NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number: NQF 2988

Measure Title: ***Medication Reconciliation for Patients Receiving Care at Dialysis Facilities.***

Date of Submission: 5/10/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.
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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: **CROWNWeb ESRD Clinical Data Repository**	☐ 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*).

The *M*edication *Reconciliation for Patients Receiving Care at Dialysis Facilities* measure was tested using data from three KCQA member dialysis organizations, each with the capacity to provide retrospective analyses from a data warehouse/repository. All pertinent data from all eligible patients (i.e., adult and pediatric in-center and home hemodialysis and peritoneal dialysis) of the participating organizations during the testing period were included in the datasets.

1.3. What are the dates of the data used in testing? **April 1-September 30, 2015.**

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: **Dialysis facility**	☒ other: **Dialysis facility**

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 entity is the dialysis facility. All facilities in each of the three participating dialysis organizations were included in the analysis. The number of contributing facilities varied by month, but was approximately 5,292 facilities in each of the six months of the study. The range of contributing facilities was 5,258 (April 2015) to 5,319 (September 2015).

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*):
All patients (i.e., adult and pediatric in-center and home hemodialysis and peritoneal dialysis) in all facilities in each of the three participating dialysis organizations were included in the analysis. This translated to approximately 323,000 to 328,000 patients for each of the six months of the study period. Demographic information such as age, sex, and race were not assessed.

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.
Not applicable.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Sociodemographic information such as income, education, and language were not assessed.

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*):
Empirical reliability testing at the measure score level was conducted using the beta-binomial test (1) on the data pulls from the three participating dialysis organizations. Each organization pulled Q2 and Q3 data for 2015 for all facilities in accordance with the measure specifications, then provided their datasets for each facility (anonymized) for each month to an independent methodologist.

The beta-binomial method is characterized as a "natural model for estimating the reliability of simple pass/fail rate measures," and so is appropriate for this KCQA metric. Using this approach, reliability represents the ability of a measure to effectively distinguish the performance of one measured entity from another. The model is based on the beta distribution for the "true" scores for the measured entity, and assumes the entity's score is a binomial random variable conditional on the entity's true value that comes from the beta distribution. The beta distribution, which can be symmetric, skewed, or U-shaped, is "a very flexible distribution on the interval from 0 to 1" (1).

Reliability as calculated for the KCQA measure is thus the ratio of signal to noise, where the signal is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of 0 implies that all variability in a measure is attributable to measurement error, while a reliability of 1 implies all the variability is attributable to real differences in performance. The higher the reliability score, the greater the confidence the measure distinguishes the performance of one dialysis facility from another. A reliability statistic of 0.7 is generally viewed as an acceptable threshold (1).

1. **Adams, JL. The reliability of provider profiling: A tutorial. RAND Health, 2009.**

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*):
For the 6-month study period, for all facilities (excluding those with <=11 patients in a given reporting month, as per the measure specifications [approximately 3.7% of facilities each month]), the mean reliability of the measure is 0.9935 (range = 0.8166-1). (Results also contained in the KCQA Testing Data Attachment.)

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?*)
As previously noted, a reliability statistic of 0.7 is generally viewed as an acceptable threshold (1). Our reliability statistic of 0.9935 is excellent, suggesting the measure is highly reliable and effectively differentiates real differences in performance among facilities.

2b2. VALIDITY TESTING
2b2.1. 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 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*)

2b2.2. 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):*
Per NQF guidance (2), face validity of the measure was assessed through a systematic and transparent process by identified experts. Specifically, two separate groups of experts in the field of ESRD and dialysis care were identified—lead (voting) representatives from KCQA member organizations and a 9-member expert panel identified by the KCQA Steering Committee. Each group completed a face validity assessment that explicitly addressed whether performance scores resulting from the measure, as specified, provide an accurate reflection of quality. Individuals responded to the following two questions:

- **How likely is it that the measure score provides an accurate reflection of medication reconciliation quality? (highly unlikely; unlikely; neither likely nor unlikely; likely; highly likely)**

- **What is the likelihood that the measure can be used to distinguish good from poor quality? (highly unlikely; unlikely; neither likely nor unlikely; likely; highly likely)**

2. NQF. *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement.* April 2015. Available at: http://www.qualityforum.org/Projects/i-m/Measure_Evaluation_Guidance/Measure_Evaluation_Guidance.aspx. Accessed March 22, 2016.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*):
The face validity assessment yielded the following:

- **KCQA Member Organizations' Lead Representatives:**
  1. **77.3% of KCQA Lead Representatives (n=22) agreed it is highly likely or likely that the measure score provides an accurate reflection of medication reconciliation quality.**
  2. **77.3% of the panel agreed that it is likely/highly likely that the measure can be used to distinguish good from poor quality.**

- **Expert Panel:**
  1. **88.9% of the 9-member panel agreed it is highly likely or likely that the measure score provides an accurate reflection of medication reconciliation quality.**
  2. **77.8% of the panel agreed it is highly likely or likely that the measure can be used to distinguish good from poor quality.**

(Results also contained in the KCQA Testing Data Attachment.)

2b2.4. 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?*):
Both the Expert Panel and KCQA Lead Representatives showed significant agreement that scores from the measure as specified will accurately reflect medication reconciliation quality and will differentiate quality among providers. Our interpretation of these results is that this measure has substantial face validity.

_____

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions

2b3.1. 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*):
***Medication Reconciliation for Patients Receiving Care at Dialysis Facilities* has one patient-level and one facility-level exclusion:**

- **Patient-Level Exclusion: Transient patients, i.e., in-center hemodialysis patients who receive <7 dialysis treatments in the facility during the calculation month.**
- **Facility-Level Exclusion: Facilities with <=11 (i.e., <12) patients during the calculation month.**

For the patient-level exclusion, the analysis was conducted on the data pulls from the three participating dialysis organizations. Again, each participating organization pulled 2015 Q2 and Q3 data for all facilities in accordance with the measure specifications, then provided their datasets for each facility (anonymized) for each month. For each facility across the three participating dialysis organizations, the overall number and percentages of patients meeting the exclusion (transient patients, i.e., in-center hemodialysis patients who receive <7 treatments during the calculation

month) was recorded for each of the 6 months. The combined dataset was then examined to identify the monthly and overall frequencies of the occurrence, as well as the variability of the exclusion.

The facility-level exclusion parameter of <=11 was empirically determined during testing specifically to assess the impact on reliability of a "small numbers" effect. The effect of the measure's reliability in the context of excluding facilities at varying thresholds, including CMS's general implementation approach of excluding facilities with <11 patients/patient events (I.e., <=10). Both the percentage of facilities that would be excluded from measurement, as well as the reliability of the measure for small facilities were analyzed:

- **Using the CMS <11 threshold resulted in the exclusion of 3.3-3.6% of facilities from the measure, depending on the month.**
- **<11 threshold reliability statistics: Minimum = 0.3615; 10th Percentile = 0.6937; Median = 0.9174; 90th Percentile = 1; Maximum = 1**
- **At the 10th percentile, the measure does not achieve the previously cited reliability threshold of 0.7.**

Additional analyses were performed to determine the sample size that would yield a reliability statistic of 0.7 for all but outliers (defined as below the 10th percentile). Based on these analyses, a reliability statistic of at least 0.7 for the 10th percentile occurs at the threshold of <=11 (i.e., <12) patients in a given reporting month:

- **<=11 threshold reliability statistics: Minimum = 0.3622; 10th Percentile = 0.7089; Median = 0.9177; 90th Percentile = 1; Maximum = 1**

Based on this analysis, KCQA specified "facilities with <=11 (i.e., <12) patients in the reporting month" as the empirically appropriate small-numbers exclusion for the measure.

2b3.2. 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*):
Findings for the exclusions analysis are as follows:

- **Transient in-center hemodialysis patients who receive <7 dialysis treatments in the facility during the reporting month:**
  - **Mean number of patients excluded per facility in each study month: April = 1.95; May = 1.94; June = 1.96; July = 1.93; Aug = 1.94; Sep = 1.93**
  - **Mean number of patients excluded per facility, per month = 1.94**
  - **Total number (and percent) of patients excluded across all facilities in each study month: April = 8,972 (2.79%); May = 8,949 (2.77%); June = 9,073 (2.80%); July = 8,942 (2.74%); Aug = 9,007 (2.76%); Sep = 8,986 (2.75%)**
  - **Mean number (and percent) of patients excluded across all facilities, per month: 8,988 (2.77%)**
  - **Total number (and percent) of patient-months excluded across all facilities over the 6-month study period = 53,928 (2.77%)**

- **Facilities with <=11 patients during the reporting month:**
  - **Number (and percent) of facilities excluded in each study month: April = 180 (3.76%); May = 185 (3.86%); June = 177 (3.68%); July = 181 (3.76%); Aug = 179 (3.71%); Sep = 180 (3.72%)**

- **Mean number (and percent) of facilities excluded over the 6-month study period = 180.33/3.75%**

2b3.3. 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*):
The frequency and variability with which the exclusions were encountered during testing is sufficient to demonstrate they are necessary to prevent unfair distortion of performance results.

2b4. 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 2b5.

2b4.1. 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

2b4.2. 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 (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*)**:**

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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):*

Descriptive statistics for the performance measure scores for all tested entities (facilities) were constructed. These statistics include the mean, standard deviation and standard error, 95% confidence interval, median, mode, range of scores, and the interquartile range of scores across the measured entities.

Meaningful difference is defined as a significant spread (>20%) between minimum and maximum scores or a significant spread between median and minimum scores, median and maximum scores, and/or the interquartile range.

2b5.2. 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*):

Descriptive statistics for the performance measure scores are as follows:

- **Mean Performance Score = 52.62%**
- **Standard Deviation = 32.83**
- **Standard Error = 0.197**
- **95% Confidence Interval = 52.24 to 53.01**
- **Median Score = 48.18**
- **Mode of Scores = 100**
- **Range of Scores = 0 to 100**
- **Interquartile Range = 27.59 to 87.62**

2b5.3. 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?*)

Results are interpreted as showing a significant spread between the minimum and maximum scores (0-100), as well as the median and minimum (0-48.14) and maximum scores (48.14-100) and the interquartile range, indicating that the measure identifies clinically and practically meaningful

differences in performance among the measured entities.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*):
Medication Reconciliation for Patients Receiving Care at Dialysis Facilities **is constructed as an "all or nothing" measure, such that an event for which any of the numerator data elements are missing does not meet the measure criteria and is counted as a measure "fail" for that patient for that month. Consequently, there are no missing data to report on this measure.**

2b7.2. 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*):
Not applicable, as noted above.

2b7.3. 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):
Not applicable, as noted above.