



Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 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's outcome captures beneficiaries with Stage 4 CKD who progress to ESRD and require chronic dialysis. This measure is for nephrology practices (sometimes referred to as "providers" in this submission) who care for patients with Stage 4 CKD.

1b.01. Developer Rationale:

sp.12. Numerator Statement:

The measure outcome is progression from Stage 4 CKD to ESRD requiring chronic dialysis in the measurement year for patients age 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.

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 advanced cancer-related ICD-10 codes from an inpatient encounter. Codes are in attached data dictionary.

Measure Type: Outcome

sp.28. Data Source:

Claims

Other (specify)

Beneficiary enrollment data (including hospice enrollment, ESRD/dialysis enrollment, and vital statistics)

sp.07. Level of Analysis:

Clinician: Group/Practice

IF Endorsement Maintenance – Original Endorsement Date:

Most Recent Endorsement Date:

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

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

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

1. Importance to Measure and Report

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

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

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]

[Response Ends]

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

Describe how and from whom input was obtained.

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

2. Scientific Acceptability of Measure Properties

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

sp.01. Provide the measure title.

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

[Response Begins]

Delay in Progression of Chronic Kidney Disease (CKD) Measure

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

The 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's outcome captures beneficiaries with Stage 4 CKD who progress to ESRD and require chronic dialysis. This measure is for nephrology practices (sometimes referred to as "providers" in this submission) who care for patients with Stage 4 CKD.

[Response Ends]

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

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

Please do not select:

- Surgery: General

[Response Begins]

Renal: Chronic Kidney Disease (CKD)

Renal: End Stage Renal Disease (ESRD)

[Response Ends]

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

[Response Begins]

Access to Care

Care Coordination

Health and Functional Status: Change

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

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

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Adults (Age >= 18)

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

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

Please do not select:

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

[Response Begins]

Clinician: Group/Practice

[Response Ends]

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

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

[Response Begins]

Ambulatory Care

[Response Ends]

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

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

[Response Begins]

None available.

[Response Ends]

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

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

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3753_DataDictionary_CKDProgDelay_01052023.xlsx

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

sp.13. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]

The measure outcome is progression from Stage 4 CKD to ESRD requiring chronic dialysis in the measurement year for patients age 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.

[Response Ends]

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

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

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

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

[Response Begins]

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).

[Response Ends]

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

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

The 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.

[Response Ends]

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

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

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

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

[Response Begins]

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.

Code	Description of Code
OTY00Z0	Transplantation of Right Kidney, Allogeneic, Open Approach
OTY00Z1	Transplantation of Right Kidney, Syngeneic, Open Approach
OTY00Z2	Transplantation of Right Kidney, Zooplastic, Open Approach
OTY10Z0	Transplantation of Left Kidney, Allogeneic, Open Approach
OTY10Z1	Transplantation of Left Kidney, Syngeneic, Open Approach
OTY10Z2	Transplantation of Left Kidney, Zooplastic, Open Approach
50360-50365	Kidney transplant
50380	Kidney transplant
S2065	Kidney transplant

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

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 practices¹.

Service	HCPCS Codes
Office/Outpatient Visit E/M	99201-99205, 99211-99215
Prolonged E/M	99354-99355
Transitional Care Management Services	99495-99496
Advance Care Planning	99497-99498

Service	HCPSC Codes
Welcome to Medicare and Annual Wellness Visits	G0402, G0438, G0439
Chronic Care Management Services	99490

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

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

The cohort excludes patients with advanced or metastatic cancer, defined as specific advanced cancer-related ICD-10 codes from an inpatient encounter. Codes are in attached data dictionary.

[Response Ends]

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

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

[Response Begins]

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.

[Response Ends]

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

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

[Response Begins]

The measure is not currently stratified.

[Response Ends]

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

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

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

[Response Begins]

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

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

43 clinical risk factors (including age)

[Response Ends]

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

Attachment: If available, please provide a sample report.

[Response Begins]

Ratio

[Response Ends]

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

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

[Response Begins]

Better quality = Lower score

[Response Ends]

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

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

[Response Begins]

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

- 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,
- 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),
- At least one occurrence of ICD-10 code N18.4: "CKD, Stage 4 (Severe)" in at least one claim during the performance year,
- 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:

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

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

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

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 $\text{median}(w_j) = 1$. 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.

[Response Ends]

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

Examples of samples used for testing:

- Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
- The sample should represent the variety of entities whose performance will be measured. The [2010 Measure Testing Task Force](#) recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.
- The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.
- When possible, units of measurement and patients within units should be randomly selected.

[Response Begins]

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

[Response Ends]

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

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Beneficiary enrollment data (including hospice enrollment, ESRD/dialysis enrollment, and vital statistics)

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

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

[Response Begins]

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 status¹.

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.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results

should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

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

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

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

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

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

AND

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

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

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

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

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

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

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

Definitions

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

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

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

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

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

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

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

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Beneficiary Enrollment data including hospice enrollment, ESRD/dialysis enrollment, and vital statistics

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

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

[Response Ends]

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

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

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

Please do not select:

- Clinician: Clinician
- Population: Population

[Response Begins]

Clinician: Group/Practice

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

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.

[Response Ends]

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

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

[Response Begins]

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

[Response Ends]

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

[Response Begins]

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
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

Dataset	Applicable Section	Description of Dataset
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.07:1. Dataset Descriptions

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
Dual eligible in 2018	*	*
No	362,861	83.46
Yes	71,903	16.54

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

*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 equation¹.

Reference:

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

[Response Ends]

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

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

[Response Begins]

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}, NQF³, 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 codes⁴. 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 dataset⁵.
- **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 Survey⁶.

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-medicare-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-medicare-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?LinkIdIdentifier=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

[Response Ends]

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

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

Choose one or both levels.

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

We 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. We are using 25 as an example minimum case count.

We used the formula for signal-to-noise reliability presented by Adams et al. to calculate individual clinician-level and TIN-level reliability scores¹. To estimate the overall signal and noise, we first calculated the ICC for the Model Participant, j, using the estimates of between-entity variance τ^2 and the formula for intraclass correlation

coefficient (ICC) presented by Shrout and Fleiss². Specifically, the signal-to-noise reliability score for Model Participant, j , R_j is calculated as:

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

while

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

n_j is the number of patients for the nephrologist j , τ^2 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 $\frac{\pi^2}{6\gamma^2}$ represents the noise and γ is the shape parameter of the Weibull distribution.

R_j 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.

[Response Ends]

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

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

[Response Begins]

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

Table 2a.11. Signal-to-Noise Reliability

[Response Ends]

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

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

[Response Begins]

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.

[Response Ends]

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

[Response Begins]

Empirical validity testing

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

[Response Ends]

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

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

[Response Begins]

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).

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

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

Empiric Validity –Model Validation

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.

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.

Reference:

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

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

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 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 the 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

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.

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

Table 2b.03:1. C-statistics, progression datasets A and B, with vs. without eGFR values

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.

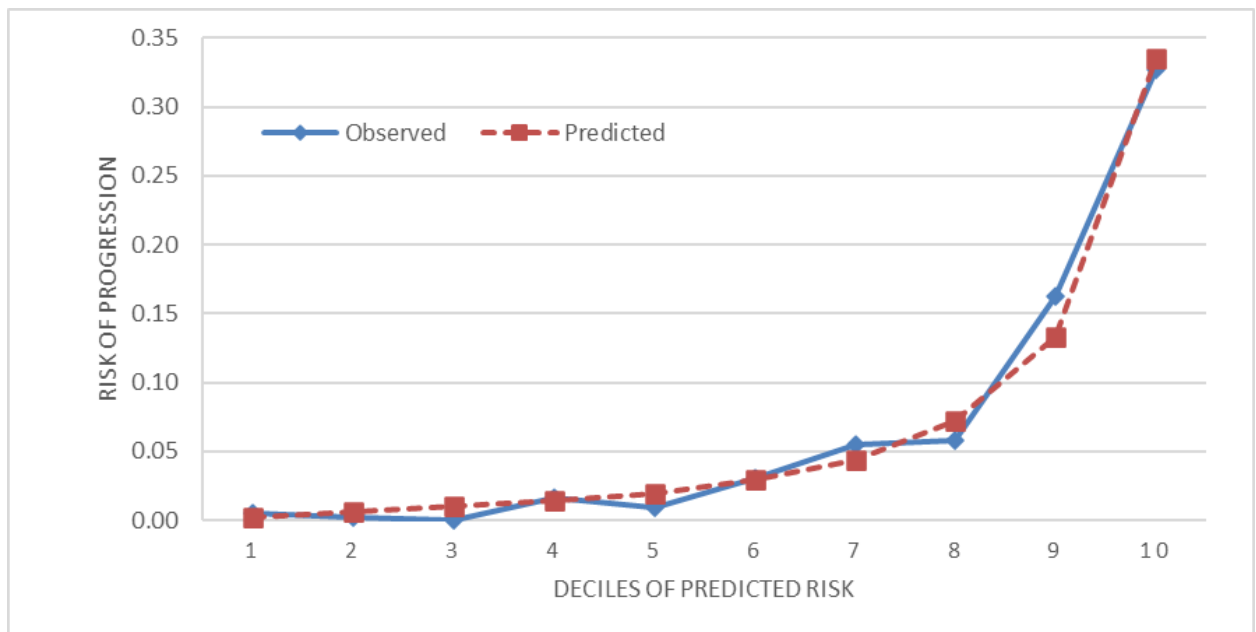


Figure 2b.03:1. Progression EHR Dataset A, without eGFR (n=5,658)

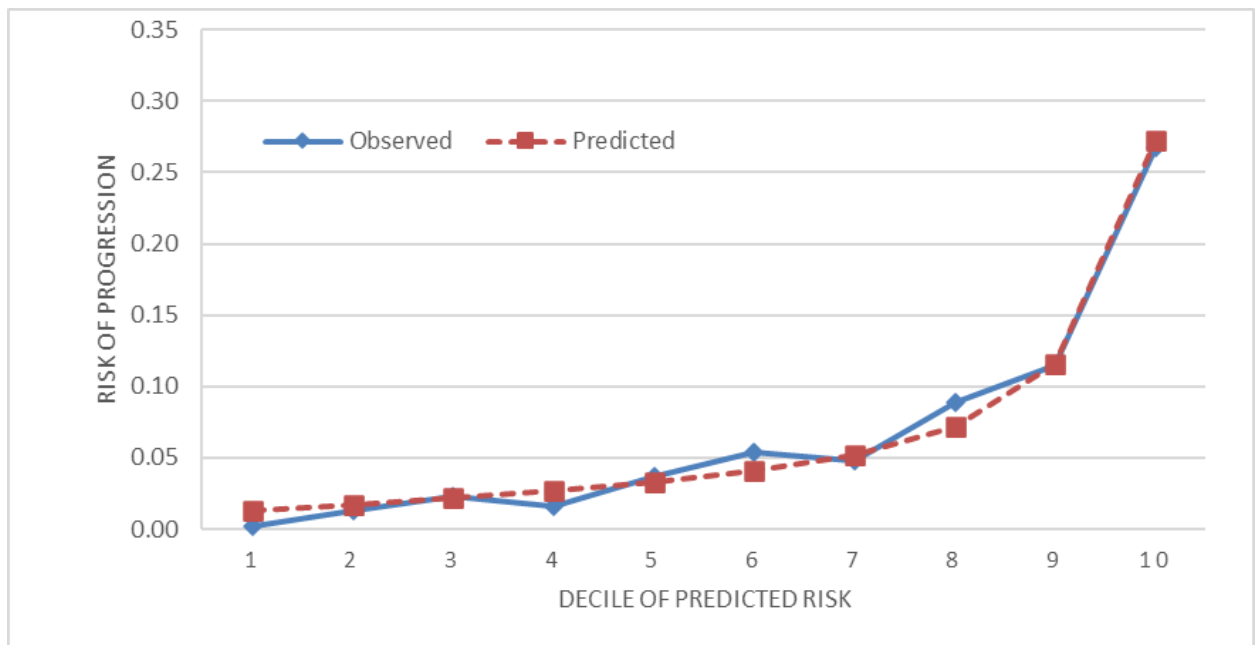


Figure 2b.03:2. Progression EHR Dataset A, with eGFR (n=5,658)

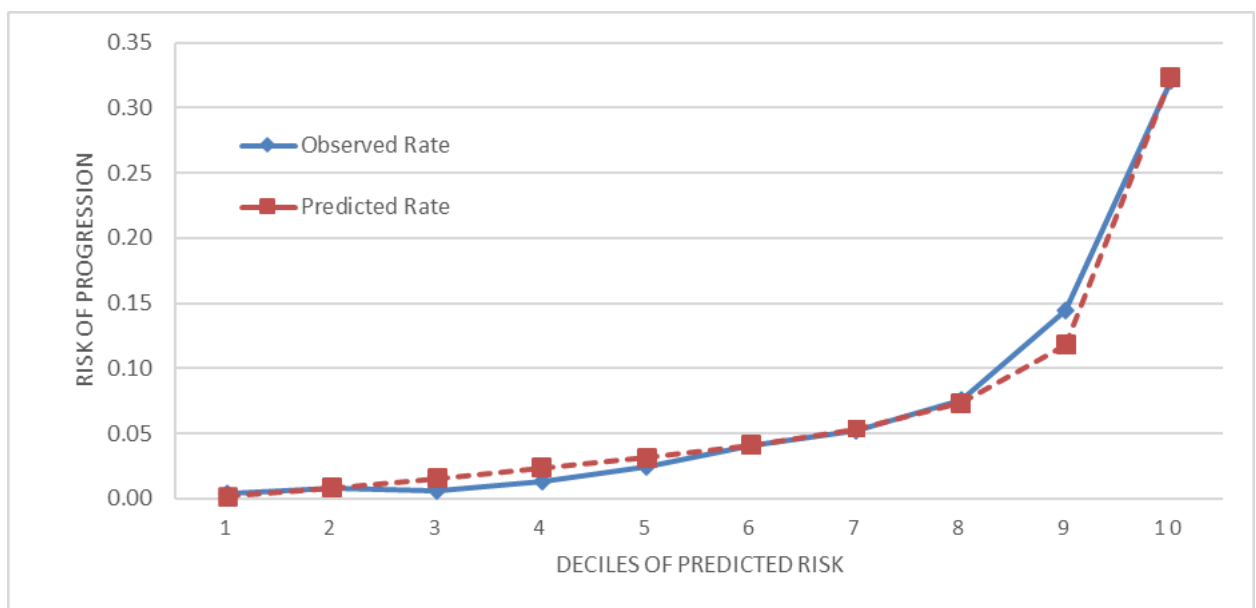


Figure 2b.03:3. Progression EHR Dataset B, without eGFR (n=14,070)

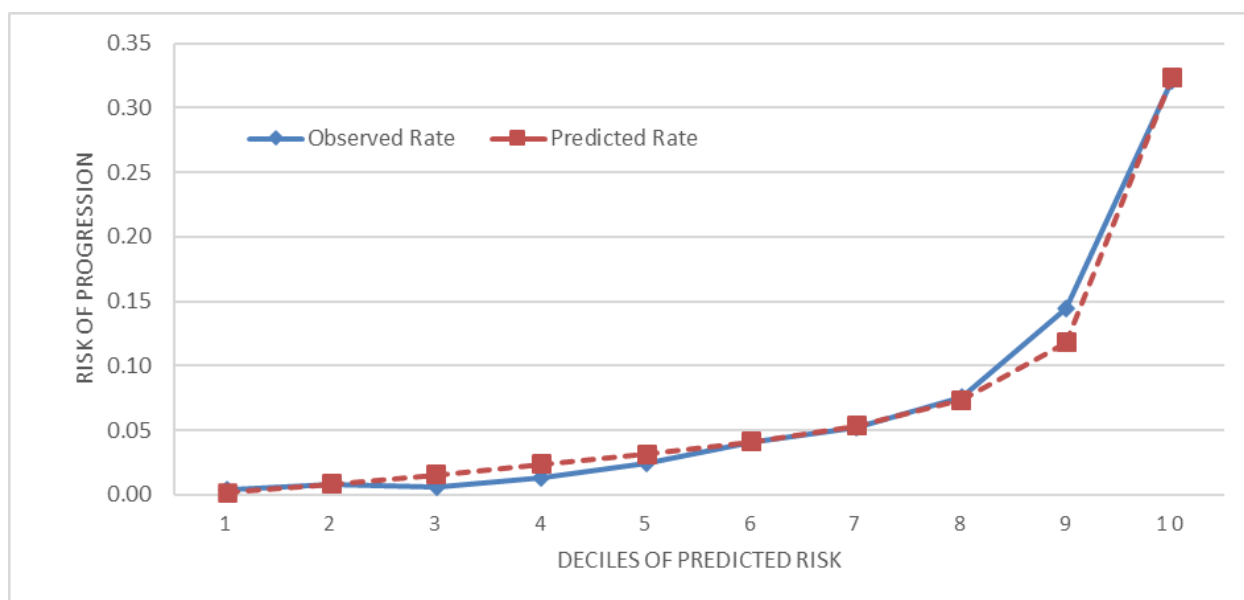


Figure 2b.03:4. Progression EHR Dataset B, with eGFR (n=14,070)

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:

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

Table 2b.03:2. TEP Face Validity Responses

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.

[Response Ends]

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

[Response Begins]

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.

Empiric Validity –Model Validation

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.

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 as disease self-management, this is true of all measures and is ideally mitigated to some extent by provider care, education, and support.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

Examination of provider-level results include measure scores for all providers and those with at least 25 patients, along with their summary statistics such as mean (SD), median (IQR), and the minimum (min) and maximum (max). We are using 25 as an example minimum case count, which aligns with CMS publicly reported outcome measures.

[Response Ends]

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

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

[Response Begins]

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.

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
Volume: Mean (SD)	152.3 (245.7)	216.6 (272.3)
Volume: Median (IQR)	61 (16 – 182)	122 (57 – 266)
Volume: min - max	1 - 2924	25 – 122

Table 2b.06. Measure Performance Statistics (RSR) for All Providers and Providers with 25 or More Patients, Progression Development Dataset

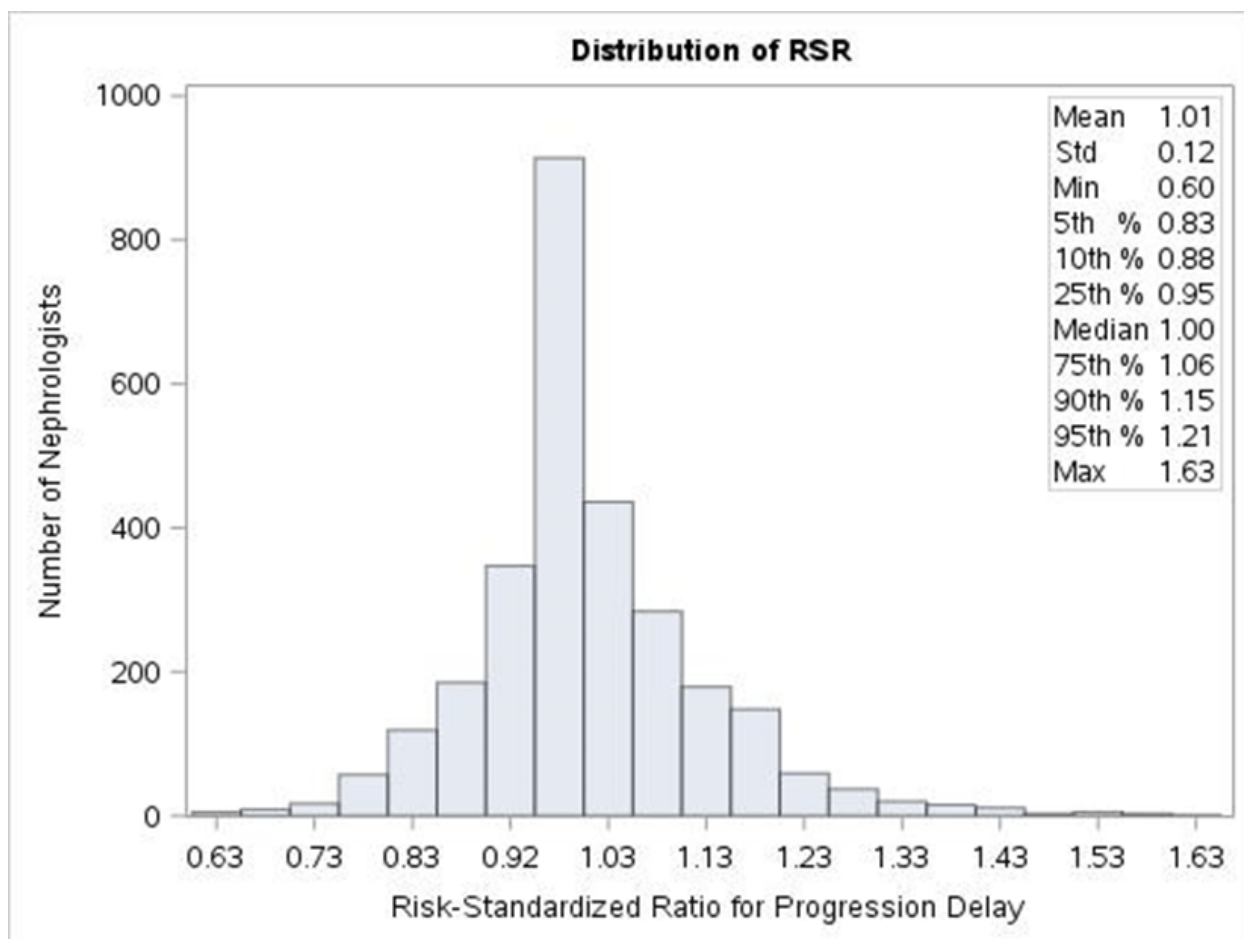


Figure 2b.06:1. Distribution of RSR, all nephrology practices (n=2,854)

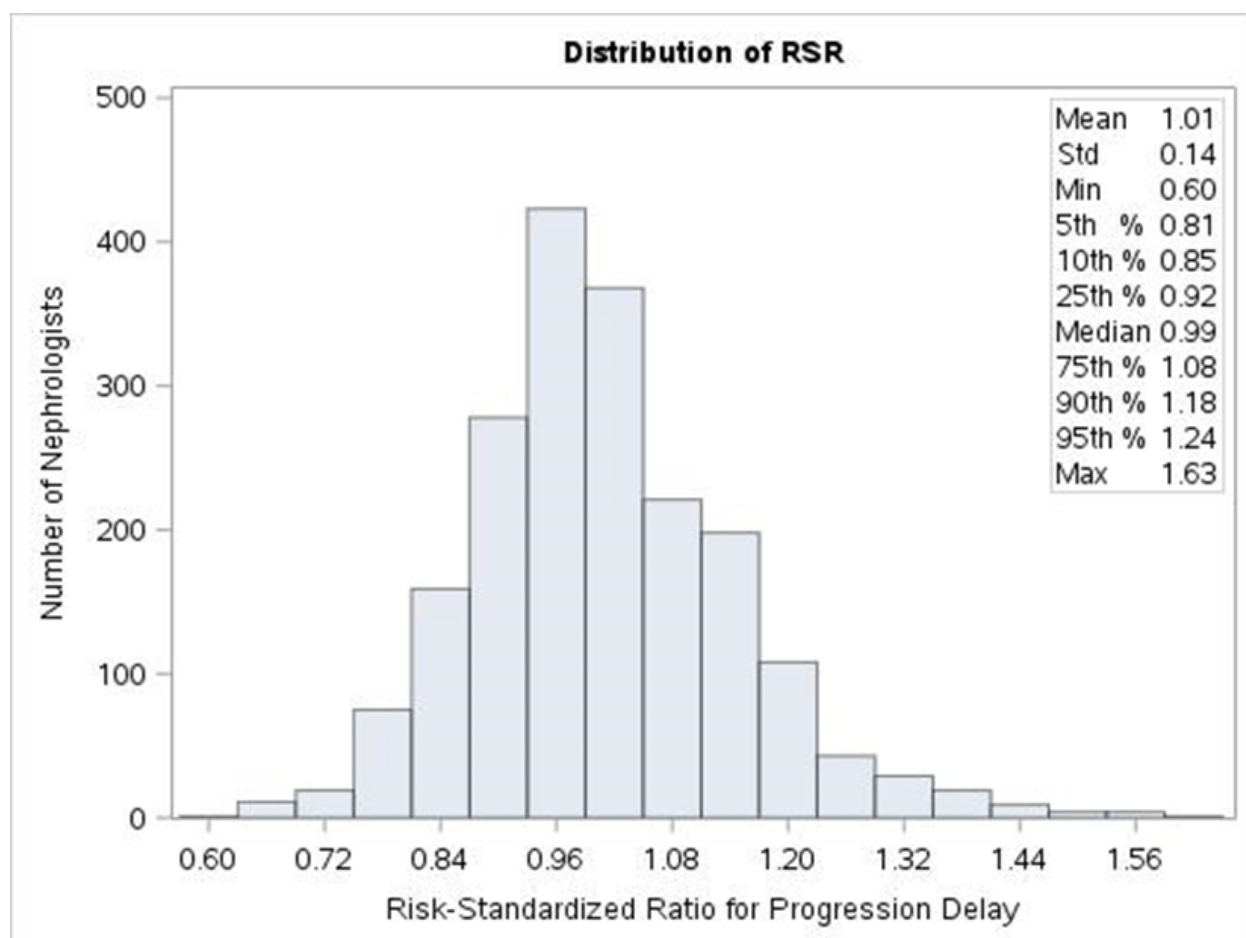


Figure 2b.06:2. Distribution of RSR, nephrology practices with 25+ patients (n=1,970)

[Response Ends]

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

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

[Response Begins]

As shown 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. 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.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or

differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

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.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

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 status¹. 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.

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It

does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

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

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

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

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

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.

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

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.

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

The 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.

[Response Ends]

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

[Response Begins]

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

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

43 clinical risk factors (including age)

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

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 a Cox hazard model with the selected risk factors.

[Response Ends]

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

[Response Begins]

Not applicable; the measure is risk-adjusted.

[Response Ends]

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

[Response Begins]

Published literature

Internal data analysis

[Response Ends]

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

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

[Response Begins]

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[1] 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. Conceptual models should illustrate 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.

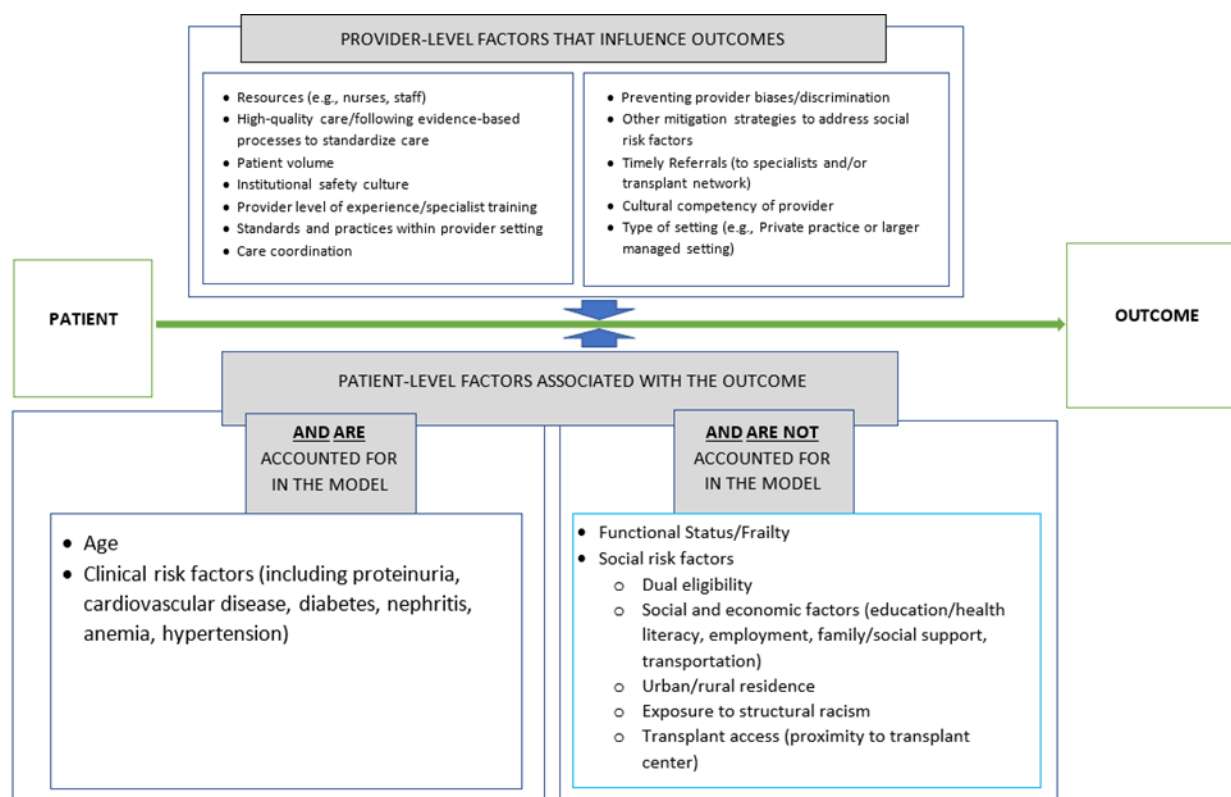


Figure 2b.23. Conceptual Model

A review of literature highlights the well-established disparities in CKD outcomes including progression from CKD to ESRD, for patients with social risk factors, including race/ethnicity as well as socioeconomic status [2-6]. 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 ESRD were higher among patients living in lower-income neighborhoods for all race/ethnicity groups, disparities by race/ethnicity were still apparent within income categories [3].

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 are also a driver of 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 [7-11]. 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). 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 [12].

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 and 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 (AHRQ) 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.

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

Table 2b.23. Candidate Social Risk Factors

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.

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. National Quality Forum: Risk Adjustment Technical Guidance Report. August 31, 2022. Available at <https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=97639>. Accessed January 5, 2023
2. Nicholas SB, Kalantar-Zadeh K, Norris KC: Socioeconomic disparities in chronic kidney disease. *Adv Chronic Kidney Dis* 22: 6–15 (2015).
3. 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.
4. Patzer RE, McClellan WM: Influence of race, ethnicity and socioeconomic status on kidney disease. *Nat Rev Nephrol* 8: 533–541 (2012).
5. 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
6. 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).
7. 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).
8. 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).
9. 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).
10. 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).
11. 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).
12. 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

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

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.

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)

#3753 Delay in Progression of Chronic Kidney Disease (CKD) Measure, Submission Last Updated: Jan 09, 2023

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)
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)

Risk Variable	Percentage	Parameter Estimates (Standard Error)	Hazard Ratio (95% Confidence Interval)
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)

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)

[Response Ends]

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

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

[Response Begins]

Social risk factor testing included:

1. Examining the distribution of SRFs among providers (Table 2b.25:1);
2. Examined bivariate (unadjusted) and multivariate (risk-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.

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%)

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

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.

Social Risk Factors	Unadjusted (Bivariate) estimate (SE)	Unadjusted Hazard Ratio (95% CI)	Adjusted (Multivariate) estimate (SE)	Adjusted Hazard Ratio (95% CI)	C-statistic[BK2] (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
Urban	0.035 (0.017)	1.036 (1.001-1.071)	0.039 (0.017)	1.039 (1.005-1.075)	0.792

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)

*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 (5th quintile).

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

Table 2b.25:3. Correlation coefficients between RSR (provider-level score) & proportion of patients in disadvantaged group by quintiles of SRF.

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

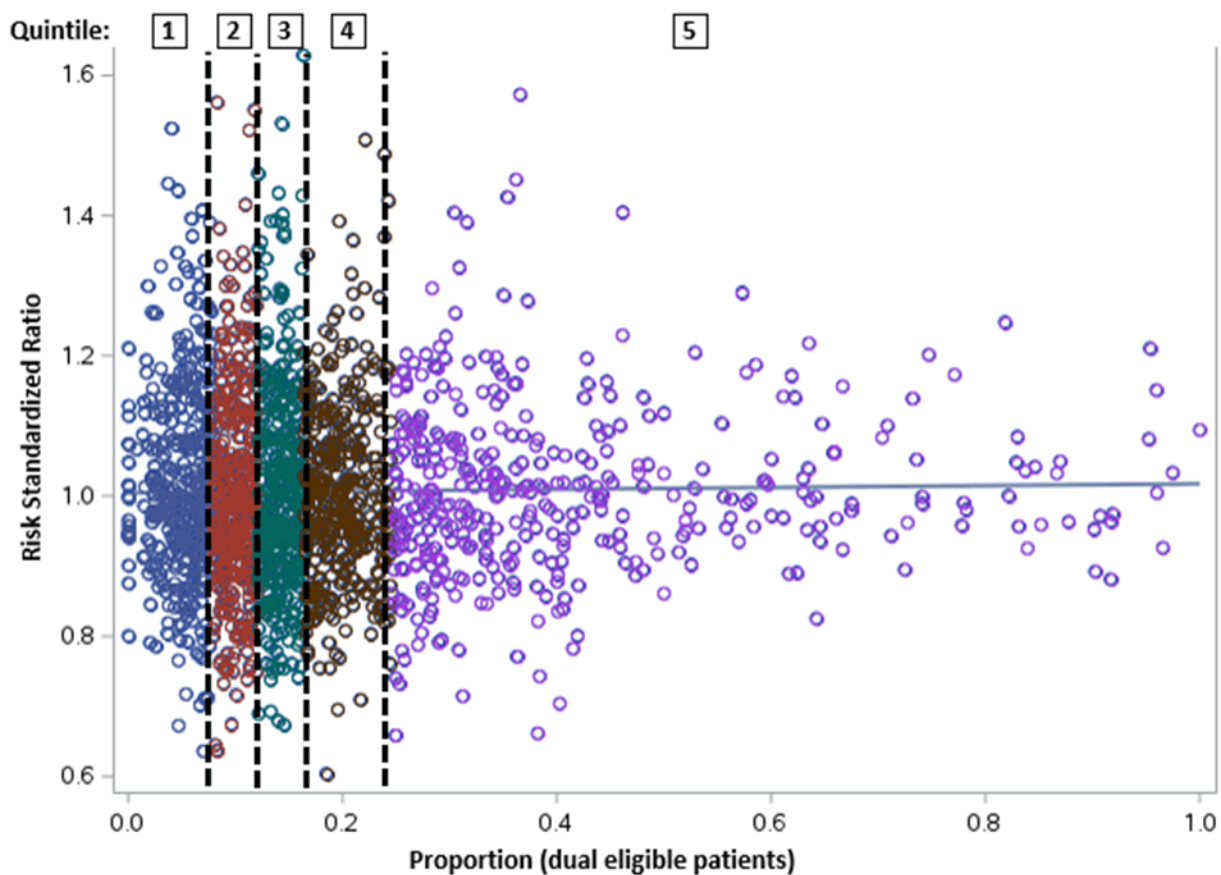


Figure 2b.25:1. Risk Standardized Ratio (RSR) by Percent Dual Eligible

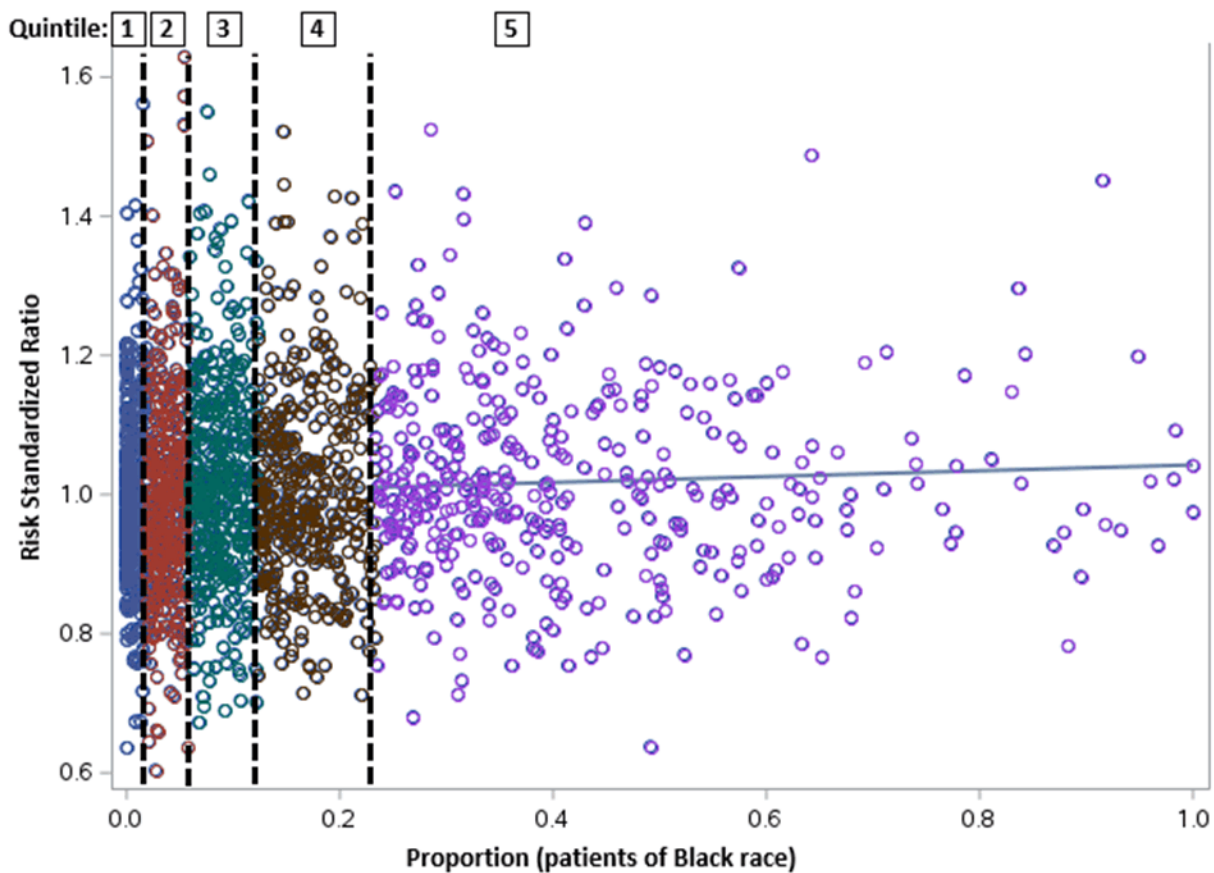


Figure 2b.25:2. RSR by Percent Patients of Black Race

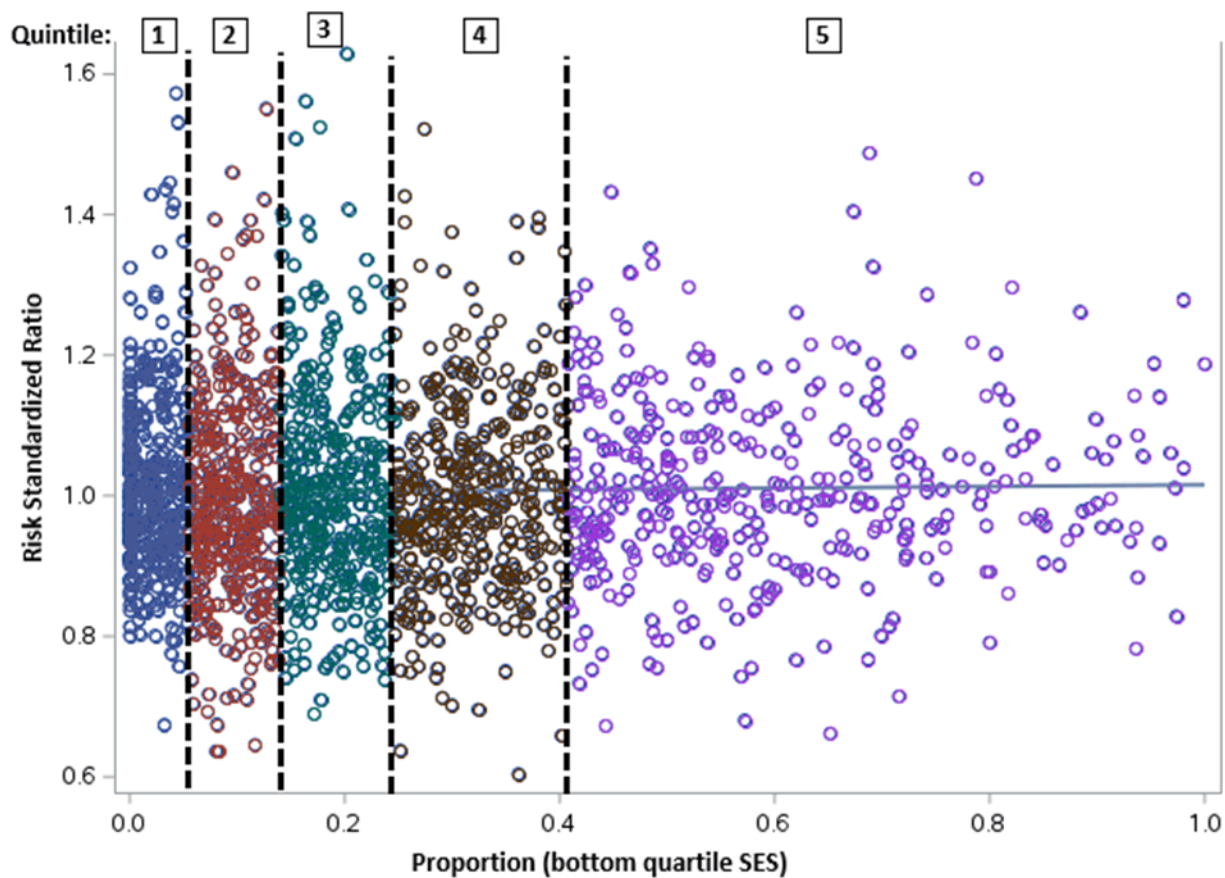


Figure 2b.25:3. RSR by Percent Patients of Low (bottom-quartile SES Index) SES

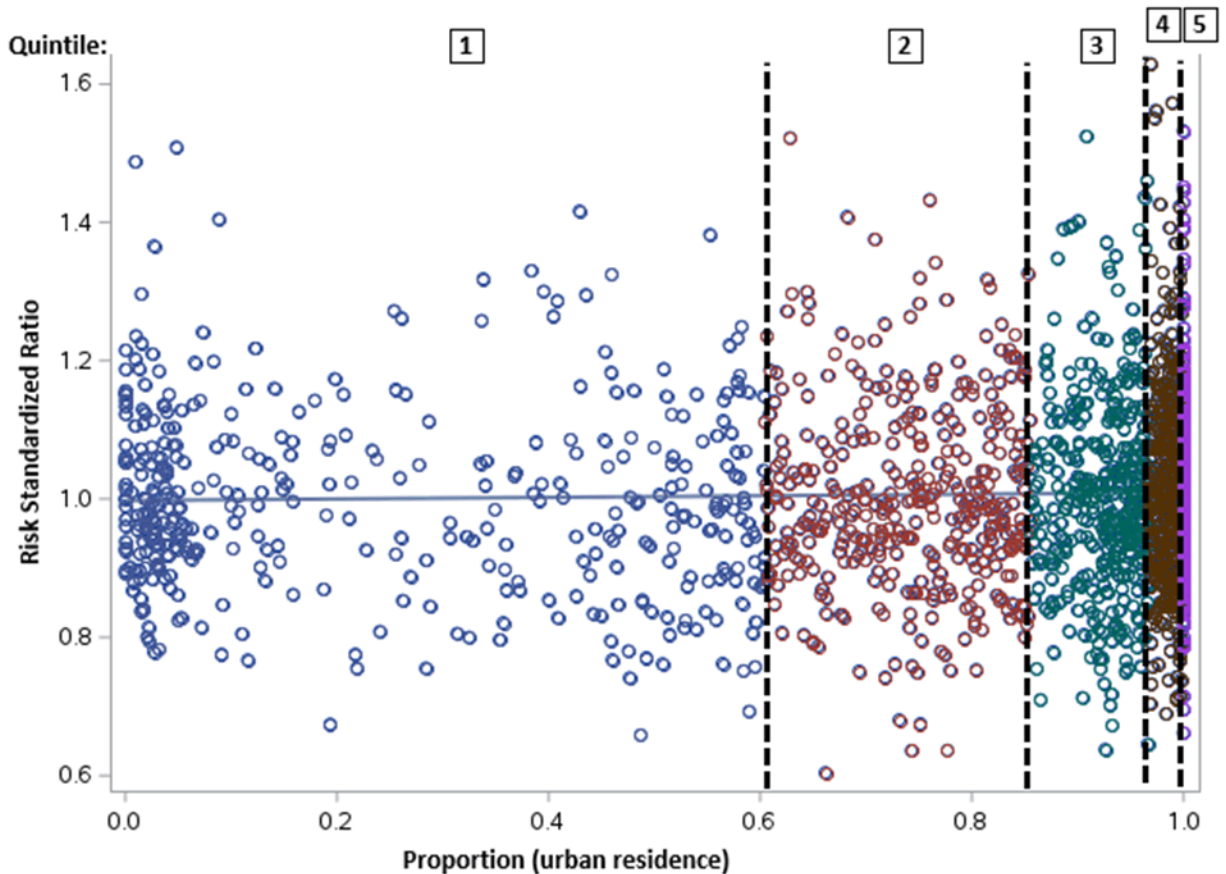


Figure 2b.25:4. RSR by Percent Patients of Urban Residence

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 SRF we tested (except for Black race) 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.

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

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 performance¹.

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.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

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.

Model Performance Statistic	Development Sample	Validation Sample
Number of Patients	217,382	217,382
Progression Delay Risk	4.714	4.745
Calibration (γ_0 , γ_1)	(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

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

[Response Ends]

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

[Response Begins]

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).

[Response Ends]

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

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

[Response Begins]

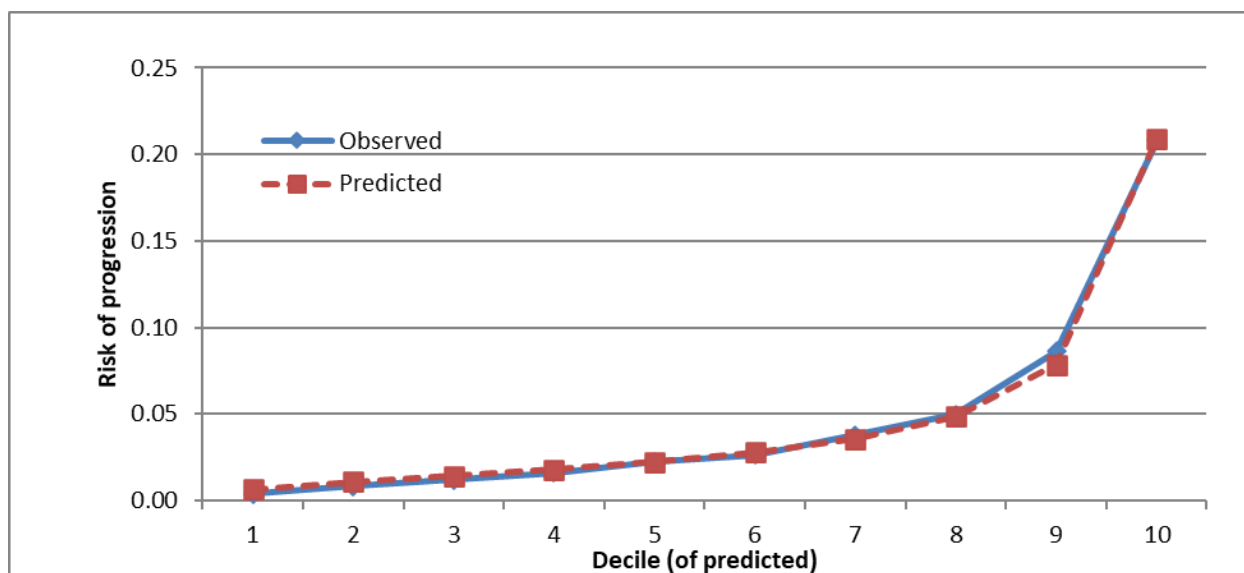


Figure 2b.29:1. Observed vs. predicted progression risk by decile of predicted (derivation sample)

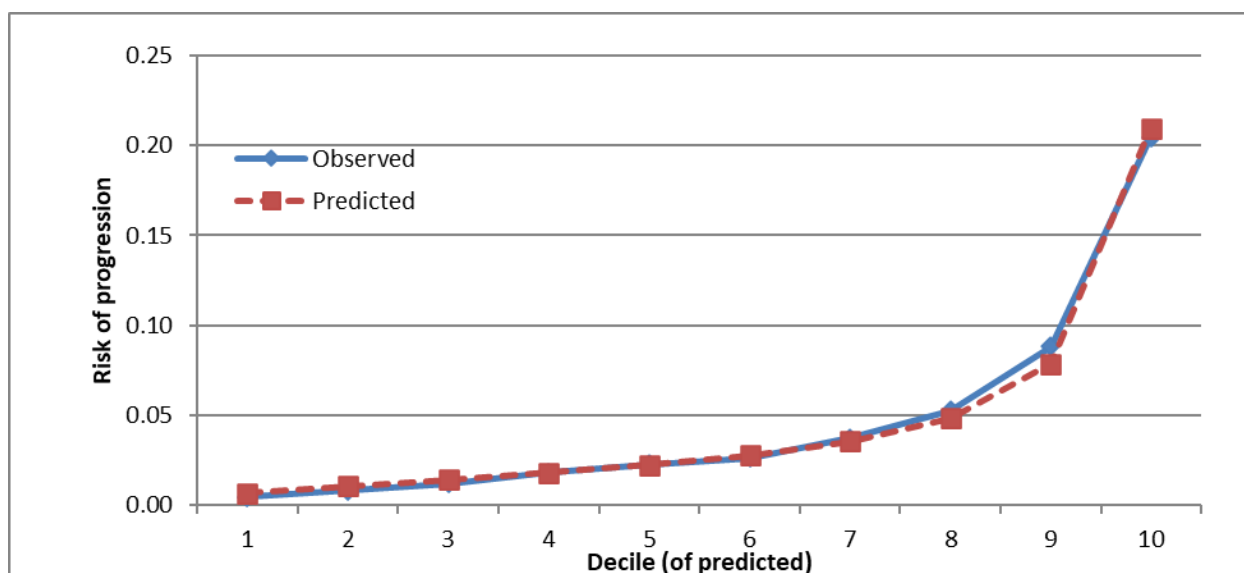


Figure 2b.29:2. Observed vs. predicted progression risk by decile of predicted (validation sample)

[Response Ends]

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

[Response Begins]

Not applicable; measure is not stratified.

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

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

[Response Begins]

Interpretation:

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/Over-fitting Statistics (γ_0 , γ_1):

If the γ_0 in the validation samples are substantially far from zero and the γ_1 is substantially far from 1, there is potential evidence of over-fitting. The validation sample calibration values of $\gamma_0 = -0.026$ and $\gamma_1 = 0.987$ 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.

[Response Ends]

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

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

[Response Begins]

No additional testing.

[Response Ends]

3. Feasibility

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]

[Response Ends]

4. Usability and Use

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

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

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

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

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

NOTE: If there are no related measures, please select N/A.

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

NOTE: If there are no competing measures, please select N/A.

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

Provide analyses when possible.

[Response Begins]

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

Available in attached file

Attachment: 3753_Methods_CKDProgDelay_01052023.pdf

Contact Information

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

Measure Steward Point of Contact: Day, Tim, timothy.day@cms.hhs.gov

Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

Measure Developer Point(s) of Contact: Bagshaw, Kyle, kyle.bagshaw@yale.edu

Peter, Doris, doris.peter@yale.edu

Additional Information

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

[Response Begins]

Available in attached file

[Response Ends]

Attachment: 3753_Methods_CKDPProgDelay_01052023.pdf

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

Describe the members' role in measure development.

[Response Begins]

Name, Credentials	Professional Role	Organization, Location
Steven Spencer ³ , 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 ^{[1]^{2,3}} , MSN, MBA-HCM, RN, CNN, CLNC	Renal Clinic Nurse	Atlanta VA Medical Center-Renal Clinic, Decatur, GA
Andrew “Drew” Wall ^{2,3} , 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 ^{2,3,4}	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
Titte R. Srinivas ² , 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 ³ , MBA	Patient; President	American Association of Kidney Patients, Bowie, MD

Name, Credentials	Professional Role	Organization, Location
Daniel E Weiner ³ MD, MS	Associate Professor of Medicine and Nephrologist; Medical Director of Clinical Research	Tufts Medical Center; and Dialysis Clinic Inc, Boston, MA
Jessie Pavlinac ³ MS, RDN-AP, CSR, LD, FAND	Adjunct Senior Instructor	Oregon Health & Science University; School of Medicine, Portland, OR
Scherly Leon ³ MD, MS, MPH	Medical Director; Nephrologist	Atlantic Dialysis, New York, NY
Milda Saunders ^{2,3} MD, MPH	Assistant Professor of Medicine; General Internist	University of Chicago Medicine, Chicago, IL

Table A2. TEP Membership List

1. TEP member did not participate in the June 2020 meeting.
2. TEP member did not participate in June 2021 meeting.
3. TEP member did not participate in July 2022 meeting.
4. TEP member dropped from TEP.

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]