 (1.015-1.082)
Hypertension (CC95)	94.91	0.253 (0.040)	1.288 (1.191-1.394)
Specified Heart Arrhythmias (CC96)	36.78	-0.185 (0.017)	0.831 (0.804-0.859)
Other and Unspecified Heart Disease (CC98)	27.58	0.117 (0.017)	1.125 (1.088-1.163)
Other Circulatory Disease (CC109)	30.57	-0.091 (0.016)	0.913 (0.884-0.942)
Pleural Effusion/Pneumothorax (CC117)	15.58	0.245 (0.020)	1.278 (1.229-1.328)
Proliferative Diabetic Retinopathy and Vitreous Hemorrhage (CC122)	1.30	0.168 (0.042)	1.183 (1.089-1.285)
Diabetic and Other Vascular Retinopathies (CC123)	7.50	0.140 (0.024)	1.151 (1.097-1.207)

Risk Variable	Percentage	Parameter Estimates (Standard Error)	Hazard Ratio (95% Confidence Interval)
Kidney Transplant Status: ICD-10-CM codes beginning with 'Z' (CC132Z; includes Z4822 Encounter for aftercare following kidney transplant; and Z940 Kidney transplant status)	2.48	-0.242 (0.032)	0.785 (0.738-0.836)
Dialysis Status (CC134)	3.08	0.283 (0.023)	1.328 (1.268-1.39)
Acute Renal Failure (CC135)	43.62	0.255 (0.017)	1.291 (1.247-1.336)
Chronic Kidney Disease, Stage 5 (CC136)	14.56	1.264 (0.016)	3.538 (3.429-3.651)
Chronic Kidney Disease, Severe (Stage 4) (CC137)	92.87	-0.094 (0.038)	0.911 (0.844-0.982)
Chronic Kidney Disease, Moderate (Stage 3) (CC138)	73.73	-0.399 (0.016)	0.671 (0.651-0.692)
Chronic Kidney Disease, Mild or Unspecified (Stages 1-2 or Unspecified) (CC139)	83.01	0.188 (0.025)	1.206 (1.148-1.268)
Unspecified Renal Failure (CC140)	10.27	0.223 (0.019)	1.249 (1.204-1.297)
Nephritis (CC141)	5.77	0.320 (0.022)	1.377 (1.319-1.437)
Urinary Tract Infection (CC144)	33.73	-0.154 (0.016)	0.857 (0.831-0.884)
Other Urinary Tract Disorders (CC145)	46.19	0.128 (0.015)	1.137 (1.104-1.171)
Other Female Genital Disorders (CC148)	5.24	-0.202 (0.034)	0.817 (0.765-0.873)
Male Genital Disorders (CC149)	22.68	0.085 (0.017)	1.088 (1.053-1.125)
Complications of Specified Implanted Device or Graft (CC176)	5.20	0.176 (0.025)	1.192 (1.135-1.253)
Other Complications of Medical Care (CC177)	8.65	-0.141 (0.024)	0.868 (0.828-0.91)
Alcohol/Cannabis Use or Use Disorder, Mild or Uncomplicated; Non-Psychoactive Substance Abuse; Nicotine Dependence (CC203)	10.11	0.113 (0.021)	1.12 (1.076-1.166)

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.

Social risk factor testing included:

1. Examining the distribution of SRFs among providers (Table 2b.25:1);
2. Examined bivariate (unadjusted) and multivariate (adjusted) relationships of SRFs with progression to dialysis (Table 2b.25:2);
3. Examined risk model performance with and without each SRF (Table 2b.25:2); and
4. Examined the relationship between measure scores and the provider-proportion of patients with SRFs (Table 2b.25:3 and Figures 2b.25:1-4).

The prevalence of SES factors in the Progression Measure cohort varies across 2,854 measured entities as shown in Table 2b.25-1. At the median provider, 14.7% of patients are dual-eligible; 18.0% have bottom-quartile AHRQ SES; 8.5% are of Black race; and 96.7% live in urban areas.

Table 2b.25:1. Provider-level distribution of social risk factors in the Progression Measure cohort (n=2,854)

SRF	Median provider-level SRF prevalence (IQR)
Dual Eligible	14.3% (7.7%-25.0%)
Low AHRQ SES	18.0% (4.6%-38.0%)
Black Race	8.5% (1.3%-22.5%)
Urban residence	96.9% (74.7%-100%)

To understand the relationship between each SRF and the outcome, we compared the bivariate (unadjusted) association for each SRFs and compared it with the association in the presence of all the clinical and demographic risk variables in the model (Table 2b.25:2). We summarize the results in the narrative section below the table.

Table 2b.25:2. Bivariate Associations and Multivariate Associations Using Cox Proportional Hazard Regression Models Between SRF and Outcome (Progression to Dialysis). Adjusted Models Include 43 Clinical Factors, Including Age (N= 434,764)

Social Risk Factors	Unadjusted (Bivariate) estimate (SE)	Unadjusted Hazard Ratio (95% CI)	Adjusted (Multivariate) estimate (SE)	Adjusted Hazard Ratio (95% CI)	C-statistic (adjusted model)
None (clinical risk model, 43 factors)	*	*	*	*	0.792
Dual Eligibility	0.388 (0.017)	1.475 (1.426-1.525)	-0.059 (0.019)	0.943 (0.909-0.978)	0.792
Low AHRQ SES	0.184 (0.016)	1.202 (1.165-1.241)	0.008 (0.016)	1.008 (0.976-1.04)	0.792
Race (Black)	0.500 (0.017)	1.649 (1.594-1.707)	0.116 (0.018)	1.123 (1.084-1.163)	0.792

Social Risk Factors	Unadjusted (Bivariate) estimate (SE)	Unadjusted Hazard Ratio (95% CI)	Adjusted (Multivariate) estimate (SE)	Adjusted Hazard Ratio (95% CI)	C-statistic (adjusted model)
Urban	0.035 (0.017)	1.036 (1.001-1.071)	0.039 (0.017)	1.039 (1.005-1.075)	0.792

*Intentionally left blank

- Dual eligibility: The unadjusted hazard ratio (1.475) suggests dual eligible patients have a higher risk of progression than those who are not dual eligible; however, once adjusted for comorbidities, dual eligible patients become significantly slightly less likely to progress to the outcome (adjusted hazard ratio 0.943). This suggests that the increased risk of progressing to dialysis associated with dual eligibility is explained by greater comorbidity among dual eligible patients.
- AHRQ SES Index: Although having lower neighborhood economic status is associated with the outcome of progression in the unadjusted model (hazard ratio 1.202), once adjusted for comorbidities, the relationship is no longer significant (hazard ratio 1.008, 95% CI, 0.976-1.04). This suggests that the increased risk of progression among those from lower-SES neighborhoods is explained by greater comorbidity among those patients.
- Black Race: Although unadjusted hazard ratio (1.649) shows a greater risk of progression among Black patients compared with non-Black patients, the adjusted model with comorbidities greatly attenuates that risk (hazard ratio 1.123). This suggests that after accounting for comorbidities, Black patients have a small but significant increased risk of progression
- Urban: Patients living in urban areas have a significantly but very slightly higher risk of progression to the outcome (unadjusted hazard ratio 1.036, adjusted ratio 1.039). The practical significance of this association is likely minimal.

We also examined model performance (c-statistic) with and without each SRF and found that the c-statistic was unchanged (Table 2b.25:2)

Finally, we examined the correlations between provider's Risk Standardized Incidence Rate (RSIR) and the proportion of their patients with each SRF, both overall and within each quintile of the SRF. As shown in Table 2b.25:3 and Figures 2b.25:1-4 below, there is no significant association between a provider's risk adjusted score and their proportion of dual eligible, low-SES, or urban patients, and a significant association only within the second quintile of providers by percent Black race. Importantly, for all of the SRFs we tested, there is no relationship between provider scores and the proportion of patients with SRFs for providers with the highest proportion of patients with SRFs (5^{th} quintile).

Table 2b.25:3. Correlation coefficients between RSR & proportion of patients in disadvantaged group by SRF.

Quintile	Dual eligible (p-value)	Black race (p-value)	Low SES (p-value)	Urban residence (p-value)
1st	0.026 (0.601)	0.024 (0.631)	0.059 (0.242)	-0.087 (0.086)
2nd	0.040 (0.431)	0.112 (0.027)	0.023 (0.653)	-0.023 (0.647)
3rd	0.036 (0.469)	0.017 (0.736)	-0.078 (0.120)	0.013 (0.798)
4th	0.038 (0.454)	0.036 (0.477)	0.056 (0.268)	-0.011 (0.835)
5th	0.064 (0.207)	-0.010 (0.842)	0.038 (0.454)	n/a*
Total	0.015 (0.515)	0.052 (0.021)	0.019 (0.410)	0.025 (0.261)

**More than 20% of providers have 100% urban patients so there is no variation to measure within the top quintile.*

Figure 2b.25:1. Risk Standardized Ratio (RSR) by Percent Dual Eligible

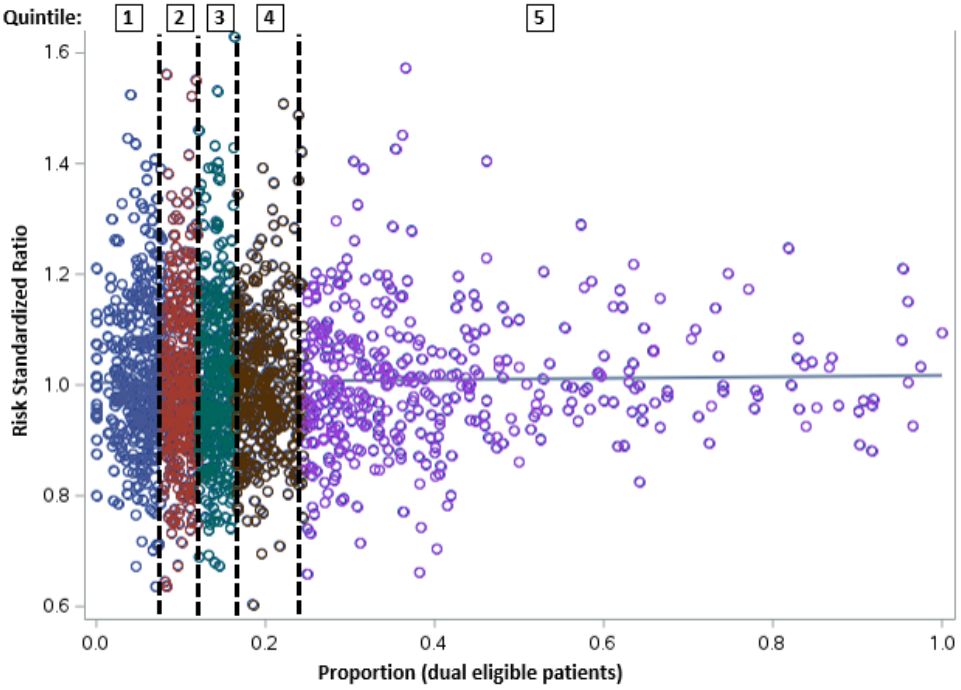


Figure 2b.25:2. RSR by Percent Patients of Black Race

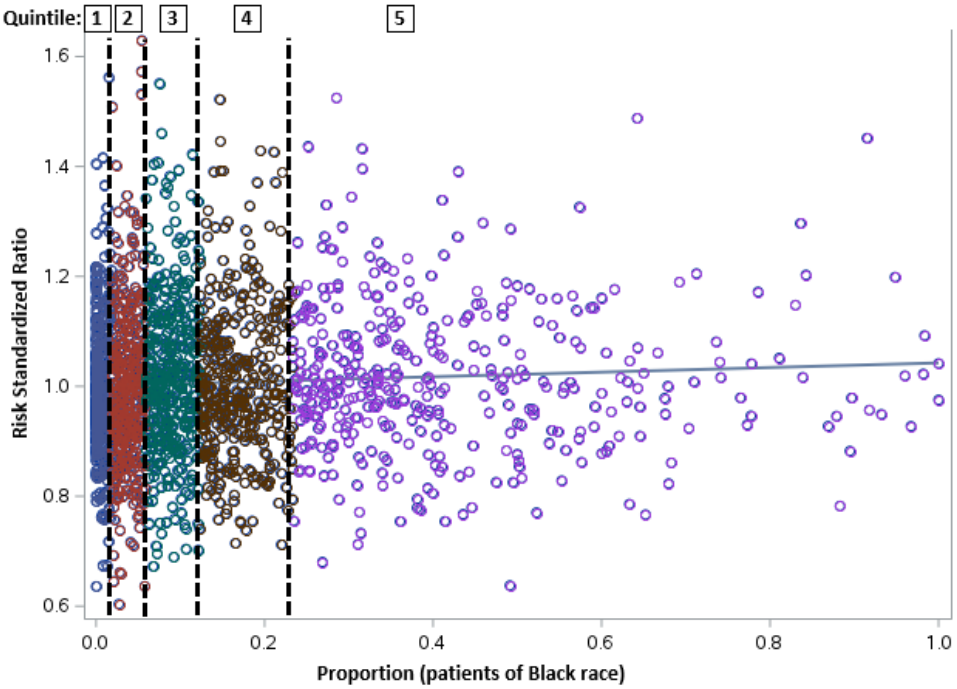


Figure 2b.25:3. RSR by Percent Patients of Low (bottom-quartile SES Index) SES

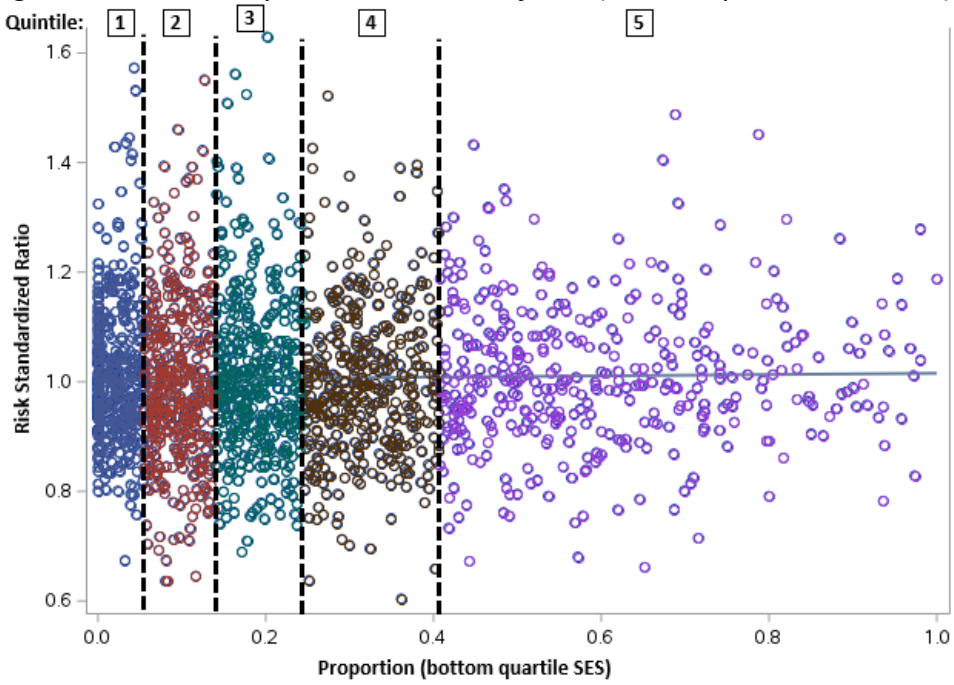
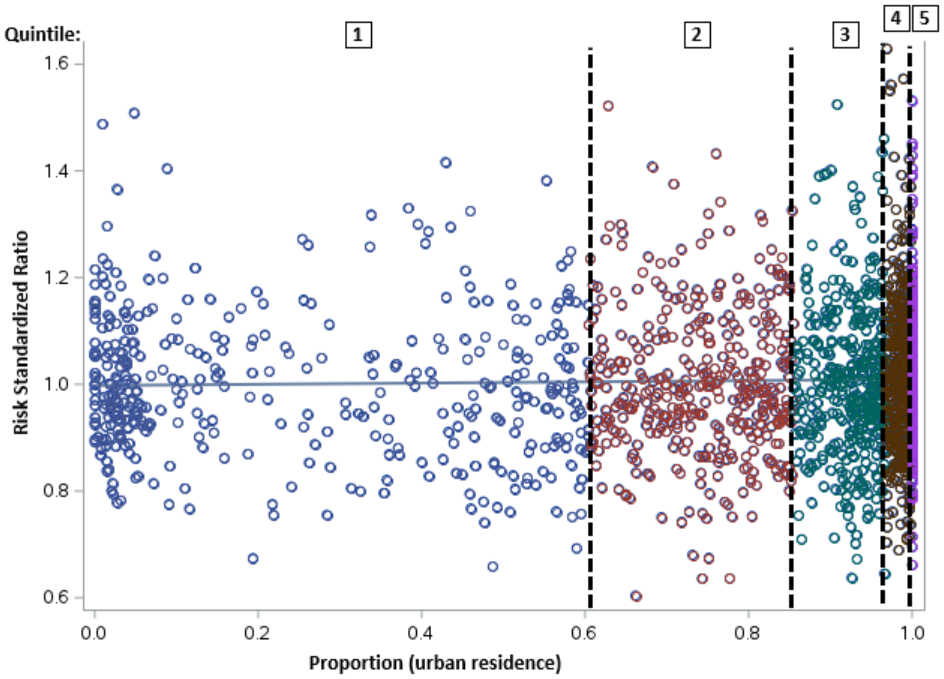


Figure 2b.25:4. RSR by Percent Patients of Urban Residence



Summary and Conclusion

Based on our conceptual model, we examined the impact of including four social risk factors (dual eligibility, low AHRQ SES, race, and urbanicity) into the CKD Progression Measure. We found that while odds of the outcome in a bivariate model are higher among patients who are dual-eligible, Black, have low SES, or reside in an urban county, the relationship between each variable and the outcome is greatly attenuated in a multivariable model (dual-eligibility hazard ratio is below 1, low AHRQ SES variable is no longer significant), suggesting that the clinical risk variables account for most of the risk. Importantly, there is no statistically significant relationship between any of the social risk factors we tested and measure scores among nephrology practices with the highest proportion of patients with social risk factors. Therefore, because there is minimal impact on provider scores, and due to the tradeoff between unintended consequences of adjusting for social risk factors and potentially masking differential care for patients with social risk factors, in particular for Black patients, we did not include social risk factors in the final model. We will revisit this decision during periodic re-evaluation of the measure.

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.

To test model performance, we divided the mortality development dataset randomly in half into a "derivation sample" and a "validation sample." We fit the model to the data in the derivation sample to specify the model coefficients in a hierarchical logistic regression model, then used those coefficients in the validation sample to confirm the model is generalizable and well-calibrated.

We computed three summary statistics for assessing model performance[1].

Discrimination Statistics

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

Calibration Statistics

(3) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)

In addition, we plotted calibration curves (figures 2b.29:1-2) comparing the observed to predicted mortality at the patient level, within each decile of predicted mortality, in both the derivation and validation samples. A well-specified and calibrated model will demonstrate a) clear increasing trend; b) close correspondence between the predictions and the observations; and c) similar results in the validation as the derivation sample.

Reference:

1. Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

2b.27) Provide risk model discrimination statistics.

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

The Harrel's C-statistic for the full Progression Development Dataset sample (n=434,764), evaluating the risk model using Cox proportional hazard model, is 0.792. Table 2b.27 below shows our model testing results for the derivation and validation samples.

Table 2b.27. Risk Model Performance, Progression Development Dataset (N=434,764)

Model Performance Statistic	Development Sample	Validation Sample
Number of Patients	217,382	217,382
Progression Delay Risk	4.714	4.745
Calibration (γ0, γ1)	(0, 1)	(-0.026, 0.987)
Discrimination- Predictive ability (lowest decile %, highest decile %)	(0.4%, 20.8%)	(0.4%, 20.4%)
C-statistic	0.800	0.798

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

Please see table 2b.27 above for calibration (overfitting) results; please see section 2b.29 below for risk-decile plots (figures 2b.29:1-2).

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.

Observed and predicted decile plots, for the Derivation and Validation samples of the Progression Development Dataset

Figure 2b.29:1. Observed vs. predicted progression risk by decile of predicted (derivation sample)

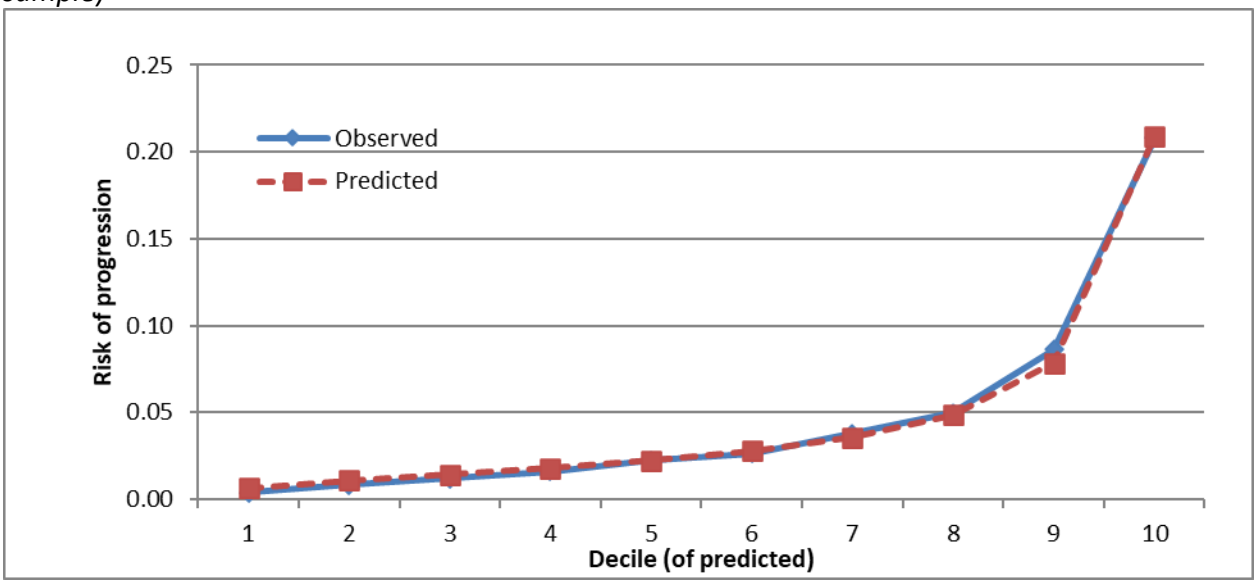
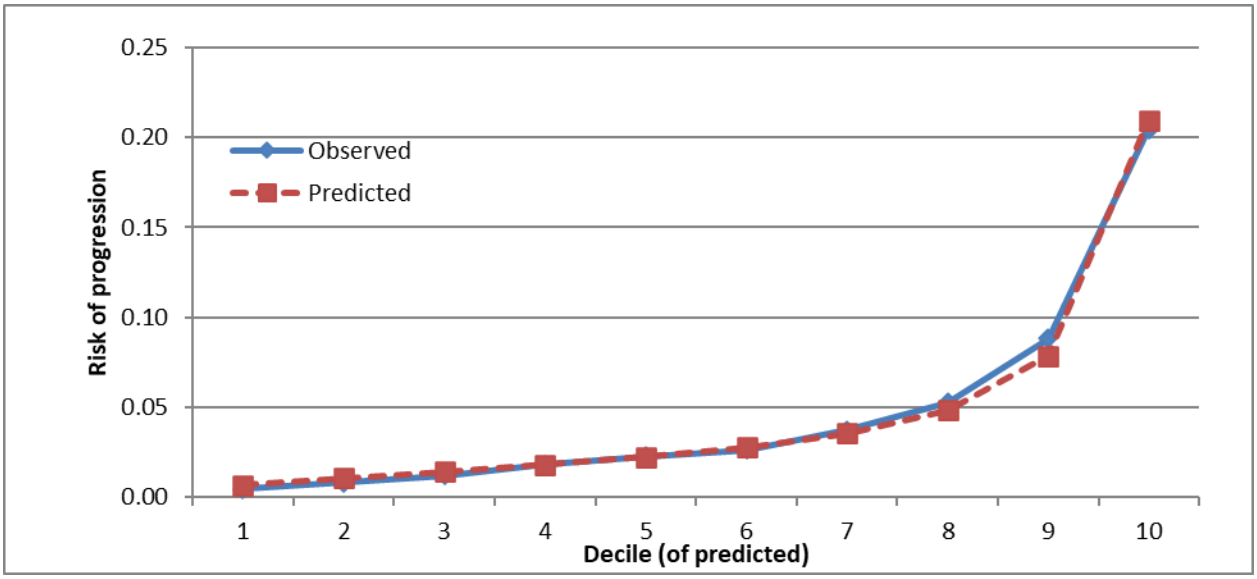


Figure 2b.29:2. Observed vs. predicted progression risk by decile of predicted (validation sample)



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

The measure is not stratified.

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?

Discrimination Statistics
The C-statistic was 0.792 in the full development dataset, 0.800 in the derivation sample and 0.798 in the validation sample, indicating good model discrimination. The model's predictive ability shows a wide range between the lowest decile and highest decile in both samples, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics
Over-fitting (Calibration γ0, γ1)
If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from 1, there is potential evidence of over-fitting. The validation sample calibration values of γ0 = -0.026 and γ1 =0.987 indicate good calibration of the model.

Risk Decile Plots
Higher deciles of the predicted outcomes are closely associated with higher observed outcomes, which show a good calibration of the model. This plot indicates excellent discrimination of the model and good predictive ability. The similar results in both the derivation and validation samples indicates the model has high generalizability.

Overall Interpretation
Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics.

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.

Empiric Validity Testing: Methods

As a form of empiric validity testing, we validated the clinical risk model containing condition data. We compared the predictive modeling of patient-level results with only claims-based clinical conditions (ICD-10 codes) shown in section 2b.24 to a model additionally including eGFR abstracted from EHR data. This testing was conducted to ensure confidence in a claims-based clinical risk model using clinical conditions The addition of any EHR data would require an additional data source and provider burden, so would need to be considered carefully.

Testing for empiric validity used the **Progression EHR Datasets A and B.** Due to these datasets not having access to Medicare enrollment data for the outcome, adaptations were made. A proxy outcome for ESRD requiring maintenance dialysis was defined as a patient having two encounters of the ICD-10 code N18.6 (ESRD), the first during the performance year and the second within the next 12-months. A patient's eGFR was defined as the median of all eGFRs values found 6-months prior and including time zero encounter (if later than January 1). The risk model used two eGFR variables: 1) "eGFR": Continuous variable of median of eGFRs, both for patients with a known eGFR and those with imputed eGFR; and 2) "eGFR Unknown indicator": An indicator variable for those without any eGFR, as a missingness indicator.

Empiric Validity Testing: Results

C-statistics for the risk model with claims-based risk factors only vs. including eGFR values for Progression EHR Datasets A and B are presented in Table 2b.03:1 below.

Table 2b.32:1. C-statistics, progression datasets A and B, with vs. without eGFR values

Dataset	C-statistic, no eGFR (claims only)	C-statistic, with eGFR	C-statistic, difference (with-without eGFR)
Progression EHR dataset A (n=5,658)	0.794	0.865	+0.071
Progression EHR dataset B (n=14,070)	0.779	0.843	+0.064

Calibration plots of observed vs. predicted risk of progression (by deciles of predicted risk), with and without eGFR, for Progression EHR Datasets A and B are presented in Figures 2b.03:1-4 below.

Figure 2b.32:1. Progression EHR Dataset A, without eGFR (n=5,658)

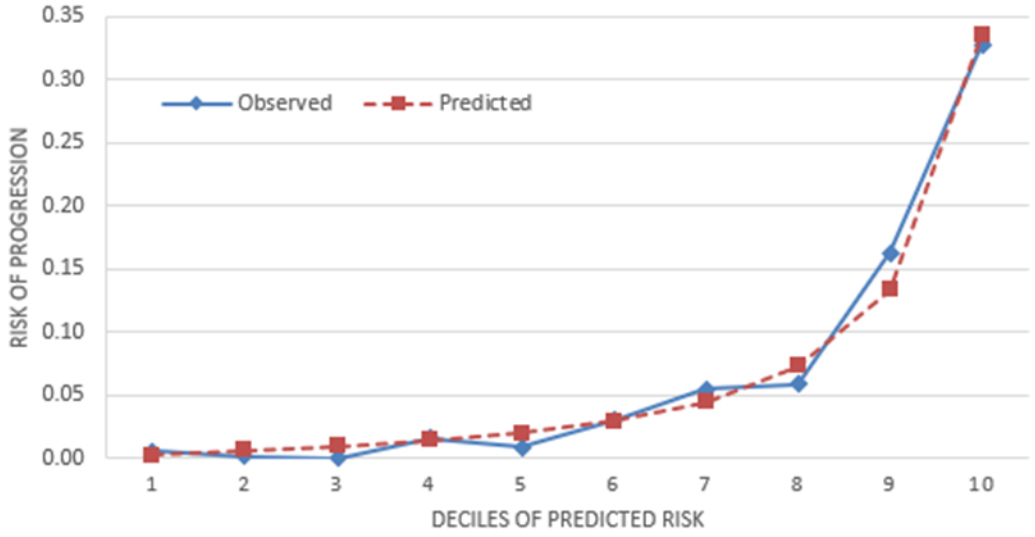


Figure 2b.32:2. Progression EHR Dataset A, with eGFR (n=5,658)

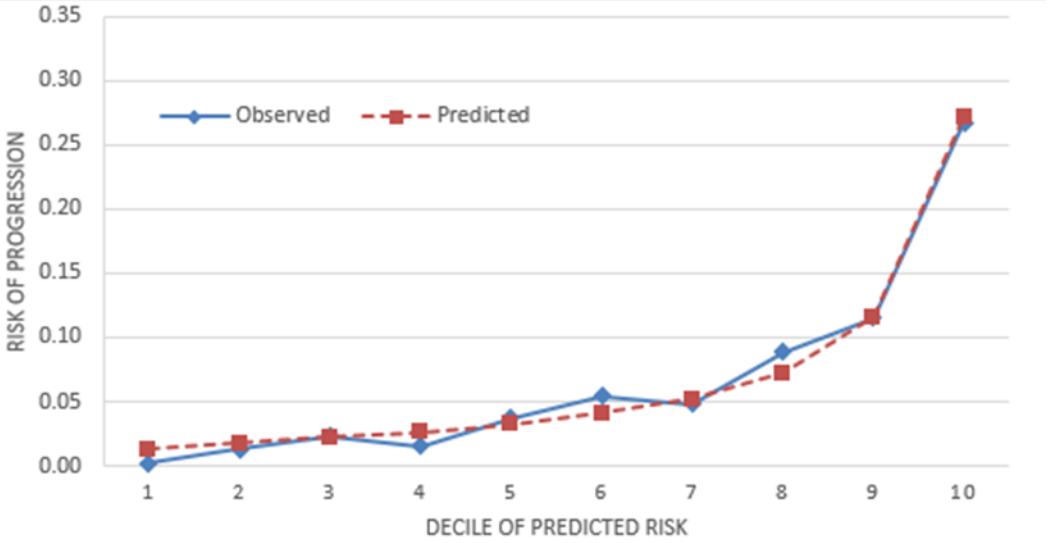


Figure 2b.32:3. Progression EHR Dataset B, without eGFR (n=14,070)

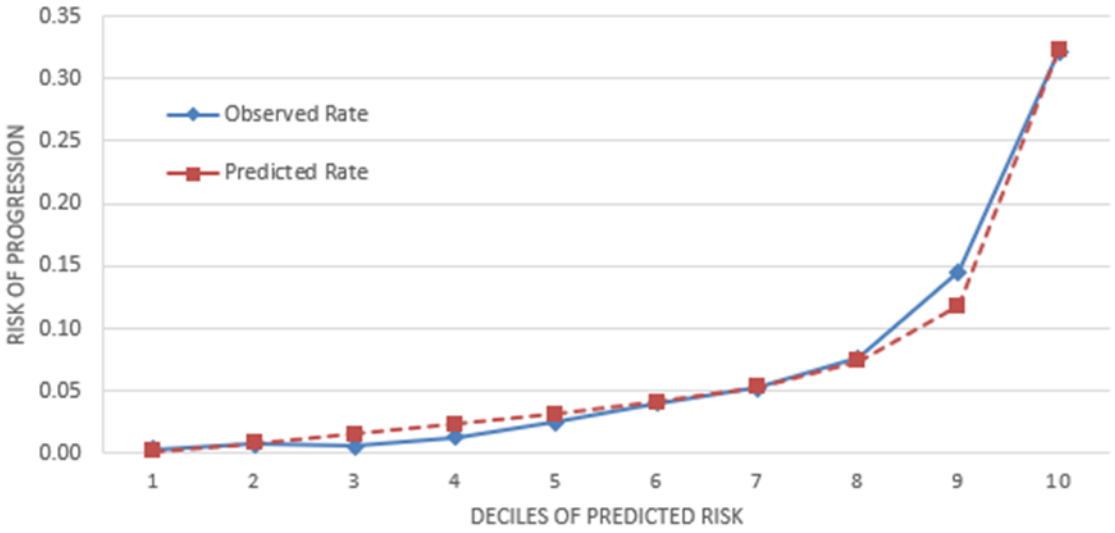
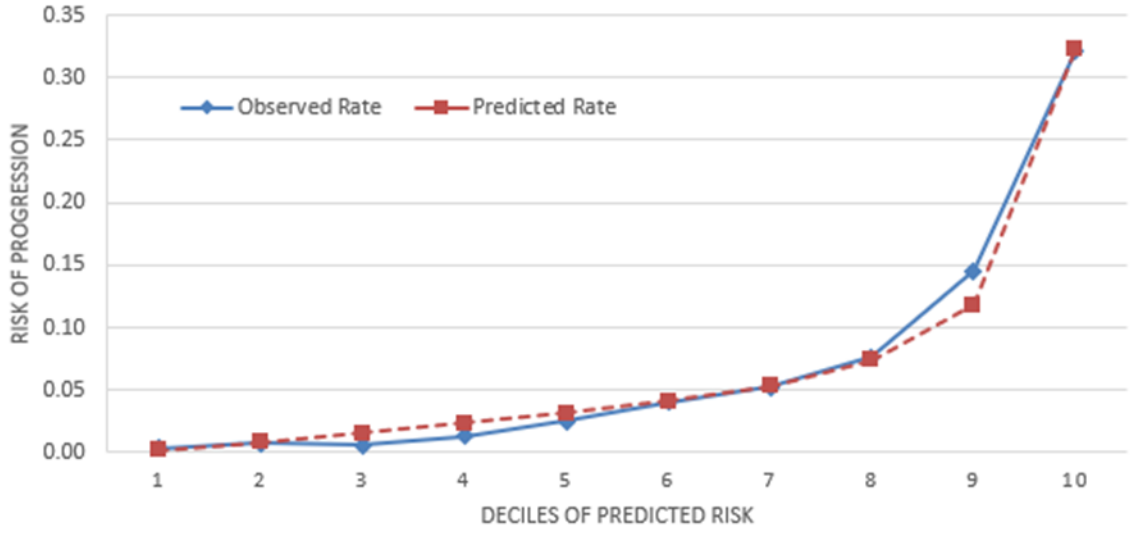


Figure 2b.32:4. Progression EHR Dataset B, with eGFR (n=14,070)



Empiric Validity Testing: Interpretation

Our testing results support the validity of the entirely claims-based risk model and do not demonstrate a key need to additionally include EHR data.

Inclusion of eGFR in the patient-level risk model resulted in a modest improvement to the c-statistic from 0.794 to 0.865 (+0.071) in EHR dataset A. In Progression EHR dataset B, the c-statistic similarly improved by +0.064 from 0.779 to 0.843. It should be noted that the values of 0.794 and 0.779 using only claims-based factors are already quite high, and we would expect adding any EHR data to result in at least a minor increase in c-statistic (which can only increase as additional variables are added). Particularly when weighed against the additional burden that would be required of providers to report eGFR values and the already strong performance

of the claims-only model, the marginal improvement of adding eGFR data to the model is not greatly meaningful.

The calibration plots (figures 2b.03:1-4) demonstrate high alignment between predictions and observation for both datasets whether including eGFR or not, suggesting that on aggregate the marginal benefit of adjusting for eGFR in addition to the claims-based risk factors is minimal. With or without eGFR, there is a clear difference between patients with higher vs. lower predicted risk that corresponds well to observed progression among those patients.

Feasibility (3.01 - 3.07)

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

☐ Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
☒ x Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
☐ Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
☐ Other (Please describe)

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. ALL data elements are in defined fields in electronic health records (EHRs)

☐ ALL data elements are in defined fields in electronic claims
☐ ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)
☒ x ALL data elements are in defined fields in a combination of electronic sources
☐ Some data elements are in defined fields in electronic sources
☐ No data elements are in defined fields in electronic sources
☐ Patient/family reported information (may be electronic or paper)

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.

Not applicable; all data elements needed to compute the performance scores are from electronic sources.

3.04) Describe any efforts to develop an eCQM.

There are currently no efforts underway to develop an eCQM.

3.05) Complete and attach the eCQM-Feasibility-Scorecard.xls file.

Not applicable; this is not an eCQM.

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.

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

Administrative data are routinely collected as part of the billing process. Because completion of claims is required for hospital reimbursement, there is little missing data. The measures do not require any additional data collection.

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.

There are no fees, licensing, or other requirements to use any aspect of this measure as specified.

Use (4a.01 – 4a.10)

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.

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**

☐ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☒ x Not in use
☐ Use unknown
☐ Other (please specify here:)

4a.02) Check all planned uses.

☐ Public reporting
☐ Public Health/Disease Surveillance
☒ x Payment Program
☐ Regulatory and Accreditation Program
☐ Professional Certification or Recognition Program
☒ x Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (internal to the specific organization)
☐ Measure Currently in Use
☐ Other (please specify here:)

Name of program & sponsor: Kidney Care Choices Model (CMS Innovation Center)

URL: https://innovation.cms.gov/innovation-models/kidney-care-choices-kcc-model

Purpose: To help reduce cost and improve quality of care for patients with late-stage CKD and ESRD while delaying the need for dialysis and encouraging kidney transplantation.

Geographic area and number and percentage of accountable entities and patients included: Kidney Care Choices is a voluntary payment model open to providers nationwide. As of 2023 there will be approximately 130 nephrology practices across the model options, caring for approximately 250,000 total aligned patients.

Level of measurement and setting: Nephrology practices (including nephrologists and non-physician clinicians who specialize in nephrology or primarily provide nephrology services)

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?

This is a newly developed measure. The measure is planned for implementation in the voluntary Kidney Care Choices model as soon as 2024.

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.

This measure is planned for use in the voluntary Kidney Care Choices Model beginning in 2024.

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.

Not applicable; the measure has not yet been implemented.

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.

Not applicable; the measure has not yet been implemented.

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

Not applicable; the measure has not yet been implemented.

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

The measure has not been implemented. To gain relevant perspectives during development, CORE included several practicing nephrologists in the TEP. The TEP convened at several key points in the development process at which these individuals offered the following feedback:

- They emphasized that delaying the progression of CKD is an important outcome for both providers and patients, and noted the lack of care coordination between providers, which could be ameliorated by the implementation of such a measure.
- They agreed that the measure construct is useful and a valid aspect of quality to measure.
- They noted there are no current measures that focus on the prevention of CKD progression.
- They agreed that the measure is valid as specified and appropriate to use in the context of a voluntary payment model.

Additionally, CORE worked closely with two practicing nephrologists (Dr. Deidra Crews of the Johns Hopkins University School of Medicine, and Dr. F. Perry Wilson of the Yale University School of Medicine) as clinical subject matter experts throughout the development process, soliciting input on a regular basis. Dr. Crews and Dr. Wilson provided their guidance on all aspects of measure development, including but not limited to building the risk model, specifying cohort exclusions, and developing the time-to-event concepts. Importantly, they helped identify specific conditions that were vital for inclusion in the risk model and advocated for the exclusion of inpatient CKD claims in identifying the cohort.

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

CORE engaged with the CMS Innovation Center Kidney Care Choices model team iteratively throughout the development process to ensure the measure specifications align with the goals and requirements of the model. Key feedback from this perspective included:

- Supported cohort definition (Medicare FFS beneficiaries with confirmed Stage 4 CKD), acknowledging some imprecision with respect to coding variations in claims.
- Agreed with inclusion of patients who were disenrolled from hospice care; prior hospice care should not be considered a blanket exclusion.

- Agreed with definition of ESRD outcome specifically as initiation of chronic dialysis (and not Stage 5 CKD), and of the method for identifying the ESRD outcome
- Supported basing risk model built off cohort (all Medicare beneficiaries with Stage 4 CKD seen by a nephrologist); the risk model based on this wider population supports the Innovation Center's desire to be able to adapt the measure beyond this payment model.

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.

Input from nephrologists in the TEP, from Dr. Crews and Dr. Wilson, and from the CMS Innovation Center Kidney Care Choices model team, played a key role in shaping development of the measure. CORE engaged iteratively with these individuals throughout the development process, incorporating feedback at all key decision points to create a measure broadly acceptable to these users.

Usability (4b.01 - 4b.03)

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.

The measure has not yet been implemented; there are no measured entities among which to assess performance or improvement.

As outlined in section 1a (Importance to Measure and Report: Evidence), delaying progression of CKD and the need for dialysis is broadly a beneficial outcome. Evidence suggests that this progression is partially within a provider's ability to control. By highlighting the risk of progression for a provider's patients compared to other practices (after adjusting for case mix factors outside of the provider's control), the measure will demonstrate opportunities for performance improvement that can lead to better outcomes in the future.

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

Not applicable; the measure has not yet been implemented.

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

Not applicable; the measure has not yet been implemented.

Related and Competing (5.01 - 5.06)

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 endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01) Search and select all endorsed related measures (conceptually, either same measure focus or target population) by going to the PQM website.

(Can search and select measures.)

We searched the National Quality Forum (NQF) Quality Positioning System (QPS) (http://www.quavisitlityforum.org/QPS/) using the keywords "kidney disease," "glomerular filtration," and "eGFR" to identify any endorsed measures that are potentially related or competing.

We identified one conceptually related endorsed measure: the Adult Kidney Disease Angiotensin Converting Enzyme (ACE) or Angiotensin Receptor Blocker (ARB) Therapy measure (NQF # 1662), developed by the Renal Physicians Association. This is a process measure assessing how often patients with CKD (any stage) and proteinuria who do not require renal replacement therapy (RRT) are prescribed ACE inhibitors or ARB therapy. The target population overlaps to some extent with that of the Progression Delay measure but is much broader (including patients with any stage of CKD in a range of care settings), while the Progression Delay measure focuses more narrowly on Stage 4 CKD patients (who share a more specific clinical profile at high risk for progression to ESRD). The focus of both measures are distinct, with the ACE/ARB Therapy measure capturing a specific process of care (the prescription of specific treatments) while the Progression Delay measure focuses on the outcome of progression to ESRD and patients' risk thereof. The ACE/ARB measure is also specified for use in a variety of settings, while the Progression Delay measure is designed to measure nephrology practices specifically.

5.02) Search and select all endorsed competing measures (conceptually, the measures have both the same measure focus or target population) by going to the PQM website.

(Can search and select measures.)

Because the ACE/ARB Therapy measure is a different type of metric (process vs. outcome), reflects a different measure focus (prescription of specific treatment to patients with CKD vs. risk of progression to ESRD), and only partially overlaps in setting and target population, it is not a competing measure with the Progression Delay measure.

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

In addition to the QPS, we searched for keywords “kidney disease,” “glomerular filtration,” and “eGFR” in a number of measure databases and search engines to identify potential competing measures that are not NQF-endorsed, including:

- CMS Measures Inventory Tool (CMIT): https://cmit.cms.gov/CMIT_public/ListMeasures
- OVID search tool
- Most current Measure Applications Partnership (MAP) Pre-Rulemaking Report and MAP List of Measures Under Consideration (MUC): https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Pre-Rulemaking.html
- American Medical Association (AMA) Physician Consortium for Performance Improvement (PCPI) Measures Directory: https://www.thepcpi.org/page/Measures-Directory
- QualityNet: http://www.qualitynet.org/
- Leapfrog Hospital Safety Grade Measures: http://www.hospitalsafetygrade.org/for-hospitals
- Quality Payment Program: https://qpp.cms.gov/mips/explore-measures/quality-measures?search=opioid%20use%20disorder
- National Committee for Quality Assurance (NCQA): https://www.ncqa.org/
- Association of State and Territorial Health Officials (ASTHO): https://www.astho.org/

We also performed a search through the Google search engine and Google Scholar for “chronic kidney disease measure,” and “progression of kidney disease measure” and reviewed the first 30 results for each search.
Ultimately we identified three additional potentially measures:

1. Adult Kidney Disease: Blood Pressure Management;
2. Adult Kidney Disease Laboratory Testing (Lipid Profile);
3. Adult Kidney Disease: Referral to Nephrologist; and

Each measure was also stewarded by the Renal Physicians Association. The first measure (Blood Pressure Management) is an intermediate outcome measure while the others are process measures; none assess the outcome of progression to ESRD. The Laboratory Testing (Lipid Profile) measure was previously endorsed by NQF, but was retired with its endorsement removed prior to development of the Progression Delay measure; the other measures were never endorsed. None of the measures are used in a CMS program.

The first two (Blood Pressure Management and Laboratory Testing (Lipid Profile)) include a target population of patients with stage 3, 4, or 5 CKD, broader than that of the Progression Delay measure (which include strictly patients with stage 4 CKD).The third (Referral to Nephrologist, including patients with CKD not receiving RRT and with eGFR < 30) has the closest alignment to the Progression Delay measure target population.

Ultimately while there may be some overlap in target population for these measures, none of these are competing measures in the context of the Progression Delay measure given the

completely distinct measure focuses, the absence of consensus-based endorsement, and the lack of use in any CMS program.

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

☐ Yes
☒ No

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

The ACE/ARB Therapy process measure, while conceptually related, bears very little overlap in specifications to the Progression Delay outcome measure and any need for harmonization is minimal. As noted these are different types of measures, designed to capture entirely distinct clinical constructs. There is some overlap in target population, but the ACE/ARB Therapy measure is designed for a broad cohort of patients across a range of clinical settings while the Progression Delay measure is focused specifically on patients with stage 4 CKD being cared for by nephrology practices.

As a strictly claims-based measure, there is no additional data collection burden associated with the Progression Delay measure.

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.

Not applicable; there are no competing measures.

Additional (1 - 9)

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.

☒ x Available in attached file
☐ No appendix
☐ Available at measure-specific web page URL identified in sp.09

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

Describe the members' role in measure development.

The TEP provided input on methodological decisions at all stages of the development process. Key feedback proved by the TEP is discussed in sections 1a.02, 2b.02-2b.04, and 4a.08-4a.10.

Table A2. TEP Membership List

Name, Credentials	Professional Role	Organization, Location
Steven Spencer, MD, MPH	Chief Medical Officer	Onslow Memorial Hospital, University of North Carolina, Jacksonville, NC
Wendy St. Peter, PharmD, FNKF, FASN, FCCP	Professor	University of Minnesota, College of Pharmacy, Minneapolis, MN
Erma Boykin, MSN, MBA-HCM, RN, CNN, CLNC	Renal Clinic Nurse	Atlanta VA Medical Center-Renal Clinic, Decatur, GA
Andrew "Drew" Wall, MS	Founder and Chief Innovation Officer	HealthMap Solutions, Inc., Tampa, FL
Eric Martinez, MD	Abdominal Transplant Surgeon; Teaching Faculty	Baylor Scott & White Health; Baylor University Medical Center, Dallas/Fort Worth, TX
Adam Weinstein, MD	CMIO (DaVita); Nephrologist (UMSMG)	DaVita; and University of Maryland Shore Medical Group, Annapolis, MD
Lisa Cormack	Caregiver	North Ridgeville, Ohio
Yaakov Liss, MD	Attending Physician Nephrology Department; Dialysis Unit Medical Director; Vice Chair	CareMount Medical Group; DaVita Celia Dill Dialysis Center; Quality Committee RPA, Brewster, NY
Derek Forfang	Patient; Public Policy Committee Chair	National Kidney Foundation, San Pablo, CA

Name, Credentials	Professional Role	Organization, Location
Titte R. Srinivas, MD	FAST; Chief of Nephrology and Hypertension Division; Medical Director Kidney and Pancreas Transplant Programs	University Hospitals Cleveland Medical Center, Cleveland, OH
Jack Lennon, MBA	Patient; Executive Director	Improving Renal Outcomes Collaboratives, Cincinnati, OH
Richard Knight, MBA	Patient; President	American Association of Kidney Patients, Bowie, MD
Daniel E Weiner MD, MS	Associate Professor of Medicine and Nephrologist; Medical Director of Clinical Research	Tufts Medical Center; and Dialysis Clinic Inc, Boston, MA
Jessie Pavlinac MS, RDN-AP, CSR, LD, FAND	Adjunct Senior Instructor	Oregon Health & Science University; School of Medicine, Portland, OR
Scherly Leon MD, MS, MPH	Medical Director; Nephrologist	Atlantic Dialysis, New York, NY
Milda Saunders MD, MPH	Assistant Professor of Medicine; General Internist	University of Chicago Medicine, Chicago, IL

3) Indicate the year the measure was first released.

Not applicable; the measure has not been released.

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

This is a newly developed measure and has not been revised. The reported specifications were finalized in April 2022.

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

Annual

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

2023

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

N/A

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

N/A

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

N/A