NATIONAL QUALITY FORUM

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 #: 1425
Corresponding Measures:
De.2. **Measure Title:** Measurement of nPCR for Pediatric Hemodialysis Patients
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** Percentage of patient months of pediatric (< 18 years old) in-center hemodialysis patients (irrespective of frequency of dialysis) with documented monthly nPCR measurements.
1b.1. **Developer Rationale:** For in-center hemodialysis patients, nPCR provides an estimate of dietary protein intake, which has been shown to provide additional information to spKt/V. Studies have shown that in adolescent patients who achieved target spKt/V levels, nPCR was associated with nutritional status. Furthermore, there is evidence that nPCR < 1 gram/kg/day is predictive of malnutrition and sustained weight loss among adolescent patients.

S.4. **Numerator Statement:** Number of patient months in the denominator with monthly nPCR measurements.
S.6. **Denominator Statement:** Number of all patient months for pediatric (less than 18 years old) in-center hemodialysis patients (irrespective of frequency of dialysis).
S.8. **Denominator Exclusions:** Exclusions that are implicit in the denominator definition include adult patients (greater than or equal to 18 years of age), all patients who have not been in the facility for the entire reporting month, and all home hemodialysis and peritoneal dialysis patients. There are no additional exclusions for this measure.

De.1. **Measure Type:** Process
S.17. **Data Source:** Claims, Registry Data
S.20. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Aug 16, 2011 **Most Recent Endorsement Date:** Oct 02, 2015

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

1. Evidence, Performance Gap, Priority – 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.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
1425_Evidence.docx
1a.1 <u>For Maintenance of Endorsement:</u> Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.
For in-center hemodialysis patients, nPCR provides an estimate of dietary protein intake, which has been shown to provide additional information to spKt/V. Studies have shown that in adolescent patients who achieved target spKt/V levels, nPCR was associated with nutritional status. Furthermore, there is evidence that nPCR < 1 gram/kg/day is predictive of malnutrition and sustained weight loss among adolescent patients.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, 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 (4b1) under Usability and Use.*
Among the 29 facilities that have at least 11 eligible pediatric patients, we generated the following statistics of their performance scores using the January – December 2017 (i.e., calendar year 2017) CROWNWeb clinical data: mean (SD) = 76.64% (32.5%), min = 0%, max = 99.3%, 25th percentile = 75.8%, 50th percentile = 90.8%, and 75th percentile = 94.1%.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, 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.**
N/A

1b.4. **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.** *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) 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 (4b1) under Usability and Use.*
Disparity analyses were performed among the entire eligible pediatric population (n=504) to examine the difference in performance scores by sex, race and ethnicity.

In particular, for each facility, the percent of patient-months by demographic group (sex, race, ethnicity, age) was calculated. Then, the facilities were divided into tertiles (Q1-Q3) based on the percentage of patient-months in the particular demographic category (i.e., a facility with percentage of females similar to the national median will be included in tertile 3). The top 33.3% of facilities in terms of rank, based on the percentages of females, were classified as Q3, while the bottom 33.3% of facilities were classified as Q1. Average (mean) performance for the measure was calculated for each tertile, and the means were examined for trend across tertiles (Q1-Q3).

The mean performance scores for percent of patient-months with a nPCR measurement in each tertile, by demographic group, are presented below. Males, non-Black, non-White, non-Hispanic are the respective reference categories. Based on the small sample size, we do not believe that the following results suggest a meaningful trend.

Range of Facility Level Tertiles by Population Group (Tertile 1-3):
Females (Q1=66.1%, Q2=70.8%, Q3=55.5%)
Black (Q1=65.5, Q2=64.2%, Q3= 60.1%)
White (Q1=63.1%, Q2=64.7%, Q3=63.5%)
Hispanic (Q1=60.2%, Q2=70.5%, Q3=69.0%)

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, 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 1b.4
N/A

2. Reliability and Validity—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.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*
Renal, Renal : End Stage Renal Disease (ESRD)

De.6. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Children, Populations at Risk

S.1. Measure-specific Web Page *(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.)*
N/A

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
For the Spring 2019 Maintenance submission: While the interdialytic time element is needed to calculate nPCR, CROWNWeb currently does not allow collection of that data element, therefore the measure currently does not require reporting of that variable at this time. We plan to revise the measure via an annual update once data collection resumes in CROWNWeb.

S.4. **Numerator Statement** *(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.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Number of patient months in the denominator with monthly nPCR measurements.

S.5. **Numerator Details** *(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 S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The number of patients in the study month where (1) nPCR value and the date the nPCR were collected and reported or (2) the following 7 components used to calculate nPCR (BUN pre-dialysis, BUN post-dialysis, pre-dialysis weight, pre-dialysis weight unit of measure, post-dialysis weight, post-dialysis weight unit of measure, and delivered minutes of BUN hemodialysis session), and the date of collection are all reported.

Note: Interdialytic time is also needed to calculate nPCR; however CROWNWeb currently does not allow collection of that data element therefore the measure does not require reporting of this variable.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of all patient months for pediatric (less than 18 years old) in-center hemodialysis patients (irrespective of frequency of dialysis).

S.7. **Denominator Details** *(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 S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
A treatment history file is the data source for the denominator calculation used for the analyses supporting this submission. This file provides a complete history of the status, location, and dialysis treatment modality of an ESRD patient from the date of the first ESRD service until the patient dies or the data collection cutoff date is reached. For each patient, a new record is created each time he/she changes facility or treatment modality. Each record represents a time period associated with a specific modality and dialysis facility. CROWNWeb is the primary basis for placing patients at dialysis facilities and dialysis claims are used as an additional source of information in certain situations. Information regarding first ESRD service date, death, and transplant is obtained from CROWNWeb (including the CMS Medical Evidence Form (Form CMS-2728) and the Death Notification Form (Form CMS-2746)) and Medicare claims, as well as the Organ Procurement and Transplant Network (OPTN).

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclusions that are implicit in the denominator definition include adult patients (greater than or equal to 18 years of age), all patients who have not been in the facility for the entire reporting month, and all home hemodialysis and peritoneal dialysis patients. There are no additional exclusions for this measure.

S.9. **Denominator Exclusion Details** *(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 S.2b.)*
N/A

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including 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 with at S.2b.)*
N/A

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
To be included in the denominator for a particular month, the patient must be on in-center hemodialysis for the entire month, must be < 18 years old at the beginning of the month, and must be assigned to that facility for the entire month. An individual patient may contribute up to 12 patient-months per year.

The numerator counts the number of patients in the study month where (1) nPCR value and the date the nPCR were collected and reported or (2) the components that allow calculation of nPCR (BUN pre-dialysis, BUN post-dialysis, pre-dialysis weight, pre-dialysis weight unit of measure, post-dialysis weight, post-dialysis weight unit of measure, and delivered minutes of BUN hemodialysis Session) and the date of collection are all known.

Note: Interdialytic time is also needed to calculate nPCR; however, CROWNWeb currently does not allow collection of that data element, therefore the measure does not require reporting of that variable.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.
N/A

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
Specify calculation of response rates to be reported with performance measure results.
N/A

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims, Registry Data

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.
CROWNWeb and Medicare claims.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: Dialysis Facility

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2. Validity – See attached Measure Testing Submission Form
1425_testing_01072019-636824726937338424.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
Yes

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
Yes

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.
No - This measure is not risk-adjusted

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.

3a. Byproduct of Care Processes
For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition
If other:

3b. Electronic Sources
The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.
ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **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 other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. 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.
IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.
Data collection is accomplished via CROWNWeb, a web-based and electronic batch submission platform maintained and operated by CMS contractors. Measures reported on DFC are reviewed on a regular basis by dialysis facility providers and rare instances of inaccurate or missing data are present based on comments reported in the DFC ticketing system.

3c.2. Describe 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)*.
N/A

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.

4a. Accountability and Transparency

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

4.1. Current and Planned Use
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 performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Not in use	Public Reporting Dialysis Facility Compare http://www.medicare.gov/dialysisfacilitycompare/

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

Dialysis Facility Compare
Purpose: Dialysis Facility Compare helps patients find detailed information about Medicare-certified dialysis facilities. They can compare the services and the quality of care that facilities provide.
Geographic area: United States
Number of accountable entities: All Medicare-certified dialysis facilities that are eligible for the measure, and have at least 11 patients (due to public reporting requirements). For the most recent update to Dialysis Facility Compare (January 2019), 8 facilities

had a score reported.
Patients included: All patients who meet the inclusion criteria to be included in the measure.

4a1.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*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.*)
N/A

4a2.1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
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.
Results of this measure are currently reported on Dialysis Facility Compare. All Medicare-certified dialysis facilities are eligible for reporting (approximately 7,000 dialysis facilities). The program has a helpdesk and supporting documentation available to assist with interpretation of the measure results.

4a2.1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
For DFC, the results are first reported to facilities via a closed preview period, where facilities can review their data prior to each of the quarterly updates of the public facing Dialysis Facility Compare website. These preview reports are posted on dialysisdata.org, where facilities can also find a detailed Guide to the Quarterly Dialysis Facility Compare Reports and other supporting documentation. Facilities can submit comments/questions about their results at any time, and can request patient lists for their facilities during the specified preview periods.

A measures manual that describes the calculations for DFC in detail is published on the CMS website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/06_MeasuringQuality.html

4a2.2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
For DFC, feedback can be provided any time through contacting the dialysisdata.org helpdesk. Preview periods allow for specific times for facilities review and comment on measure calculations, and provide an opportunity to request a patient list.

4a2.2.2. **Summarize the feedback obtained from those being measured.**
We reviewed the comments and questions submitted during the DFC preview periods that have taken place since the last maintenance (2016-present). We have received only a handful of clarification questions since the measure was added to DFC, likely due to the very small number of facilities affected.

4a2.2.3. **Summarize the feedback obtained from other users**
N/A

4a2.3. **Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
The revisions to this measure are based on data availability and not direct feedback. As described above, we have not received much in the way of feedback on this measure, likely due to the small number of facilities that have their results publicly reported.

Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results

could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. **Refer to data provided in 1b 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, what are the reasons? 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.
Given that small scale observational studies have shown an association between nPCR and nutritional status among malnourished adolescent patients who achieved target spKt/V levels, we would expect that public reporting of this measure may engage facilities to better monitor the nutrition status of their pediatric patients. CY 2017 was the first year of public reporting; this may be too short of a time frame to observe meaningful trends, particularly because of the small number of facilities for which the measure is calculated.

Q1: N = 29, Mean = 75.59%, Std Dev =32.25%, Min = 0.0%, Max = 100.0%
Q2: N = 29, Mean = 77.07%, Std Dev =32.88%, Min = 0.0%, Max = 100.0%
Q3: N = 29, Mean = 78.84%, Std Dev =33.90%, Min = 0.0%, Max = 100.0%
Q4: N = 29, Mean = 76.20%, Std Dev =33.41%, Min = 0.0%, Max = 100.0%

4b2. Unintended Consequences

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

4b2.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
None that we are aware of.

4b2.2. **Please explain any unexpected benefits from implementation of this measure.**
None that we are aware of.

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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-
Co.3 **Measure Developer if different from Measure Steward:** University of Michigan Kidney Epidemiology and Cost Center
Co.4 **Point of Contact:** Casey, Parrotte, parrotte@med.umich.edu

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:** 04, 2019
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 04, 2020

Ad.6 **Copyright statement:** N/A
Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** After the submission of the testing attachment on January 7, we noticed a typo in 2b4.1 (Meaningful Differences). The description of the analysis mentions the wrong event (hypercalcemia), which was included in error. The description of the analysis performed is otherwise accurate.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1425
Measure Title: Measurement of nPCR for Pediatric Hemodialysis Patients
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 4/2/2019

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one

step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☐ Outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: measurement of nPCR

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL Diagram or 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.

In the 2006 Kidney Disease Outcomes Quality Initiative (KDOQI) Guidelines, Clinical Practice Guideline for pediatric hemodialysis adequacy (Guideline 8.2.2) specifies nPCR should be measured monthly. The 2008 KDOQI Clinical Practice Guideline Update for nutrition in children with CKD Recommendation 1.1 states that the nutritional status and growth of all children with CKD stages 2-5 be evaluated on a periodic basis. Recommendation 1.2 states that nPCR should be evaluated in hemodialyzed adolescents. Small scale observational studies have shown an association between nPCR and nutritional status among malnourished adolescent patients who achieved target spKt/V levels [1,2]. Additionally, in adolescent patients, nPCR levels < 1 gram/kg/day were found to be an earlier and more sensitive marker than serum albumin levels in predicting malnutrition and sustained weight loss [3].

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

N/A

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

N/A

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

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

X Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	National Kidney Foundation. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for 2006 Updates: Hemodialysis Adequacy, Peritoneal Dialysis Adequacy and Vascular Access. Am J Kidney Dis 48:S1-S322, 2006 (suppl 1). http://www.kidney.org/PROFESSIONALS/kdoqi/guideline_upHD_PD_VA/index.htm
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	8.2.2 Assessment of nutrition status is an essential component of HD adequacy measurement. nPCR should be measured monthly by using either formal urea kinetic modeling or algebraic approximation. (B) 2008 KDOQI CPR RECOMMENDATION 1: EVALUATION OF GROWTH AND NUTRITIONAL STATUS 1.1 The nutritional status and growth of all children with CKD stages 2 to 5 and 5D should be evaluated on a periodic basis. (A) 1.2 The following parameters of nutritional status and growth should be considered in combination for evaluation in children with CKD stages 2 to 5 and 5D. (B) i. Dietary intake (3-day diet record or three 24-hour dietary recalls)

	ii. Length- or height-for-age percentile or standard deviation score(SDS) iii. Length or height velocity-for-age percentile or SDS iv. Estimated dry weight and weight-for-age percentile or SDS v. BMI-for-height-age percentile or SDS vi. Head circumference-for-age percentile or SDS (=3 years old only) Normalized protein catabolic rate (nPCR) in hemodialyzed adolescents with CKD stage 5D.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	N/A
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	The 2006 KDOQI Guideline 8.2.2 rating strength grade is 'B'. The recommendation for Grade B guidelines states 'It is recommended that clinicians routinely follow the guideline for eligible patients. There is moderate to strong evidence that the practice improves health outcomes.'
Provide all other grades and definitions from the recommendation grading system	The rating system defined in the KDOQI Guidelines was used to grade the strength of the Guideline recommendation. KDOQI defined grades as follows: Grade A: It is strongly recommended that clinicians routinely follow the guideline for eligible patients. There is strong evidence that the practice improves health outcomes. Grade B: It is recommended that clinicians routinely follow the guideline for eligible patients. There is moderately strong evidence that the practice improves health outcomes. Grade CPR: It is recommended that clinicians consider following the guideline for eligible patients. This recommendation is based on either weak evidence or on the opinions of the Work Group and reviewers that the practice might improve health outcomes.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	N/A

Estimates of benefit and consistency across studies	N/A
What harms were identified?	N/A
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	In May 2014, an additional literature search was performed. A recent comprehensive review on the subject [4] is included in the citations below as a result of that search. This review continues to be supportive of the concept of monitoring nPCR as part of evaluation of Protein Energy Wasting (PEW) in children/adolescents on dialysis. 1. Goldstein, Baronette, et al. nPCR assessment and IDPN treatment of malnutrition in pediatric hemodialysis patients. Pediatric Nephrology (2002) 17:531-534. 2. Orellana P, Juarez-Congelosi M, Goldstein SL. Intradialytic parenteral nutrition treatment and biochemical marker assessment for malnutrition in adolescent maintenance hemodialysis patients. J Ren Nutrition 2005 Jul;15(3):312-7. 3. Juarez-Congelosi M, Orellana P, Goldstein SL: Normalized protein catabolic rate versus serum albumin as a nutrition status marker in pediatric patients receiving hemodialysis. J Ren Nutr 17:269-274, 2007. 4. Mastrangelo A, Paglialonga F, Edefonti A. Assessment of nutritional status in children with chronic kidney disease and on dialysis. Pediatr Nephrol. 2014 Aug;29(8):1349-58. doi: 10.1007/s00467-013-2612-7. Epub 2013 Sep 5.

1a.4 OTHER SOURCE OF EVIDENCE

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

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.
N/A

1a.4.2 What process was used to identify the evidence?

N/A

1a.4.3. Provide the citation(s) for the evidence.
N/A

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 1425
Measure Title: Measurement of nPCR for Pediatric Hemodialysis Patients
Date of Submission: 1/7/2019
Type of Measure:

☐ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☒ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

Instructions

- 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.
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures**, section **2b3** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at 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 version 7.1 of the Measure Testing Attachment.

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.

2a2. Reliability testing [10] 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 [11] 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; [12]

AND
If patient preference (e.g., informed decisionmaking) 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). [13]

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

OR
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

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

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

Notes
10. 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).
11. 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.
12. 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.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.
15. 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.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. 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*).

CROWNWeb and Medicare Claims Data from January 2013 to December 2013

For the Spring 2019 maintenance submission, CROWNWeb and Medicare Claims Data from January 2017 to December 2017.

1.3. What are the dates of the data used in testing? January 2013 to December 2013

For the Spring 2019 maintenance submission, January 2017 to December 2017

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were 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*)

The measured entities used in testing and analysis include reported nPCR and the necessary data elements needed for calculating nPCR for 455 in-center hemodialysis (ICH) pediatric patients from 30 dialysis facilities with at least 11 eligible pediatric patients across all regions of the United States.

Public reporting of this measure on DFC or in the ESRD QIP would be restricted to facilities with at least 11 eligible patients for the measure. We have applied this restriction to all the reliability and validity testing reported here.

Facilities vary in size, and include anywhere from 11 to 28 eligible ICH pediatric patients. The data elements include "nPCR" or the combination of "Kt/V hemodialysis collection date", "BUN pre-dialysis", "BUN post-dialysis", "pre-dialysis weight", "pre-dialysis weight unit of measure", "post-dialysis weight", "post-dialysis weight unit of measure", "delivered minutes of BUN hemodialysis session", and "interdialytic time."

For the Spring 2019 maintenance submission, the measured entities used in testing and analysis include reported nPCR and the necessary data elements needed for calculation of nPCR. There are 511 in-center hemodialysis (ICH) pediatric patients from 29 dialysis facilities that have had at least 11 eligible pediatric patients across all regions of the United States.

Public reporting of this measure on DFC or in the ESRD QIP would be restricted to facilities with at least 11 eligible patients in order for the measure to comply with restrictions on reporting of potentially patient identifiable information related to small cell size. We have applied this restriction to all the reliability and validity testing reported here.

Facilities vary in size, and include anywhere from 11 to 42 eligible ICH pediatric patients.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Testing was performed on all Medicare and non-Medicare pediatric, ICH patients available in CROWNWeb from 2013. The sample included 455 patients from 225 facilities. The table below shows the number and percent of pediatric ICH patients by race, sex, and Hispanic ethnicity.

Race	Frequency	Percent
Asian	23	5.05%
Black	147	32.31%

White	274	60.22%
Native American	5	1.10%
Pacific Islander	4	0.88%
Mid East Arabian	1	0.22%
Other/Multi-racial	1	0.22%
Sex		
Female	202	44.40%
Male	253	55.60%
Ethnicity		
Hispanic	163	35.82%
Non-Hispanic	292	64.18%

For the Spring 2019 maintenance submission, testing was performed on all Medicare and non-Medicare pediatric, ICH patients available in CROWNWeb from 2017. The sample included 511 patients from 29 facilities. The table below shows the number and percent of pediatric ICH patients by race, sex, and Hispanic ethnicity.

Race	**Frequency**	**Percent**
Asian/Pacific Islander	37	7.24%
Black	146	28.57%
White	317	62.04%
American Indian/Alaskan Native	3	0.59%
Other/Multi-racial	8	1.57%
Sex		
Female	237	46.38%
Male	274	53.62%
Ethnicity		
Hispanic	169	33.07%
Non-Hispanic	339	66.34%
**3 missing		

1.7. 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 reported below.

N/A

1.8 What were 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.

N/A

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level 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*)

January 2013 – December 2013 CROWNWeb data were used to calculate the inter-unit reliability (IUR) for the overall 12 months to assess the reliability of this measure. The NQF-recommended approach for determining measure reliability is a one-way analysis of variance (ANOVA), in which the between and within facility variation in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the measure variability that is attributable to the between-facility variance. The yearly based IUR was estimated using a bootstrap approach, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by ANOVA. We note that the method for calculating the IUR was developed for measures that are approximately normally distributed across facilities. Since this measure is not normally distributed, the IUR value should be interpreted with some caution.

For the Spring 2019 maintenance submission, we followed the same methodology as that used in the previous submission for the data from January 2017 – December 2017

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

The overall IUR was 0.985, which indicates that about 98.5% of the variation in the measure can be attributed to the between facility differences and 1.5% to within facility variation.

For the Spring 2019 maintenance submission, the overall IUR was 0.963, which indicates that about 96.3% of the variation in the measure can be attributed to the between facility differences and 3.7% to the within facility variation.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The IUR suggests this measure is reliable. However, since the distribution of performance scores is skewed, the IUR value should be interpreted with some caution.

For the Spring 2019 maintenance submission, the IUR suggests that this measure is reliable. However, since the distribution of performance scores is skewed, the IUR value should be interpreted with some caution.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

- ☐ **Empirical validity testing**
- ☒ **Systematic assessment of face validity of <u>performance measure score</u> 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*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. 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)*

Concurrent validity was used as a method for testing the association between facility percentage of reporting nPCR month and mean nPCR value. Using calendar year 2013 CROWNWeb data, average facility-mean nPCR value was compared between the two groups using a two-sided two-sample t-test. Facilities were then categorized into one of two groups:

1) Facilities with 100% reporting of nPCR among their pediatric patients;

2) Facilities with less than 100% reporting of nPCR among their pediatric patients

nPCR values outside the range of [0.2, 1.8] were excluded.

This measure was also reviewed and approved by a Clinical TEP in 2010.

For the Spring 2019 maintenance submission, we employed methods similar to those used in the previous submission. The current analysis is based on January 2017 – December 2017 data, and employed the following categorization:

1) Facilities with >= 85% reporting of nPCR among their pediatric patients;
2) Facilities with <85% reporting of nPCR among their pediatric patients.

We performed a validity analysis to examine the association between facility percentage of reporting nPCR month and mean nPCR value via the means of two-sample t-test. We would expect that facilities with at least 85% reporting of nPCR among their pediatric patients are likely paying attention to this parameter in their clinical management (i.e., assessment of protein intake) of pediatric dialysis patients.

We also maintain this measure on the basis of face validity, as the measure was reviewed and approved by a Clinical TEP in 2010.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Among facilities with at least 11 eligible pediatric patients with recorded nPCR values, facilities with 100% reporting of recorded nPCR values had a mean serum albumin of 3.77, while facilities with less than 100% reporting of recorded nPCR values had a mean serum albumin of 4.0. Using a t-test, these values were statistically significant (p-value 0.02).

For the Spring 2019 maintenance submission, among facilities with at least 11 eligible pediatric patients and recorded nPCR values, facilities with 85% or higher reporting of recorded nPCR values had a mean nPCR of 0.9974, while facilities with less than 85% reporting of recorded nPCR values had a mean nPCR of 0.6587. According to the t-test (Satterthwaite version), the mean nPCR values of these two groups were not statistically significant (p-value=0.13)

2b1.4. What is 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?*)

These findings are somewhat unexpected, and in the opposite direction of analyses previously conducted. This difference may have resulted from a larger sample utilized for the current analyses (previous analyses were conducted over a limited timeframe). We speculate that the observed findings may have resulted if facilities are more likely to collect necessary data elements for nPCR assessment in patients for which nutritional concerns exist. These results therefore do not necessarily contradict the importance of evaluating nPCR.

Give no evidence that facility-specific nPCR differs by reporting percentage, we found no evidence of invalidity.

In addition to these results, along with the clinical importance of evaluating nPCR, we also propose maintaining the measure based on face validity (as determined by the Technical Expert Panel that initially developed the measure).

_____

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*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*)

N/A

2b2.2. What were 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*)

N/A

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., 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*)

N/A

_____

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

N/A

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?
N/A

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?
N/A

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**
N/A

2b3.5. 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 controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

N/A

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
N/A

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
N/A

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
N/A

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

N/A

_____

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b4.1. 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 related to performance gap in 1b)*

Differences in measure performance were evaluated separately for each facility using patient level analyses. The proportion of patients with yearly based percent of patients with reporting of nPCR was compared between one facility and the overall national distribution, and repeated for each individual facility.

Note that the monthly based measure is a simple average of binary outcomes across individuals in the facility, for which the binary outcome equals to 0 (failure = fail to report nPCR) if the value is missing. The differences in proportions can be compared using Fisher's Exact tests or its normal approximation. The yearly based measure, however, is not a simple average of binary outcomes and we instead used a re-sampling based exact test, with re-sampling generated from the population distribution of the patient level outcomes. More details for the testing method are provided in Appendix. Due to non-symmetric of the measure distributions, one-sided test with significance level 0.025 is used (corresponding to cutoff=0.05 in two-sided test). To calculate the p-value, we assess the probability that the facility would experience a number of events more extreme than that observed if the null hypothesis were true.

For the Spring 2019 submission, we reproduced the significance analysis using data from January – December 2017. We have revised the description of the analysis to be clearer:

Testing was performed on the yearly based performance score. We used a re-sampling based exact test, with re-sampling generated from the population distribution of the patient level outcomes. Note that a one-sided test with significance level 0.025 is used (corresponding to cutoff=0.05 in a two-sided test) due to non-symmetric structure of the measure's distribution. To calculate the p-value, we compute the probability that the facility would experience a number of events (i.e., percentage with hypercalcemia) more extreme than that observed if the null hypothesis were true, with the null hypothesis being that the facility's distribution of hypercalcemia will follow the overall national distribution.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *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*)

Proportion of facilities with significant p-values (0-as expected/better than expected; 1-worse than expected; cutoff=0.025) is shown as follows:

	# of Facilities	Percent of facilities	Median Performance Score
As Expected/Better than Expected	23	76.67%	100.00%
Worse than Expected	7	23.33%	24.49%

For the Spring 2019 maintenance submission, the proportion of facilities with significant p-values (0-as expected/better than expected; 1-worse than expected; cutoff=0.025) is shown as follows:

	# of Facilities	Percent of facilities
As Expected/Better than Expected	27	93.1%
Worse than Expected	2	6.9%

2b4.3. What is 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? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Significance testing identifies 7 facilities (23.3%) with worse than expected performance at a median of 24.5% of patients with reporting of nPCR data elements. The clear separation in measure performance between facilities identified with worse than expected performance versus those with as expected or better than expected performance provides support for the ability to identify clinically important differences in performance on this measure through significance testing.

For the Spring 2019 submission, significance testing identifies that 27 facilities (93.1%) have achieved expected performance, and 2 facilities (6.9%) had worse than expected performance. Between facilities identified with worse than expected performance versus those with as expected or better than expected performance, there exists a clear separation that provides support for the ability to identify clinically important differences in performance on this measure through significance testing.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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 eMeasures). 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.***

2b5.1. 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; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to 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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Missing is the outcome and this measure is reporting the percentage of non-missing. Thus, the missing data does not cause bias in this measure.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches*

for handling missing data that were considered and pros and cons of each)

N/A

2b6.3. What is 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 nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., 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, provide rationale for the selected approach for missing data*)

N/A