NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0258

Corresponding Measures:

De.2. Measure Title: Consumer Assessment of Healthcare Providers and Systems (CAHPS) In-Center Hemodialysis Survey (ICH CAHPS)

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services

De.3. Brief Description of Measure: This is a survey-based measure and one of the family of surveys called CAHPS Surveys (Consumer Assessment of Healthcare Providers and Systems)that are focused on patient experience. The questionnaire asks End Stage Renal Disease (ESRD) patients receiving in-center hemodialysis care about the services and quality of care that they experience. Patients assess their dialysis providers, including nephrologists and medical and non-medical staff, the quality of dialysis care they receive, and information sharing about their disease. The survey is conducted twice a year, in the spring and fall with adult in-center hemodialysis patients. Publicly-reported measures focus on the proportion of survey respondents at each facility who choose the most favorable responses.

Three multi-item measures:

a. M1: Nephrologists' Communication and Caring (NCC)
b. M2: Quality of Dialysis Center Care and Operations (QDCCO)
c. M3: Providing Information to Patients (PIP)

Three Global items:

a. M4: Rating of the nephrologist
b. M5: Rating of dialysis center staff
c. M6: Rating of the dialysis facility

The first three measures are created from six or more questions from the survey that are reported as one measure score. The three global items are single-item measures using a scale of 0 to 10 to report the respondent's assessment.

The results are reported on Dialysis Facility Compare (DFC) on the Medicare.gov website.

1b.1. Developer Rationale: Quality health care for people with Medicare is a high priority for Executive Branch leadership, the Department of Health and Human Services, and the Centers for Medicare & Medicaid Services (CMS). CAHPS surveys are a crucial component of patient-centered care and a valuable feedback tool to help CMS continually improve the products and services it purchases for beneficiaries. They are especially critical for Medicare beneficiaries

with End-Stage Renal Disease (ESRD) who are a vulnerable population with multiple co-morbidities who rely on dialysis for life-sustaining treatment.

Until the creation of the ICH CAHPS Survey, very little was known about the provider-patient interaction or the quality of care received from the perspective of ESRD patients. ESRD patients form an especially vulnerable, minority population that is totally reliant on the ESRD facility and its predominantly non-professional staff for life-sustaining care. Additionally, this patient population is characterized by lower than average cognitive function, high incidence of mental health disorders, and an average of 3.5 co-morbidities. Many patients are reluctant to provide feedback for fear of retribution; others are reluctant to report facilities to ESRD Networks and/or state survey agencies because they perceive that these bodies are not responsive to patient concerns. In addition, many patients are not able to switch to another facility if they are unhappy with their care, making them a captive population, because there is not another one close enough, or one that has any openings in its schedule. Finally, some patients just don't understand what mechanisms are available for them to provide feedback on facility practices

After years of voluntary use by dialysis facilities, recent MIPPA legislation links the ICH CAHPS with the ESRD QIP. Section 1881(h) of the Social Security Act (the Act) states that the ESRD QIP "shall include, to the extent feasible, a measure (or measures) of patient satisfaction as the Secretary shall specify." The ICH CAHPS survey will support the Institute of Medicine's dimensions of care that focus on patient-centered care emphasizing "patient empowerment, improved patient-provider interaction, improved access, quality and outcomes." CAHPS is also directly responsive to recommendations by the Office of Inspector General (OIG), Government Accountability Office (GAO) and MedPAC to collect and monitor patient satisfaction with care and other access indicators to determine whether patients face obstacles in obtaining needed care.

S.4. Numerator Statement: There are a total of six ICH CAHPS measures. Three of them are multi-item measures and three are global ratings. Each measure is composed of the responses for all individual questions included in the measure. Missing data for individual survey questions are not included in the calculations. Only data from a "completed survey" is used in the calculations. Each measure score is at the facility level and averages the proportion of respondents who chose each answer option for all items in the measure. Each global rating is be scored based on the number of respondents in the distribution of top responses; e.g., the percentage of patients rating the facility a "9" or "10" on a 0 to 10 scale (with 10 being the best).

S.6. Denominator Statement: Patients receiving in-center hemodialysis at sampled facility for the past 3 months or longer are included in the sample frame.

The denominator for each question is composed of the sample members that responded to the particular question.

Proxy respondents are not allowed.

Only complete surveys are used. A complete survey is defined as one where the sampled patient answered at least 50 percent of the questions that are applicable to all sample patients: Q1-Q20, Q22, Q23, Q25-Q37, Q39-Q41 (Appendix provides more details about these questions.)

S.8. Denominator Exclusions: Exclusions:

a. Patients less than 18 years of age
b. Patients not receiving dialysis at sampled facility for 3 months or more
c. Patients who are receiving hospice care
d. Any surveys completed by a proxy (mail only mode or mixed mode)
e. Any ineligible patients due to death, institutionalization, language barrier, physically or mentally incapable.

De.1. Measure Type: Outcome: PRO-PM

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Facility, Other, Population : Regional and State

IF Endorsement Maintenance – Original Endorsement Date: Nov 15, 2007 **Most Recent Endorsement Date:** Jan 07, 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? NA

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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

Evidence Summary

- Brief background: This is a patient-reported outcome-based performance measure (PRO-PM) that uses survey data from patients ages 18+. The questionnaire asks ESRD patients receiving in-center hemodialysis care about the services and quality of care that they experience.
- Measure developer provided an updated logic model depicting the relationship between
    - Nephrologist communication and caring, improvement in patient understanding, and better outcomes for the patient resulting in improved experience of care
    - Center operations, opportunity for quality improvement, and better outcomes
    - Providing information to patient, improvement in patient understanding, and better outcomes for the patient resulting in improved experience of care
    - Global ratings measures, opportunity for quality improvement, and better outcomes
- Developer attests to use of focus groups with patients and family members during the development of the measure to suggest its importance to patients.
- Developer points to studies that have found that patient reports about care are moderately related to HEDIS measures and predictive of clinical outcomes as evidence that there are structures, processes, interventions or services that could be introduced to improve ICH CAHPS performance.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

Question for the Committee:

- *Does the Committee agree that there at least one thing that the provider can do to achieve a change in the measure results?*
- *Does the Committee agree that the target population values the measured outcome and finds it meaningful?*

Guidance from the Evidence Algorithm

Measure assesses performance on a patient-reported outcome (Box 1) → Empirical data suggest a structure, process, intervention or service may improve measure performance (Box 2) → PASS

Preliminary rating for evidence: ☒ Pass ☐ No Pass

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

Maintenance measures – increased emphasis on gap and variation

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

- Developer provided in the Appendix to their submission several tables with summary statistics for 3 years of survey data representing 2.7M individuals.
- As an example, for gap, summary statistics for the Fall 2017 data for the six domains of the ICH CAHPS survey:
  - Means ranged from 60.0 - 67.4 % of top box scoring
  - SDs ranged from 5.8 - 12.5
- Year over year performance appears to have stabilized in many domains, but continues to improve in others

Disparities

- Developer identified racial, language and disabilities disparities for six of the ICH CAHPS outcome measures
- Used multivariate regression analysis conducted using patient-level data from the 2016 CAHPS Spring Survey

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence: *For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.*

- high
- PASS

1b. Performance Gap: *Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?*

- Sufficient (and highly valued) performance gap. Disparities data limited by exclusions but showed disparities.
- Performance gap mean 60-67.4. Moderate opportunity for Improvement

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

2b. Validity: Testing**;** Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability**;** Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

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

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

Evaluators: NQF Scientific Methods Panel

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel. A summary of the measure and the Panel discussion is provided below.

- **Ratings for reliability:** 2 high, 3 moderate, 0 low and 1 insufficient → measure passes with moderate reliability
  - Testing included score-level and data element testing
  - Panelists noted that methodological approach as follows:
  - Inter-class reliability (ICR) (patient level) was computed on sum scores to determine how much of the variation in the sum is the result of true variation among dialysis facilities versus possible error. This was computed using one-way analysis of variance, regressing scores on dialysis facilities and then transforming the F-statistic as ICR = (F-1)/F
  - A second facility-level reliability calculated intraclass correlation coefficients (ICCs) for dialysis facilities using the intracc.sas macro. ICC(1,k) is another method of estimating the ratio of between dialysis facility variability to within-dialysis facility variability when patients differ across sites. The ICC(1,k) was calculated as
  - ICC = (Between-Facility Mean Square − Within-Facility Mean Square) / Between-Facility Mean Square
  - In addition, multi-item scales were assessed for internal consistency (Cronbach's alpha)
  - Generally, values of 0.70 or higher indicate acceptable reliability in all of these approaches.
  - Internal consistency results, across all multi-item measures, exceeded 0.70. SMP members noted that "ICR and ICC results rarely exceeded 0.70 but were consistently in the 0.60-0.70 when including only those centers with >30 patients responding. This may restrict the value of this measures to those facilities with sufficiently large volume to obtain >30 surveys in the study period."
  - Note: Measure developer used Cronbach's alpha to test for reliability at the data element level, but only for multi-item measures. Measure developer did not perform data element testing for single item measures within the survey.

- **Ratings for validity:** 2 high, 4 moderate, 0 low and 0 insufficient → measure passes with moderate validity
    - Testing included score-level and data element testing
    - SMP described the analysis as follows: "The developer used confirmatory factor analysis to evaluate the appropriateness of the components of the composite measures and correlational analyses were used to evaluate the relationship between each composite measure and global rating items (doctors, dialysis staff and dialysis center), composite intercorrelations and item correlations with composite scores. These analyses were used to confirm that the measures retain the relationships between the items and the composites as when they were developed and as such, are appropriate for re-evaluation."
    - SMP notes that "Correlational analyses demonstrated that the items and composite measures show consistent and positive expected relationships, which is evidence of construct validity. Similarly, the CFA confirmed the composite structure."
    - SMP member noted: "Proxy responses and hospice patients are excluded. Will it be known or reported how many such patients there are by facility, or can that information be used to weight the data?"

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The SMP is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The SMP is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

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

Combined Methods Panel Scientific Acceptability Evaluation

Measure Number: 0258

Measure Title: ICH CAHPS NQF#: 0258

Type of measure:

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

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

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐☒ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☐☐ **Other**Patient Experience of Care Survey Data

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

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

MP#4: No concerns.

MP#6: None

MP#1: None.

MP#2: No major concern. Six measures. Three multi-item and three global ratings. Each measure encompasses the responses for all questions included in the particular measure. Only complete data is used in the calculations. A complete survey is defined as one where the sampled patient answered at least 50 percent of the questions that are applicable to all sample patients. The measures score averages the proportion of those responding to each answer choice in all questions. Each global rating is scored based on the number of respondents in the distribution of top responses; e.g., the percentage of patients rating the facility a "9" or "10" on a 0 to 10 scale. Denominator is patients with ESRD receiving in-center hemodialysis at sampled facility for the past 3 months or longer. Proxy respondents are not allowed.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No** NA

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

   MP#6: comparable to the other CAHPS surveys. No concerns

   MP#2: Inter-class reliability (ICR) (patient level) was computed on sum scores to determine how much of the variation in the sum is the result of true variation among dialysis facilities versus possible error. This was computed using one-way analysis of variance, regressing scores on dialysis facilities and then transforming the F-statistic as ICR = (F-1)/F

   A second facility-level reliability calculated intraclass correlation coefficients (ICCs) for dialysis facilities using the intracc.sas macro.[1] ICC(1,k) is another method of estimating the ratio of between dialysis facility variability to within-dialysis facility variability when patients differ across sites. The ICC(1,k) was calculated as

   ICC = (Between-Facility Mean Square − Within-Facility Mean Square) / Between-Facility Mean Square

[1] Hamer, R. M. (1990). *Compute six intraclass correlation measures*. Available at http://support.sas.com/kb/25/031.html#ref

In addition, multi-item scales were assessed for internal consistency (Cronbach's alpha)

Generally, values of 0.70 or higher indicate acceptable reliability in all of these approaches.

Submission document: Testing attachment, section 2a2.2

MP#4: Methods used, including recoding are appropriate.

MP#3: Data element level reliability test was conducted at the patient level, not at the level that the measures are specified. Internal consistency test at the patient level is appropriate if a patient level summary score is first calculated based on his/her answers to multiple items and then facility level scores are calculated based on the patient level summary scores. For the multi-item measures specified in the application, a facility level summary score is first calculated for each survey item based on all eligible patients' answers to that item, measure score is the average of multiple item-specific facility scores.

Measure score reliability testing using ICC was appropriate.

MP#1: Cronbach's alpha was used to evaluate the internal consistency of the composite items and facility-level (interclass) SNR analysis was conducted, with the latter being most appropriate for this criterion.

MP#5: The methods used for reliability testing were generally acceptable, using standard and well-accepted methods, at both data element and measure score levels.

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

   MP#6: Inter-Class Reliability and Intra-class Correlations demonstrate good results at the facility level with a minimum of 60 responses.

   MP#2:

   a. **Internal consistency results, across all multi-item measures, exceeded 0.70. ICR and ICC results rarely exceeded 0.70 but were consistently in the 0.60-0.70 when including only those centers with >30 patients responding. This may restrict the value of this measures to those facilities with sufficiently large volume to obtain >30 surveys in the study period.**

   Submission document: Testing attachment, section 2a2.3

   MP#4: Testing sample is adequate with minimum population of 60 patient respondents. Moderate to high confidence that measure results are reliable. Appropriate test methods at element level and measure score level.

   MP#3: ICC testing results were acceptable.

   MP#1: The interclass coefficient (facility-level analysis) indicates that the six measures, when using the top box approach, are reasonably free from measurement error, although the reliability values were marginal for Providing Information to Patients and Global rating of kidney doctors (0.65). The results were reported for facilities with at least 30 respondents; it's unclear if this caveat is carried through the reporting of the measures. In other words, the measure is likely to be less reliable when there are fewer than 30 patients who report data for these measures.

   MP#5: The results of reliability testing were acceptable. At the measure score level, two closely-related metrics were presented – "Intra-class reliability" and intra-class correlation. The results are essentially the same, so choice of metric doesn't seem to make a meaningful difference. It's just a little unusual to see both.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

☐☒ **Yes**

☒ **No** Data elements reliability testing is not consistent with the measures as specified.☒☐ **No**

☐ **Not applicable** (data element testing was not performed)

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

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

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

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

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

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

    MP#3: NQF guideline requires that both data element reliability and measure score reliability be provided for instrument-based measure. Data element reliability testing for the measures with multiple items is not appropriate. No data element reliability testing is reported for global rating items.

    MP#1: Facility level analyses using the top box scoring method indicate that for two of the six measures, reliability is substandard.

    MP#2:

    a. **Restrict to facilities providing data on >30 patients?**

    MP#6: Internal Consistency results for the top box coded elements were adequate for the composite scores. Site level ICR/ICC for 30 or more responses were adequate

    MP#5: See response to item 7 above.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    MP#2:

    a. Proxy responses and hospice patients are excluded. Will it be known or reported how many such patients there are by facility, or can that information be used to weight the data?

    Submission document: Testing attachment, section 2b2.

    MP#4: No concerns with the exclusions to scope of the sample frame.

    MP#3: No concern.

    MP#1: None.

    MP#5: None

    MP#6: No exclusions

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

    Submission document: Testing attachment, section 2b4.

    MP#4: No concerns.

    MP#3: No concern

    MP#5: The developers find that a substantial number of sites are either significantly above or below the national mean in scores on essentially all the measures derived from this survey. They have not been able, or not attempted to, show that observed differences are meaningful.

MP#6: None.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
MP#4: No concerns.
MP#5: N/A
15. **Please describe any concerns you have regarding missing data.**

MP#2: Missing data for individual survey questions are not included in the calculations. The denominator for each question is composed of the sample members that responded to the particular question. This can be a problem for questions disproportionately left blank

MP#6: Agree with submitters that the patient-mix adjustment model accounted for any bias in missing survey responses and does not warrant further investigation

Submission document: Testing attachment, section 2b6.

MP#4: No concerns.

MP#3: No concern.

MP#1: N/A

MP#5: None

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** Patient mix adjustment ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒☐ Yes ☐ No ☐☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☒ Yes ☐ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐☒ Yes ☐ No NA

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐☒ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐☒ Yes ☐ No

16e**. Assess the risk-adjustment approach**

MP#4: Overall testing was sufficient.

MP#6: Robust patient mix adjustment model addresses critical aspects of risk adjustment for the population

MP#3: Given that dependent variables are binary indicators (top-box or not, bottom-box or not), it is not clear why ordinary least square regression instead of logistic regression was used to assess the effects of survey mode and patient characteristics. The results between OLS and logistic regression might be similar, but it would be useful to provide rationale.

MP#1: The risk-adjustment approach, while parsimonious, appears to be effective and appropriate.

MP#2: A re-evaluation of the 2014 patient mix was made in 2018 and it was determined to retain the original patient mix adjusters. The current patient mix adjusters are: Overall health; Overall mental health; Heart disease; Deaf or serious difficulty hearing; Blind or serious difficulty seeing; Difficulty concentrating, remembering, or making decisions; Difficulty dressing or bathing; Age; Sex; Education; Does the patient speak a language other than English at home; Did someone help the patient complete this survey; Total number of years on dialysis. Comprehensive set of adjuster variables

MP#5: The approach is generally acceptable**.** Social factors associated with survey responses are appropriately included in the adjustment model.

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

19. **Validity testing level:** ☐☒ **Measure score** ☐☒ **Data element** ☒☐ **Both**

20. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

    ☒ **Empirical validity testing of the measure score**

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

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

    MP#2: Composite model tested for fit and performed extremely well. Linear scores convert to ICH (In-Ctr Hemodialysis) CAHPS Star ratings. Uses CAHPS approach to patient case-mix adjustment

    MP#6: confirmatory factor analysis (CFA) of the items comprising the multi-item measures and correlations of each item with sum scores for its own multi-item measure, as well as with the other two multi-item measures were appropriate for testing validity

    Submission document: Testing attachment, section 2b2.2

    MP#4: Appropriate

    MP#3: Multiple data element validity tests were conducted and reported. Both CFA and discriminant among composites analysis were only done at the patient level, other analyses were done at both patient and facility level. Facility level top-box score analysis is appropriate for validity testing specific to the measures as specified.

    For the measure score validity testing, the validity of global rating is assumed, not tested. Composite measures scores were then correlated with the global rating scores.

    MP#1: The developer used confirmatory factor analysis to evaluate the appropriateness of the components of the composite measures and correlational analyses were used to evaluate the relationship between each composite measure and global rating items (doctors, dialysis staff and dialysis center), composite intercorrelations and item correlations with composite scores. These analyses were used to confirm that the measures retain the relationships between the items and the composites as when they were developed and as such, are appropriate for re-evaluation.

    MP#5: The developers rely on correlations among measures in the survey to establish measure score-level validity – a modest level of correlation (neither too high nor too low) is viewed as acceptable evidence of validity. There is no evidence presented linking measure scores to any independent measure of quality of care at the clinic level.

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

    Submission document: Testing attachment, section 2b2.3

MP#4: Results of testing suggest sufficient validity based on the 14 risk adjustment factors cited.

MP#3: The results of facility level data element validity testing were good for two measures (NCC and QDCCO) but somewhat moderate for one measure (PIP).

The results of measure score validity testing were positive, but these were conditioned on the assumed validity of the global ratings.

MP#1: Correlational analyses demonstrated that the items and composite measures show consistent and positive expected relationships, which is evidence of construct validity. Similarly, the CFA confirmed the composite structure.

MP#5: Results are generally acceptable, showing moderate correlations among scores and between specific domain scores and overall ratings of care.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

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

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

    Submission document*: Testing attachment, section 2b1.*

    ☐☒ **Yes**

    ☒☐ **No**

    ☒☐ **Not applicable** (data element testing was not performed)

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

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

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

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

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

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

    MP#4: Moderate to high validity rating based on the testing results.

    MP#6: No concerns, all tests performed supported validity

    MP#3: Validity of the global rating measures were assumed. It would have been much more appropriate had the hospital level CFA and discriminant validity tests were conducted.

    MP#1: developers convincingly show evidence that supports the overall construct of these measures, without some form of criterion validity, this reviewer feels moderate validity rating is appropriate.

    MP#5: The validity of this version of CAHPS and other versions rests largely on assessments of face validity. There is no information presented linking the CAHPS scores to any separate, independent measure of quality of care at the hospital level. The patterns of correlations do demonstrate adequate validity of the measure at the individual patient or data element level and do provide weak evidence for validity at the measure score level.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☒☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

MP#2: Good fit to composite model, with expected associations between components and composite

ADDITIONAL RECOMMENDATIONS

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

    MP#5: This set of measures, like all the other CAHPS measure sets, claims to be a set of outcome measures. These are not outcome measures. They do not reflect the state of a patient after treatment; they use the patient report to provide data on care processes. These are process measures, not outcome measures, even though the data come from patient surveys. A satisfaction survey would be an outcome measure, but these are "experience of care" surveys using the patient as a data source about care processes. Since users like CMS make distinctions in their P4P programs between process and outcome measures, often assigning greater weight to outcome measures, this is a very important distinction and the NQF endorsement process should make clear that these are not outcome measures.

    The developers provide information on use of the measures to create a star rating system for dialysis facilities. Since we have no guidelines or criteria for evaluating this specific use of measures, I would advise that NQF make no comment on, or endorsement of, the star rating system as described here. In fact, any final communication of an NQF endorsement decision should be clear that the endorsement, if given, does NOT cover the star rating system.

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Specifications: *Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?*

- None.
- No concerns regarding specifications

2a2. Reliability testing: *Do you have any concerns about the reliability of the measure?*

- No
- Methods and results are reasonable. Score of Moderate for reliability

2b2. Validity testing: *Do you have any concerns with the testing results?*

- Yes. The exclusions for patients with disabilities, on hospice, needing assistance to complete survey, non-English language are potentially distorting. They also represent populations for whom inadequate care is more likely. Efforts should be made to reduce exclusions and expand the survey.
- No concerns with testing results Moderate validity

Validity- Threats to Validity: *Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data). 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce*

comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- all adequate
- No threats to validity

Other Threats to Validity: *Other Threats to Validity (Exclusions, Risk Adjustment). 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?*

- Exclusions as above. Committee should discuss.
- No threats to validity

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- Developer cites challenges with sampling due to large number of facilities (over 6,000) and the number of patients using the facilities tending to be small (median ~50)
- Developer fails to mention the need to use vendors to perform data collection and the associated time and cost for survey administration, which is the sole responsibility of the facility

Questions for the Committee:

- Has the developer adequately addressed feasibilities challenges?

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

- A low rating is assigned because measure developer has not evaluated the burden on plans associated with measure implementation in the form of fees from retention of an approved CAHPS vendor to administer the surveys.
- Based on the information submitted there is low confidence or certainty that the criterion is met.
- Note: this is not a must pass criteria per NQF's current rules.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: *Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?*

- Expensive and difficult process. Low feasibility. Consideration of other sampling techniques On line.
- Moderate feasibility. Operationally feasible, Fee's fro PRO-PM's are necessary

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Developer notes that ICH CAHPS is currently being used within two federal quality and performance programs: Dialysis Facility Compare and ESRD QIP

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

Feedback on the measure by those being measured or others

- Focus groups with in-center hemodialysis patients were conducted for CMS in February 2016 (Baltimore) and April 2016 (San Antonio)
- "Current dialysis patients, in particular, emphasized that the survey questions and reported multi-item measures captured exactly the kind of information that was important to know and that they would look for about dialysis centers."
- Feedback from informal meetings with patient groups reflect high interest in the survey. Suggestions for improvement include using the web to collect survey data, shortening the questionnaire and conducting the survey annually instead of twice a year.

Additional Feedback: N/A

Questions for the Committee:

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

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

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

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

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

Improvement results

- The developer noted the average score for the multi-item measures increased over time while the average scores for the three ratings questions dropped from 2015 to 2016 and then increased in 2017. Specific results are included in 1b2., performance gap.

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

Unexpected findings (positive or negative) during implementation

Potential harms

- Measure developer offered no potential harms, though they did acknowledge receiving complaints from some patients about being surveyed more than once per year, as well as feedback on shortening the questionnaire.

Additional Feedback: N/A

***Questions for the Committee**:*

- Do you agree with the developer that there are no potential unintended consequences?

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

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a. Use: *4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?*

- adequate
- PASS

4b. Usability: *4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.*

- This is not an "outcome" measure, it is an "experience of care measure" which is valuable for patient reported experience however, there is no outcomes tied to improved scores (eg fewer complications, fewer missed appointments, better compliance with diet/meds, hospitalizations, complications)
- High opportunity for improvement

Criterion 5: Related and Competing Measures

Related or competing measures

- The following measures are all related, though not necessarily competing:
  - NQF 0005 CAHPS Clinician and Group Surveys V3.0
  - NQF 0006 CAHPS Health Plan Survey V5.0
  - NQF 0166 Hospital CAHPS Survey
  - NQF 0258 CAHPS In-Center Hemodialysis Survey
  - NQF 0517 CAHPS Home Health Care Survey
  - NQF 1741 CAHPS Surgical Care Survey
  - NQF 2548 Child Hospital CAHPS Survey
  - NQF 2967 CAHPS Home- and Community-Based Services Survey

Harmonization

N/A

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

Related and Competing: *Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?*

- No

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June/13/2019

- **No NQF members have submitted support/non-support choices as of this date**

1. Evidence and Performance Gap – 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

NQF_evidence_attachment_ICH_CAHPS_Updated_4-19-2019_-ALT-TXT.docx

1a.1 For Maintenance of Endorsement: 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

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0258

Measure Title: In-Center Hemodialysis CAHPS Survey

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 4/18/2019

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

Outcome

☒ Outcome:

☒ Patient-reported outcome (PRO): Multi-item measures Nephrologists' Communication & Caring; Quality of Dialysis Center Care and Operations; Providing Information to Patients; Single item global ratings: Rating of the nephrologist; Rating of dialysis center staff; Rating of the dialysis facility.

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:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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.

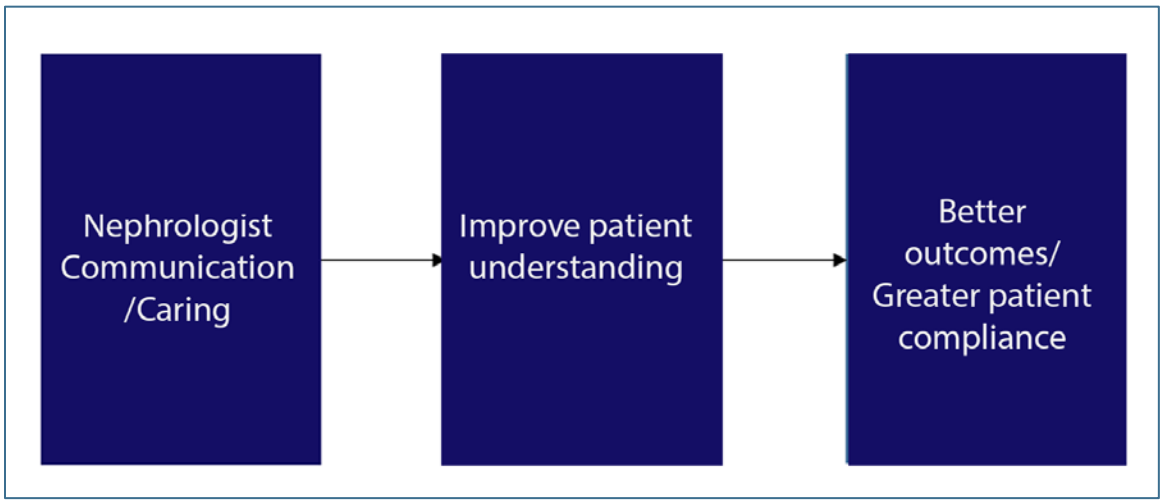
The ICH CAHPS Survey displays six measures on the Dialysis Facility Compare page of the Medicare web site (http://www.medicare.gov). All of these measures are calculated as top-box scores at the facility level, which

means we present the proportion of respondents at each facility who provided the most favorable responses to the items in the measure. For this reason, higher scores on these measures are expected to be associated with better dialysis facility quality. These measures were developed based on feedback from hemodialysis patients on what was important to them in defining high-quality care. During survey development we also consulted patient representatives, clinical professionals, and academics.

Nephrologists' Communication & Caring

This multi-item measure covers the interactions with the patient's nephrologist including listening carefully, explaining things in a way that was easy to understand, showing respect, spending enough time, caring about patient as a person, and whether the nephrologist was informed and up-to-date about health care received from other doctors . We expect higher scores on the measure to result in greater patient understanding and therefore better quality outcomes, including patients being more compliant with their therapy.

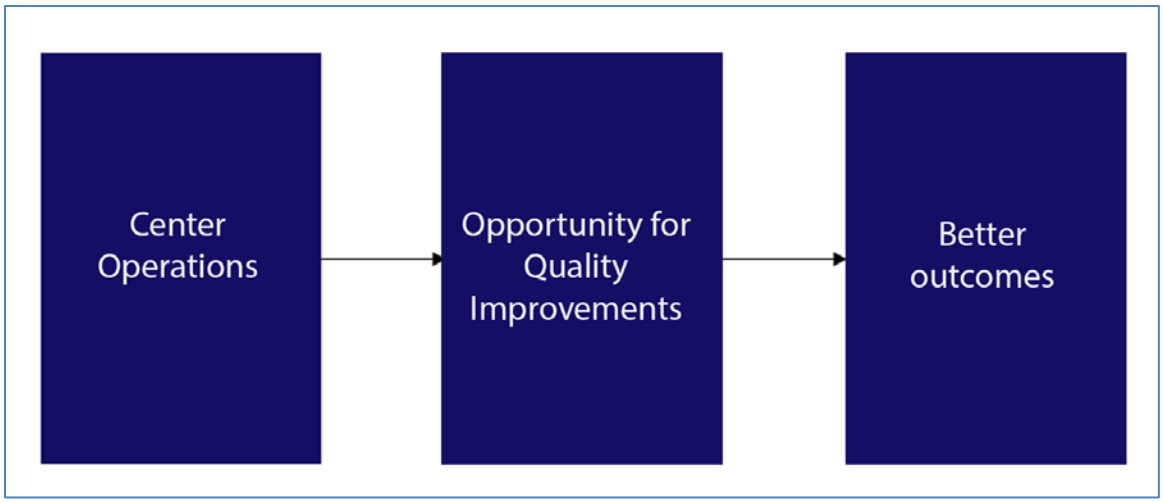
Figure 1 Logical Relationship of Nephrologist Communications to Quality Outcomes



Quality of Dialysis Center Care and Operations

This multi-item measure focuses on patient communication and interactions with dialysis center staff and their experiences at the dialysis center and with the staff. The measures addresses interactions between patients and staff, including showing respect, spending enough time, providing a caring and comfortable environment, ensuring privacy, minimizing pain, providing monitoring as needed, managing problems during dialysis, discussing dietary issues, ensuring cleanliness of the facility, discussing blood test results in understandable ways, and minimizing wait times. The results of this measure provide facilities with an opportunity to monitor and improve quality and therefore patient outcomes.

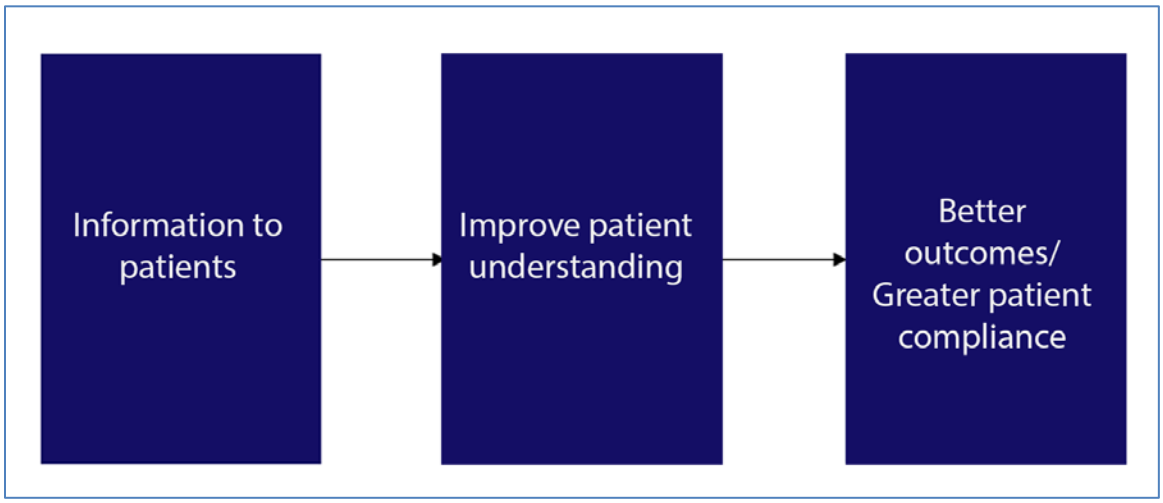
Figure 2 Logical Relationship of Center Operations Measure to Better Outcomes



Providing Information to Patients

This multi-item measure focuses on providing information to patients so that they have the knowledge and ability to manage their care. This measure includes several questions that focus on empowering dialysis patients by providing information about their rights, how to handle health problems at home, how to handle emergencies while being dialyzed, treatment options and information about which treatment is appropriate for them. The hypothesized relationship to outcomes is that providing information to patients will improve patient understanding and therefore produce better outcomes, and in particular, greater compliance.

Figure 3 Logical Relationship of Information Provided and Better Outcomes



Global Ratings Measures

All three global ratings measures are single-item measures. All three hypothesize the same relationship between the measure and outcomes, specifically that ratings provide dialysis facilities with indicators of opportunities for quality improvement. Improvements in these areas should lead to better outcomes.

Global Rating – Rating of the nephrologist

Rating the nephrologist on a scale of 0 (Worst kidney doctor possible) to 10 (Best kidney doctor possible).

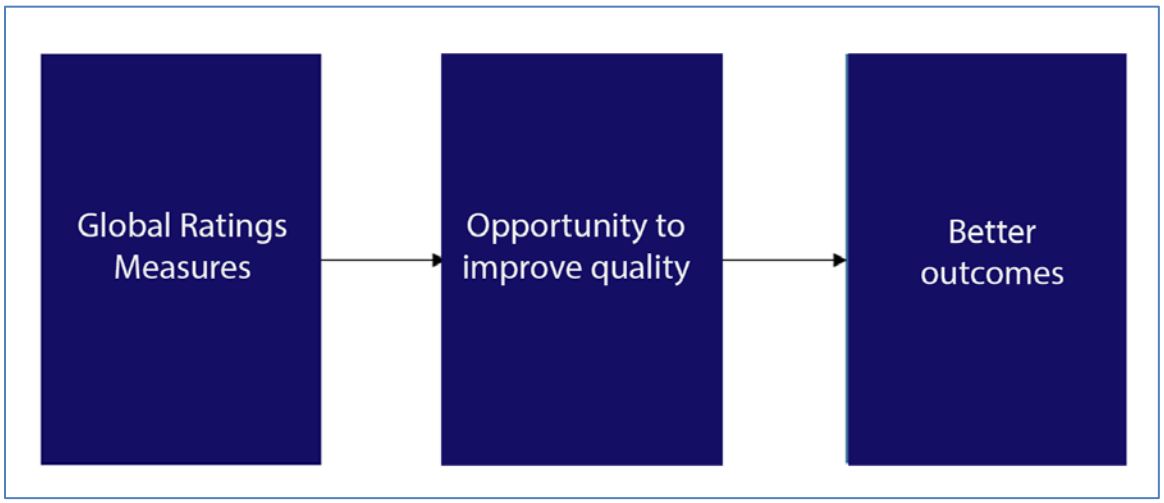
Global Rating – Rating of the dialysis center staff

Rating of the dialysis center staff on a scale of 0 (Worst dialysis staff possible) to 10 (Best dialysis center staff possible).

Global Rating – Rating of the dialysis facility

Rating of the dialysis facility on a scale of 0 (Worst dialysis facility possible) to 10 (Best dialysis facility possible).

Figure 4 Relationships of Global Ratings Measures and Outcomes



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.)

The ICH CAHPS survey was created to address accountability, drive quality improvement, and provide performance measures for public reporting. The survey items were developed following a rigorous process starting with a literature review, focus groups with patients and their family members, focus groups with nephrologists and facility staff, review of existing surveys and a Technical Expert Panel, a draft survey was developed and then tested. The focus groups, in particular, asked participants what they most wanted to know about the patient experience of in-center hemodialysis patients. The individual survey items were created with this in mind.

In addition, we know that at least some in-center hemodialysis facilities and large dialysis organizations (LDOs) are using survey results for quality improvement. Approved survey vendors are offering services related to helping their clients improve their CAHPS scores.

****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.

Our current research related to the In-Center Hemodialysis Survey has focused on the relationships between patient demographics and patient-reported outcomes. The following descriptions apply to recent unpublished research by RTI International under contract to CMS.

Results from Subgroup Analysis Using Multiple Regression Analysis at the Individual Patient Level and ICH CAHPS Outcome Scores as the Dependent Variables

- The independent variable for the percentage of the dialysis facility's patients who are Black showed consistently statistically significant and negative coefficients across seven of the eight regression models for the response categories for the dialysis facilities with the higher percentages of patients who are Black. This indicates that those dialysis facilities with higher percentages of Black patients had lower scores on seven patient experience of hemodialysis care outcome measures after controlling for race at the individual level. We also found a nonlinear relationship for six of the eight outcomes, where the highest percent Black categories had larger negative effects on the outcome scores.
- The independent variable for patient-level race showed statistically significant and negative coefficients for American Indian or Alaska Native, Asian, and Mixed race patients, when assessed against the White only omitted response category.
- We also found negative effects for non-English language, for the ICH CAHPS variable for the main language spoken at home. This result indicates that additional efforts toward linguistic and cultural sensitivity may be needed for ICH care.
- In addition to the racial and ethnic effects, we found several consistently statistically significant effects for other policy-relevant subgroups of patients. They included the negative impacts on the ICH CAHPS outcomes found for patients with four types of disabilities, for larger dialysis facilities with 30 or more stations, and for facilities with for-profit ownership. They also included the positive effects for lower patient to nurse ratios and less rural counties.
- The results of this study also point to several areas where dialysis facilities may be able to intervene to improve their quality as reflected in the ICH CAHPS outcome scores, in addition to considering ways to mitigate the racial disparities. These include increasing the numbers of nurses treating dialysis patients, providing additional care or accommodations for patients with disabilities, and providing assistance for patients whose language at home is not English.

Results from Subgroup Analysis Using Multiple Regression Analysis at the Dialysis Facility Level and Changes Over Time in the ICH CAHPS Outcome Scores as the Dependent Variables

This is a brief summary of findings from our work:

- The independent variable for the percentage of the dialysis facility's patients who are Black showed consistently statistically significant and negative coefficients across five of the six regression models for the response categories for higher percentages of patients who are Black. This indicates that those facilities with higher percentages of Black patients had lower scores on five of the six patient experience of hemodialysis care outcome measures.
- In addition, the independent variable for the percentage of the dialysis facility's patients who were Other race showed consistently statistically significant and negative coefficients across all six regression models for the response category for the highest percentage of patients who were Other race. It also had negative and significant results for the response category for the next highest percentage of Other race patients for three of the six regression models. This indicates that those facilities with the higher percentages of Other race patients had lower scores across all six ICH CAHPS outcome measures.
- The other policy-oriented variable that showed consistently significant results was for the change in the nurse to patient ratio. This variable was provides evidence that increasing the numbers of nurses in relation to the numbers of patients at dialysis facilities can improve these ICH CAHPS patient experience of care outcomes.

Relevant citations:

1. Norris K. and L. Agodoa. (2005). Unraveling the racial disparities associated with kidney disease. Kidney International 68: 914-924.

2. Nicholas S., Kalantar-Zadeh K. and K. Norris. (2013). Racial disparities in kidney disease outcomes. Seminars in Nephrology 33(5): 409-415.

3. Trivedi A., Zaslavsky A., Schneider E. and J. Ayanian. (2005). Trends in the quality of care and racial disparities in Medicare managed care. New England Journal of Medicine 353(7): 692-729.

4. Kucirka L., Grams M., Lesser J., et al. (2011). Association of race and age with survival among patients undergoing dialysis. JAMA 306(6): 620-626.

5. Sehgal A. (2003). Impact of quality improvement efforts on race and sex disparities in hemodialysis. JAMA 289(8): 996-1545.

6. Himmelfarb J. and T. Ikizler. (2010). Hemodialysis. New England Journal of Medicine 363(19): 1833-45.

7. Rodrigue J., Pavlakis M., Egbuna O., et al. (2012). The "House Calls" trial: a randomized controlled trial to reduce racial disparities in live donor kidney transplantation: rationale and design. Contemporary Clinical Trials 33:811-818.

8. Ravani P., Palmer S., Oliver M., et al. (2013). Associations between hemodialysis access type and clinical outcomes: a systematic review. Journal of the American Society of Nephrology 24: 465-473.

9. Zarkowsky D., Arhuidese I., Hick C., et al. (2015). Racial/ethnic disparities associated with initial hemodialysis access. JAMA Surgery 150(6): 529-536.

10. Rodriguez H., von Glahn T., Grembowski D., Rogers W. and D. Safran. (2008). Physician effects on racial and ethnic disparities in patients' experiences of primary care. Journal of General Internal Medicine 23(10): 1666-72.

11. Goldstein E., Elliott M., Lehrman W., Hambarsoomian K. and L. Giordano. (2010). Racial/ethnic differences in patients' perceptions of the quality of inpatient care using the HCAHPS survey. Medical Care Research and Review 67(1): 74-92.

We also want to note that CAHPS surveys provide information about selected aspects of care. One would not expect a large correlation between CAHPS surveys and clinical performance indicators because they provide complementary information about care. However, quality of care as measured in the CAHPS surveys should correlate significantly with aspects of clinical performance that are similar to the CAHPS domains. Indeed, some studies have found that patient reports about care are moderately related to HEDIS measures and predictive of clinical outcomes.

Other studies demonstrate that some performance indicators are unrelated to CAHPS measures. For example, a study of individuals 65 and over in vulnerable communities reported that measures of technical quality were not significantly associated with the CAHPS Health Plan Survey's global rating of care item.

Relevant citations:

- Fremont AM, Cleary PD, Hargraves JL, Rowe RM, Jacobson NB, Ayanian, JZ. Patient-Centered Processes of Care and Long-Term Outcomes of Myocardial Infarction. J Gen Int Med. 2001 Dec;16(12):800-8.

- Schneider EC, Zaslavsky AM, Landon BE, Lied TR, Sheingold S, Cleary PD. National quality monitoring of Medicare health plans: The relationship between enrollees' reports and the quality of clinical care. Med Care. 2001 Dec;39(12):1313-25.

- Chang JT, Hays RD, Shekelle PG, MacLean CH, Solomon DH, Rueben DB., Roth CP, Kamberg CJ, Adams J, Young RT, Wenger NS. Patients' global ratings of their health care are not associated with the technical quality of their care. Ann Intern Med. 2006 May 2;144(9):665-72.

- Rao M, Clarke A, Sanderson C, Hammersley R. Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study. BMJ. 2006 Jul 1;333(7557):19.

1a.3. SYSTEMATIC REVIEW(S) 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)

☐ 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**	
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.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

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.

Patient experience of health care is an important aspect of provider quality. Empirical studies have found a positive relationship between patient experience and other quality measures. Acknowledging the value of patient experience, CMS includes it in their Meaningful Measures framework. This framework emphasizes those measures most critical to providing high-quality care and improving individual outcomes.

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

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

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.

Quality health care for people with Medicare is a high priority for Executive Branch leadership, the Department of Health and Human Services, and the Centers for Medicare & Medicaid Services (CMS). CAHPS surveys are a crucial component of patient-centered care and a valuable feedback tool to help CMS continually improve the products and services it purchases for beneficiaries. They are especially critical for Medicare beneficiaries with End-Stage Renal Disease (ESRD) who are a vulnerable population with multiple co-morbidities who rely on dialysis for life-sustaining treatment.

Until the creation of the ICH CAHPS Survey, very little was known about the provider-patient interaction or the quality of care received from the perspective of ESRD patients. ESRD patients form an especially vulnerable, minority population that is totally reliant on the ESRD facility and its predominantly non-professional staff for life-sustaining care. Additionally, this patient population is characterized by lower than average cognitive function, high incidence of mental health disorders, and an average of 3.5 co-morbidities. Many patients are reluctant to provide feedback for fear of retribution; others are reluctant to report facilities to ESRD Networks and/or state survey agencies because they perceive that these bodies are not responsive to patient concerns. In addition, many patients are not able to switch to another facility if they are unhappy with their care, making them a captive population, because there is not another one close enough, or one that has any openings in its schedule. Finally, some patients just don't understand what mechanisms are available for them to provide feedback on facility practices

After years of voluntary use by dialysis facilities, recent MIPPA legislation links the ICH CAHPS with the ESRD QIP. Section 1881(h) of the Social Security Act (the Act) states that the ESRD QIP "shall include, to the extent feasible, a measure (or measures) of patient satisfaction as the Secretary shall specify." The ICH CAHPS survey will support the Institute of Medicine's dimensions of care that focus on patient-centered care emphasizing "patient empowerment, improved patient-provider interaction, improved access, quality and outcomes." CAHPS is also directly responsive to recommendations by the Office of Inspector General (OIG), Government Accountability Office (GAO) and MedPAC to collect and monitor patient satisfaction with care and other access indicators to determine whether patients face obstacles in obtaining needed care.

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.*

The combined spring and fall publicly reported average top-box score has increased from 2015 to 2017 for the Nephrologists' communication and caring multi-item measure (65.6 in 2015, 66.6 in 2016, and 67.4 in 2017), quality of

dialysis center care and operations multi-item measure (61.0 in 2015, 61.3 in 2016, and 62.3 in 2017), and providing information to patient's multi-item (78.2 in 2015, 79.2 in 2016, and 79.9 in 2017). The average scores for the three ratings questions dropped from 2015 to 2016 and then increased in 2017. The range of the scores within a given score has remained consistent across the three years. For example, the IQR for the rating of the nephrologist was 16 in 2015, 15 in 2016 and 16 in 2017. Similarly, for the rating of the dialysis center staff, the IQR was 18 in 2015 and 17 in both 2016 and 2017. For the rating of the dialysis facility the IQR was 17 for all three years. The IQR remained the same for the nephrologist communication and caring multi-item (12 for all three years) and for the quality of dialysis center care and operations (10 for all three years) and nearly the same for the providing information to patients multi-item (7 in 2015 and 8 in 2016 and 2017).

The composition of the respondents has remained mostly consistent across the six survey periods used to create the publicly reported scores (see Table 4) except for the percentage of respondents speaking English at home. In the spring 2015 81.3% of the respondents spoke English at home. A decreasing trend in the percentage of respondents speaking English at home is observed with 77.2% of the respondents speaking English at home in Fall 2017. Similarly, the percent of Hispanics responding to the survey has dropped from 71% in Spring 2015 to 67.2% in Fall 2017. Also changing is the distribution of respondents receiving dialysis for 3-4 years. In Spring 2015 25.1% of respondents received dialysis for 3-4 years. An increasing trend in this category is observed with 27.1% receiving dialysis for 3-4 years in Fall 2017.

Found in Table 3 in the Appendix are summary statistics (number of CCNs reported, mean, standard deviation, max, min, IQR and deciles) for the publicly reported top-box patient mix adjusted scores using the 2015 Spring/Fall, 2016 Spring/Fall, and 2017 Spring/Fall survey response data. Included in the descriptive analysis were CCNs that met the criteria for public reporting (30 or more survey completes, completes in both survey periods and facility served survey eligible patients). The characteristics of the respondents are found in Table 4 in the Appendix.

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.

NA

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.*

We found racial, language, and disabilities disparities for six ICH CAHPS outcome measures in multivariate regression analysis conducted using patient-level data from the 2016 ICH CAHPS Spring Survey (N = 107,582). The outcome measures included the three ICH CAHPS multi-item outcome measures and the three single-item ICH CAHPS global rating outcome measures

The regression models included 32 independent variables that controlled for patient, survey administration, regional, dialysis facility characteristics, and dialysis facility patient profile factors. As a result, the racial, language, and disabilities disparities found for these ICH CAHPS outcome measures cannot be explained as resulting from those other factors.

Racial disparities were found at the dialysis facility-level for the percentage of a dialysis facility's patients who are Black. This variable showed consistently statistically significant and negative coefficients across five of the six regression models for the response categories for the dialysis facilities with the higher percentages of patients who are Black. These facility-level racial disparities were found after controlling for Black race at the individual patient level.

Racial disparities were also found at the patient level for three other racial groups. The regression models found statistically significant and negative coefficients for American Indian or Alaska Native patients (for all six ICH CAHPS outcomes measures), Asian patients (five of the six outcome measures), and Mixed race patients (four of the six outcome measures).

Language disparity was found at the patient level for non-English language, as the main language spoken at home. This variable showed statistically significant and negative coefficients for all six ICH CAHPS outcome measures.

Disabilities disparities were found at the patient level. These included four types of disabilities that showed statistically significant and negative coefficients, including difficulty remembering (for all six ICH CAHPS outcomes measures), difficulty dressing (for all six ICH CAHPS outcomes measures), blindness (five of the six outcome measures), and deafness (four of the six outcome measures) patients.

Table 4 in the Appendix summarizes the distribution of respondent characteristics.

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

NA

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

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

Person-and Family-Centered Care

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

Populations at Risk : Individuals with multiple chronic conditions

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.)*

https://ichcahps.org/SurveyandProtocols.aspx

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.

Attachment **Attachment:** ICH_SurveyStandard_English_Nov2018.docx

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.

Patient

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.

No

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.

No important changes to the measure specification.

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).

There are a total of six ICH CAHPS measures. Three of them are multi-item measures and three are global ratings. Each measure is composed of the responses for all individual questions included in the measure. Missing data for individual survey questions are not included in the calculations. Only data from a "completed survey" is used in the calculations. Each measure score is at the facility level and averages the proportion of respondents who chose each answer option for all items in the measure. Each global rating is be scored based on the number of respondents in the distribution of top responses; e.g., the percentage of patients rating the facility a “9” or “10” on a 0 to 10 scale (with 10 being the best).

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).

Multi-Item Measures

Each of the multi-items measures is produced by combining responses to all of the questions included in the measure.

Step 1 – Identify relevant cases: include only cases where survey status is a "completed survey" and include only cases with non-missing values on each of the individual questions.

Step 2 - Calculate the proportion of cases in each of the response categories for each question.

Step 3 – Combine responses from each of the questions to form the measure by calculating the average proportion responding to each category across all of the questions in the measure.

Measure: M1 - Nephrologists’ Communication – Q3,Q4,Q5,Q6,Q7, and Q9;

Measure: M2 - Quality of Dialysis Center Care and Operations:

q10,Q11,Q12,Q13,Q14,Q15,Q16,Q17,Q21,Q22,Q24,Q25,Q26,Q27,Q33,Q34, and Q43

Measure: M3 - Providing Information to Patients: Q19,Q28,Q29,Q30,Q31,Q36,Q38,Q39,and Q40

The measures include a "top-box" score which reflects the average proportion of respondents who chose the most favorable option in answering questions in the measure. The "middle-box" score refers to the average proportion of respondents who chose mid-level responses. Items with a binary response will not have a middle box score. The "bottom-box" score refers to the average proportion of respondents who chose least favorable responses.

Global Ratings:

Global Item – M4 - Rating of nephrologists : Q8

Global Item – M5 - Rating of the dialysis center staff: Q32

Global Item – M6 - Rating of the dialysis facility: Q35

Step 1 – Identify relevant cases: Include only cases where survey status is a completed survey and include only cases with non-missing values on the overall rating question.

Step 2 – Calculate the proportion of cases in each of three re-coded response categories that represent top-,middle-, and bottom-box scores

The numerator is the number of respondents for whom the global rating (Xi) is 0-6.

The denominator is the total number of respondents that responded to this question (Wi)

Proportion of respondents who gave a rating of 0-6 (bottom box score):

The numerator is the number of respondents for whom the global rating (Xi) is 0-6.

The denominator is the total number of respondents (Wi).

The proportion can be defined as follows:

Let X1i = 1 when Xi is 0-6

= 0 otherwise

P1 = (SumiX1i) / SumiWi

Proportion of respondents who gave a rating of 7 or 8 (middle box score):

The numerator is the number of respondents for whom the global rating (Xi) is 7 or 8.

The denominator is the total number of respondents (Wi).

The proportion can be defined as follows:

Let X2i = 1 when Xi is 7 or 8

= 0 otherwise

P2 = (SumiX2i) / SumiWi

Proportion of respondents who gave a global rating of 9 or 10:

The numerator is the number of respondents for whom the global rating (Xi) is 9 or 10.

The denominator is the total number of respondents.

The proportion can be defined as follows:

Let X3i = 1 when Xi is 9 or 10

= 0 otherwise

P3 = (SumiX3i) / SumiWi

A facility's score on the global rating item is the proportion of cases in each response category.

Star Ratings

A linear mean is also calculated on the same question items above. Rather than recoding the item into a binary response, all levels for an item are used. The item is then transformed on a 0 to 100 scale and an average is calculated. This puts all question items, regardless of the number of responses, on the same 0 to 100 scale. A factor analysis is then conducted on each facility's linear means and assigns them to one of five groupings. The group with the lowest linear means gets 1-star. The group with the next highest linear means gets 2-stars. And the process repeats until you get to the fifth group with the highest possible linear means which gets 5-stars. A Star Rating is generated for each of the three global items as well as each of the three multi-item measures. Finally, an overall Star Rating is calculated which is a simple average of the six previous Star Ratings, rounded up. i.e. if a facility had 3 3-stars and 3 4-stars, the overall Star Rating would be (3+3+3+4+4+4)/6 = 3.5, which is rounded up to 4-stars.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Patients receiving in-center hemodialysis at sampled facility for the past 3 months or longer are included in the sample frame.

The denominator for each question is composed of the sample members that responded to the particular question.

Proxy respondents are not allowed.

Only complete surveys are used. A complete survey is defined as one where the sampled patient answered at least 50 percent of the questions that are applicable to all sample patients: Q1-Q20, Q22, Q23, Q25-Q37, Q39-Q41 (Appendix provides more details about these questions.)

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).

See information in S.6 for details.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Exclusions:

a. Patients less than 18 years of age
b. Patients not receiving dialysis at sampled facility for 3 months or more
c. Patients who are receiving hospice care
d. Any surveys completed by a proxy (mail only mode or mixed mode)
e. Any ineligible patients due to death, institutionalization, language barrier, physically or mentally incapable.

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.)*

All data for measure calculations is based on surveys that are completed by any of the approved modes: telephone only, mail only or mixed mail/telephone follow up. A survey is considered complete if at least 50 percent of the core survey questions are answered by the respondent. Missing data for individual survey questions are not included in the calculations.

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.)*

Not applicable.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Other

If other: The ICH CAHPS survey data is adjusted for public reporting using survey mode and 13 patient characteristics. Usually patient experience surveys are adjusted for factors not under the control of the provider that impact response tendencies. This is called patient mix or case mix adjustment. We conduct these adjustments so meaningful comparisons between ICH facilities can be made. The 2014 Mode Experiment was conducted to determine the set of patient mix adjusters. A re-evaluation of patient mix was made in 2018 and it was determined to retain the original patient mix adjusters. The current patient mix adjusters are: Overall health; Overall mental health; Heart disease; Deaf or serious difficulty hearing; Blind or serious difficulty seeing; Difficulty concentrating, remembering, or making decisions; Difficulty dressing or bathing; Age; Sex; Education; Does the patient speak a language other than English at home; Did someone help the patient complete this survey; Total number of years on dialysis. The coefficients for patient mix adjustment are published on the survey website after each Dialysis Facility Compare refresh. They can be found at: https://ichcahps.org/Home.aspx in the Quick Links section.

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.)*

1. Only surveys that meet the completeness criteria of greater than or equal to 50% will be included in the calculation of measures/global ratings.
2. Each of the three multi-item measures consists of 6 or more questions that are reported as one measure score. Scores are created by first determining the proportion of answers to each response option for all questions in the measure. The final measure score averages the proportion of those responding to each answer choice in all questions. Only questions that are answered by survey respondents will be included in the calculation of measure scores.
3. Statistical adjustments are made for mode of administration, and the set of patient-mix characteristics noted in S.11a. The statistically adjusted score for the three ratings questions and a given individual survey question that is included in one of the three ICH CAHPS Survey multi-item measures is the sum of a series of products in the equation shown below.

= y + a1(h1 - m1) + a2(h2 - m2) + a3(h3 - m3) + . . . + a28(h28 - m28) + a29*h29 + a30*h30

where

is the facility's adjusted score (top or bottom box) for a ratings question or the individual ICH CAHPS question included in the multi-item measure.

y is the facility's "raw score," or mean on the respective unadjusted top or bottom box ICH CAHPS ratings question or question included in the multi-item measure.

a1 to a28 are the national-level patient characteristic adjustments, for the global ratings questions and individual questions that comprise the multi-item measures.

a29 to a30 are the national-level survey mode adjustments for the global ratings questions and the individual questions that comprise the multi-item measures.

h1 to h28 are the facility's mean proportions of patients with each of the patient characteristics in the same row.

h29 to h30 are the facility's proportion for a given mode. This value will always be 0 or 1 because within a given facility all surveys are completed by either phone, mail, or mixed mode.

m1 to m28 are the national mean proportions of patients with each of the patient characteristics.

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.

For national implementation of the survey, the sample for each Medicare-certified ICH facility for the ICH CAHPS survey is selected using patient-level data that ICH facilities submit to CMS via the Consolidated Renal Operations in a Web-enabled Network (CROWNWeb). The sample is selected at the CCN level (CMS certification number), a unique provider identification number assigned to each Medicare-certified ICH facility. For each semi-annual survey patients who received care during the sampling window who meet survey eligibility criteria will either be chosen randomly or selected with certainty depending on the number of survey-eligible patients the ICH facility served during the preceding year.

For statistical precision, a target minimum of 200 completed ICH CAHPS surveys has been set for each ICH facility over each 12 month reporting period. The target number of 200 completed surveys is expected to produce a confidence interval that has a bound of ±-0.07. Please note that all facilities may not meet this target, but CAHPS measures are still calculated for them. A facility must have a total of 30 completed responses over the last two data collection periods to have scores reported on the Dialysis Facility Compare web site on Medicare.gov.

If a facility's patient volume in the preceding year is large enough, the number of patients sampled for each semiannual survey period will be sufficient to yield a minimum of 200 completed surveys in a 12 month period. If a facility does not serve enough survey-eligible patients over a given 12 month period to yield 200 completed surveys from two semiannual surveys, the sample will include all of the facility's survey-eligible patients.

To achieve these targets, we sample up to 240 survey-eligible patients from each CCN each survey period. If the number of patients within a CCN is over 240, then a simple random sample is drawn. But if the number is fewer than 240, then we conduct a census for that CCN. There are over 6,000 CNNs and our overall sample is typically between 330,000-350,000 patients.

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.

Results are based on the semiannual administration of the ICH CAHPS survey to ESRD patients meeting eligibility criteria. There are three modes of administration: mail only, telephone only, and mixed mode (mail followed by telephone). Detailed guidelines for survey administration may be found at https://ichcahps.org with detailed protocols for each mode of survey administration as well as specifications for submission of results to the ICH CAHPS Data Center.

There is no minimum response rate on the survey. All approved vendors must follow all of the survey administration protocols. If those protocols are followed, there is no minimum response rate.

Detailed response rates can be found in the Appendix.

Calculation of response rate:

RR = # Completed Surveys /(Patients Sampled – Ineligible Patients)

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Instrument-Based 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.

The survey instrument is the In-Center Hemodialysis CAHPS survey.

Modes: mail only, telephone only, or mixed mode. For the mail-only mode, data is collected for a 12-week period. For ICH CAHPS Spring surveys, data collection activities will be conducted from April through mid-July. Fall surveys will be conducted from October through mid-January. A second wave mailing is sent to non-respondents four weeks after the first mailing. For the telephone-only mode, data collection occurs during the same 12-week period as the mail survey. Vendors may make a maximum of 10 attempts to contact a patient by telephone. For the mixed-mode survey, the data collection period is the same as the other modes. The respondent is first mailed a questionnaire. If the respondent does not reply within four weeks follow-up telephone calls are made. The vendor may make up to 10 attempts to contact the respondent by telephone.

Languages of administration: English, Spanish, Chinese, Samoan, and Simplified and Traditional Chinese (only English or Spanish may be conducted by telephone mode or mixed-mode).

Please see https://ichcahps.org/SurveyandProtocols.aspx for the English version of the survey and translations.

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)*

Available at measure-specific web page URL identified in S.1

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

Facility, Other, Population : Regional and State

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Post-Acute Care

If other:

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

2. Validity – See attached Measure Testing Submission Form

ICH_CAHPS_0258_NQF_Measures_Testing_Form_FINAL_1-21-2019-636886012884683602.docx

2.1 <u>For maintenance of endorsement</u>

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 <u>For maintenance of endorsement</u>

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 <u>For maintenance of endorsement</u>

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.

Yes - Updated information is included

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:**
Measure Title: ICH CAHPS NQF#: 0258
Date of Submission: 1/4/2019

Type of Measure:

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

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: Patient Experience of Care Survey Data	☐ other:

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 data used for all measures are from patients who completed the ICH CAHPS Survey in 2017. The ICH CAHPS survey is administered semiannually to patients who received in-center dialysis during quarters 2 (April–June) and 4 (October–December) of the calendar year and is designed to measure their experience of the received dialysis treatment.

1.3 What are the dates of the data used in testing?

The Spring 2017 ICH CAHPS Survey is based on patients who received dialysis from October to December 2016; the Spring 2017 sample was drawn in April 2017. The Fall 2017 ICH CAHPS Survey is based on patients who received dialysis from April to June 2017; the Fall 2017 sample was drawn in October 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 (ICH facility)	☒ hospital/facility/agency (ICH facility)
☐ health plan	☐ health plan
☐ other:	☒ other: State and National

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)*

Approximately 350,000 eligible patients are selected each round of the survey from over 6,000 ICH dialysis facilities. Nearly all of these facilities have a small number of patients and a census is selected each survey period. Fewer than 20 facilities have a large enough patient pool for a simple random sample of 240 patients to be drawn. Thus, this is a "near census" of the population. **Table 1.5** shows the number of facilities per state/territory:

Table 1.5. Frequency of ICH dialysis centers per state or territory

ST	n	%	ST	n	%	ST	n	%	ST	n	%	ST	n	%
AK	9	0.1	**GU**	4	0.1	**ME**	18	0.3	**NJ**	165	2.6	**SD**	27	0.4
AL	163	2.6	**HI**	28	0.4	**MI**	197	3.1	**NM**	46	0.7	**TN**	168	2.7
AR	67	1.1	**IA**	65	1.0	**MN**	115	1.8	**NV**	44	0.7	**TX**	578	9.1
AZ	111	1.8	**ID**	27	0.4	**MO**	145	2.3	**NY**	265	4.2	**UT**	38	0.6
CA	567	8.9	**IL**	248	3.9	**MP**	3	0.1	**OH**	280	4.4	**VA**	171	2.7
CO	70	1.1	**IN**	153	2.4	**MS**	76	1.2	**OK**	79	1.2	**VI**	5	0.1
CT	49	0.8	**KS**	53	0.8	**MT**	14	0.2	**OR**	62	1.0	**VT**	8	0.1
DC	19	0.3	**KY**	116	1.8	**NC**	199	3.1	**PA**	278	4.4	**WA**	92	1.5
DE	25	0.4	**LA**	155	2.4	**ND**	14	0.2	**PR**	45	0.7	**WI**	125	2.0
FL	388	6.1	**MA**	76	1.2	**NE**	29	0.5	**RI**	15	0.2	**WV**	40	0.6
GA	310	4.9	**MD**	140	2.2	**NH**	17	0.3	**SC**	139	2.2	**WY**	9	0.1

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)*

Table 1.6. Sample distributions by demographic characteristics, 2017 ICH CAHPS Spring & Fall survey

Demographic	Spring 2017 Sample		Fall 2017 Sample	
	Number Sampled	Percent Sampled	Number Sampled	Percent Sampled
Age				
18-34	11,426	3.30%	11,386	3.20%
35-49	45,759	13.10%	45,917	13.10%
50-64	117,056	33.60%	118,073	33.60%
65-74	92,787	26.70%	94,424	26.80%
75-84	60,818	17.50%	61,848	17.60%
85+	20,178	5.80%	20,134	5.70%
Total	348,024	100.00%	351,782	100.00%
Gender				
Male	197,170	56.70%	199,931	56.80%
Female	150,854	43.30%	151,851	43.20%
Total	348,024	100.00%	351,782	100.00%
Race				
American Indian/ Alaskan	4,158	1.20%	3,826	1.10%
Asian	14,458	4.20%	13,112	3.70%
Black	120,650	34.70%	111,197	31.60%
White	188,206	54.10%	168,376	47.90%
Unknown	15,445	4.40%	50,559	14.40%
Hawaii / Pacific Islander	5,107	1.50%	4,712	1.30%
Total	348,024	100.00%	351,782	100.00%
Ethnicity				
Hispanic	64,662	18.60%	59,761	17.00%
Non-Hispanic	268,704	77.20%	242,187	68.80%
Unknown	14,658	4.20%	49,834	14.20%
Total	348,024	100.00%	351,782	100.00%

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.

Patient-mix adjustment, which is common on most CAHPS surveys, is used to create a level playing field when making comparisons across facilities. Patient-mix adjustment controls for patient characteristics that impact how patients respond to the survey questions. It reduces the variation between facilities that is due to the demographic profile of the patients they serve. The patient-mix adjustment coefficients are updated twice a year for ICH CAHPS, as we update the data on a rolling basis on the Dialysis Facility Compare at http://www.medicare.gov. The patient-mix adjustments for ICH CAHPS can be found here: https://ichcahps.org/Portals/0/PublicReporting/ICHCAHPS_PublicReportingCoefficientsFall2017Spring2017.docx

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.*

The ICH CAHPS Survey has a set of "About You" questions, some of which are included as patient-mix adjusters that are used for the analyses. The final set of patient mix adjusters (see **section 2b3** for more details) includes 13 patient characteristics plus survey mode. They are:

- Mode of survey administration
- Overall health
- Overall mental health
- Heart disease
- Deaf or serious difficulty hearing
- Blind or serious difficulty seeing
- Difficulty concentrating, remembering, or making decisions
- Difficulty dressing or bathing
- Age
- Sex
- Education
- Does the patient speak a language other than English at home
- Did someone help the patient complete this survey
- Total number of years on dialysis

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)***

Data Recodes

Prior to psychometric analysis, recodes at the item level were necessary to provide response direction consistency in each composite. For all dichotomous (yes/no) questions, the survey value associated with each response is 1/2, which differs in positive response direction from the remaining questions (never = 0 and always = 4). Therefore, the yes/no questions were recoded to 1/0, which preserves direction. In addition, some questions have a value of "99" to indicate nonresponse and were recoded to missing. Finally, Q21 (In the last 3 months, how often did dialysis center staff insert your needles with as little pain as possible?) contains a response that is similar to a "not applicable" category (I insert my own needles). This response was also coded to missing.

Top-box scaling procedures are used in the public reporting results and therefore psychometric confirmation is also needed with this scaling scheme. All analyses conducted with the ICH CAHPS composites with survey format data were also conducted with top-box scaling. Top-box variables represent the most positive responses versus all other responses. The calculations for top-box vary between the ratings measures and the multi-item measures. For the ratings measures, scores of 9 or 10 are considered the most favorable and receive a code of 1; all other scores receive a code of 0. To calculate the top-box codes for the multi-item measures, polytomous data is transformed into dichotomous data, the most positive response was selected to receive a code of one and indicates the top-box. All remaining responses receive a code of zero. Survey format data uses the scales from the questionnaire without any recoding.

Internal Consistency: Composite Alpha

Internal consistency of ICH CAHPS composites is assessed to provide a measure of how consistently the items are measuring each construct. Cronbach's alpha coefficient was used, at the patient level, to assess internal consistency reliability, which has an estimate range from zero to one. The Kuder-Richardson Formula 20 (KR-20), a special case of Cronbach's alpha, was used for composites with dichotomous items. An alpha (or KR-20) estimate of zero indicates no measurement consistency and one indicates perfect consistency. The cutoff criterion for the current examination is 0.80 for adequate reliability. An estimate of 0.90 is preferred, which indicates good consistency.[2] Internal consistency was computed with two other aspects of reliability at the item level: item total correlations and the alpha estimate if the item was removed from the composite. If the item-total correlation is low, or if the alpha with the item removed increases, it is an indicator that the item is not strongly related to the whole composite and may not be perfectly applicable with the other data elements.

Inter-Class Reliability and Intra-class Correlations: Site-level Reliability

Inter-class reliability (ICR) is calculated at the patient level on composite scores and determines how much of the variation in the composites is the result of true variation among dialysis facilities or potentially measurement error. This statistic is computed by fitting a one-way analysis of variance regressing composites on dialysis facilities and then transforming the F-statistic as follows:

$$ICR = (F-1)/F$$

We further investigated facility-level reliability by calculating intraclass correlation coefficients (ICCs) for dialysis facilities using the intracc.sas macro.[3] ICC(1,k) is another method of estimating the ratio of between dialysis facility variability to within-dialysis facility variability when patients differ across sites. The ICC(1,k) is calculated as follows[4]:

$$ICC = \text{(Between-Facility Mean Square} - \text{Within-Facility Mean Square)} / \text{Between-Facility Mean Square}$$

Higher ICR and ICC(1,k) estimates indicate a better ability to differentiate between dialysis facilities. Values of 0.70 or higher indicate acceptable site-level reliability.

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)*

Internal Consistency

Internal consistency estimates (in the title), item-total correlations, and alpha with item removed estimates are presented in two sets of tables. The first set present the survey format results and the second present the top box results. Specifically, **Tables 2a2.3.1a-c** contain the survey format analysis for the Nephrologists'

[2] Nunnally, J. (1978). *Psychometric methods.* New York, NY: McGraw-Hill.

[3] Hamer, R. M. (1990). *Compute six intraclass correlation measures*. Available at http://support.sas.com/kb/25/031.html#ref

[4] Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Communication and Caring (NCC), Quality of Dialysis Center Care and Operations (QDCCO), and Providing Information to Patients (PIP) composites, respectively. **Tables 2a2.3.2a-c** present the top-box coding analysis results for the NCC, QDCCO, and PIP composites.

Table 2a2.3.1a. Nephrologists' Communication and Caring (NCC) composite reliability (standardized values) information—Survey Format analysis (α=0.91)

Composite Item	Item total correlation	Alpha with item removed
Q3. In the last 3 months, how often did your kidney doctors listen carefully to you?	0.77	0.88
Q4. In the last 3 months, how often did your kidney doctors explain things in a way that was easy for you to understand?	0.78	0.88
Q5. In the last 3 months, how often did your kidney doctors show respect for what you had to say?	0.78	0.88
Q6. In the last 3 months, how often did your kidney doctors spend enough time with you?	0.77	0.88
Q7. In the last 3 months, how often did you feel your kidney doctors really cared about you as a person?	0.80	0.88
Q9. Do your kidney doctors seem informed and up-to-date about the health care you receive from other doctors?	0.54	0.92

Table 2a2.3.1b. Quality of Dialysis Center Care and Operations (QDCCO) composite reliability (standardized values) information—Survey Format analysis (α=0.93)

Composite Item	Item total correlation	Alpha with item removed
Q10. In the last 3 months, how often did the dialysis center staff listen carefully to you?	0.77	0.92
Q11. In the last 3 months, how often did the dialysis center staff explain things in a way that was easy for you to understand?	0.76	0.92
Q12. In the last 3 months, how often did the dialysis center staff show respect for what you had to say?	0.79	0.92
Q13. In the last 3 months, how often did the dialysis center staff spend enough time with you?	0.78	0.92
Q14. In the last 3 months, how often did you feel the dialysis center staff really cared about you as a person?	0.81	0.92
Q15. In the last 3 months, how often did dialysis center staff make you as comfortable as possible during dialysis?	0.77	0.93
Q16. In the last 3 months, did dialysis center staff keep information about you and your health as private as possible from other patients?	0.46	0.93
Q17. In the last 3 months, did you feel comfortable asking the dialysis center staff everything you wanted about dialysis care?	0.60	0.93
Q21. In the last 3 months, how often did dialysis center staff insert your needles with as little pain as possible?	0.51	0.93
Q22. In the last 3 months, how often did dialysis center staff check you as closely as you wanted while you were on the dialysis machine?	0.72	0.92
Q24. In the last 3 months, how often was the dialysis center staff able to manage problems during your dialysis?	0.64	0.93

Composite Item	Item total correlation	Alpha with item removed
Q25. In the last 3 months, how often did dialysis center staff behave in a professional manner?	0.74	0.92
Q26. In the last 3 months, did dialysis center staff talk to you about what you should eat and drink?	0.29	0.93
Q27. In the last 3 months, how often did dialysis center staff explain blood test results in a way that was easy to understand?	0.56	0.93
Q33. In the last 3 months, when you arrived on time, how often did you get put on the dialysis machine within 15 minutes of your appointment or shift time?	0.50	0.93
Q34. In the last 3 months, how often was the dialysis center as clean as it could be?	0.52	0.93
Q43. In the last 12 months, how often were you satisfied with the way they handled these problems?	0.64	0.93

Table 2a2.3.1c. Providing Information to Patients (PIP) composite reliability (standardized values) information—Survey Format analysis (α=0.73)

Composite Item	Item total correlation	Alpha with item removed
Q19. The dialysis center staff can connect you to the dialysis machine through a graft, fistula, or catheter. Do you know how to take care of your graft, fistula or catheter?	0.30	0.73
Q28. As a patient you have certain rights. For example, you have the right to be treated with respect and the right to privacy. Did this dialysis center ever give you any written information about your rights as a patient?	0.42	0.71
Q29. Did dialysis center staff at this center ever review your rights as a patient with you?	0.51	0.69
Q30. Has dialysis center staff ever told you what to do if you experience a health problem at home?	0.46	0.70
Q31. Has any dialysis center staff ever told you how to get off the machine if there is an emergency at the center?	0.26	0.73
Q36. You can treat kidney disease with dialysis, kidney transplant or with dialysis at home. In the last 12 months, did your kidney doctors or dialysis center staff talk to you as much as you wanted about which treatment is right for you?	0.52	0.69
Q38. In the last 12 months, has a doctor or dialysis center staff explained to you why you are not eligible for a kidney transplant?	0.37	0.72
Q39. Peritoneal dialysis is dialysis given through the belly and is usually done at home. In the last 12 months, did either your kidney doctors or dialysis center staff talk to you about peritoneal dialysis?	0.37	0.72
Q40. In the last 12 months, were you as involved as much as you wanted in choosing the treatment that is right for you?	0.45	0.70

Table 2a2.3.2a. Nephrologists' Communication and Caring composite reliability (standardized values) information—Top-Box analysis (α=0.86)

Composite Item	Item total correlation	Alpha with item removed
Q3. In the last 3 months, how often did your kidney doctors listen carefully to you?	0.71	0.83
Q4. In the last 3 months, how often did your kidney doctors explain things in a way that was easy for you to understand?	0.72	0.83
Q5. In the last 3 months, how often did your kidney doctors show respect for what you had to say?	0.71	0.83
Q6. In the last 3 months, how often did your kidney doctors spend enough time with you?	0.66	0.84
Q7. In the last 3 months, how often did you feel your kidney doctors really cared about you as a person?	0.73	0.83
Q9. Do your kidney doctors seem informed and up-to-date about the health care you receive from other doctors?	0.44	0.88

Table 2a2.3.2b. Quality of Dialysis Center Care and Operations composite reliability (standardized values) information—Top-box analysis (α=0.90)

Composite Item	Item total correlation	Alpha with item removed
Q10. In the last 3 months, how often did the dialysis center staff listen carefully to you?	0.69	0.89
Q11. In the last 3 months, how often did the dialysis center staff explain things in a way that was easy for you to understand?	0.68	0.89
Q12. In the last 3 months, how often did the dialysis center staff show respect for what you had to say?	0.70	0.89
Q13. In the last 3 months, how often did the dialysis center staff spend enough time with you?	0.68	0.89
Q14. In the last 3 months, how often did you feel the dialysis center staff really cared about you as a person?	0.73	0.88
Q15. In the last 3 months, how often did dialysis center staff make you as comfortable as possible during dialysis?	0.70	0.89
Q16. In the last 3 months, did dialysis center staff keep information about you and your health as private as possible from other patients?	0.34	0.90
Q17. In the last 3 months, did you feel comfortable asking the dialysis center staff everything you wanted about dialysis care?	0.44	0.89
Q21. In the last 3 months, how often did dialysis center staff insert your needles with as little pain as possible?	0.41	0.90
Q22. In the last 3 months, how often did dialysis center staff check you as closely as you wanted while you were on the dialysis machine?	0.64	0.89
Q24. In the last 3 months, how often was the dialysis center staff able to manage problems during your dialysis?	0.57	0.89
Q25. In the last 3 months, how often did dialysis center staff behave in a professional manner?	0.64	0.89

Composite Item	Item total correlation	Alpha with item removed
Q26. In the last 3 months, did dialysis center staff talk to you about what you should eat and drink?	0.23	0.90
Q27. In the last 3 months, how often did dialysis center staff explain blood test results in a way that was easy to understand?	0.53	0.89
Q33. In the last 3 months, when you arrived on time, how often did you get put on the dialysis machine within 15 minutes of your appointment or shift time?	0.39	0.90
Q34. In the last 3 months, how often was the dialysis center as clean as it could be?	0.47	0.89
Q43. In the last 12 months, how often were you satisfied with the way they handled these problems?	0.53	0.89

Table 2a2.3.2c. Providing Information to Patients composite reliability (standardized values) information—Top-box analysis (α=0.73)

Composite Item	Item total correlation	Alpha with item removed
Q19. The dialysis center staff can connect you to the dialysis machine through a graft, fistula, or catheter. Do you know how to take care of your graft, fistula or catheter?	0.30	0.73
Q28. As a patient you have certain rights. For example, you have the right to be treated with respect and the right to privacy. Did this dialysis center ever give you any written information about your rights as a patient?	0.42	0.71
Q29. Did dialysis center staff at this center ever review your rights as a patient with you?	0.51	0.69
Q30. Has dialysis center staff ever told you what to do if you experience a health problem at home?	0.46	0.70
Q31. Has any dialysis center staff ever told you how to get off the machine if there is an emergency at the center?	0.26	0.73
Q36. You can treat kidney disease with dialysis, kidney transplant or with dialysis at home. In the last 12 months, did your kidney doctors or dialysis center staff talk to you as much as you wanted about which treatment is right for you?	0.52	0.69
Q38. In the last 12 months, has a doctor or dialysis center staff explained to you why you are not eligible for a kidney transplant?	0.37	0.72
Q39. Peritoneal dialysis is dialysis given through the belly and is usually done at home. In the last 12 months, did either your kidney doctors or dialysis center staff talk to you about peritoneal dialysis?	0.37	0.72
Q40. In the last 12 months, were you as involved as much as you wanted in choosing the treatment that is right for you?	0.45	0.70

Inter-Class Reliability and Intra-class Correlations

Table 2a2.3.3 provides the ICR and ICC(1,k) results for each composite for both analysis waves (survey format and top-box). With an average of 49 patient respondents per site (SD = 27), ICR and ICC(1,k) analyses were constrained to facilities that had 30 respondents or more. This results in more robust estimates per facility.

Table 2a2.3.3. ICH CAHPS ICR and ICC—Site-level reliability with facility respondents ≥ 30

	ICR	ICC
Survey Format Analysis		
Nephrologists' Communication and Caring	0.708	0.709
Quality of Dialysis Center Care and Operations	0.729	0.727
Providing Information to Patients	0.647	0.647
Global rating of kidney doctors	0.682	0.681
Global rating of dialysis center staff	0.702	0.702
Global rating of dialysis center	0.738	0.737
Top-Box Analysis		
Nephrologists' Communication and Caring	0.688	0.687
Quality of Dialysis Center Care and Operations	0.692	0.691
Providing Information to Patients	0.647	0.647
Global rating of kidney doctors	0.649	0.648
Global rating of dialysis center staff	0.677	0.676
Global rating of dialysis center	0.709	0.708

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?)*

Internal Consistency

Two of the three composites (NCC & QDCCO) in the ICH CAHPS show adequate internal consistency estimates or better (alpha = 0.85 at minimum) when looking at the survey format, or the top-box coded data elements (items). However, one composite (PIP) falls short of Nunnally's[5] criterion of 0.80 in the KR-20 estimates. Consultation with additional standards for scientific acceptability indicates that for established scales, 0.7 is indeed acceptable (COSMIN)[6] for more established data elements. Therefore, the reliability evidence for all three ICH CAHPS composites, using internal consistency as an evaluative tool, is acceptable.

Inter-Class Reliability and Intra-class Correlations

All ICR and ICC(1,k) estimates are close to or above the critical cutoff of 0.7, with the exception of Providing Information to Patients (PIP) in both the survey format and the top-box analyses and the global rating of kidney doctors in the top box format. In other words, two of the three composites (NCC & QDCCO) and the three global ratings in the ICH CAHPS show adequate site-level reliability when looking at the survey format coding. The estimates dip in the top-box, which is not surprising given the reduction in variability from a 4-response o4 10-response scale to dichotomous. PIP and the global rating of kidney doctors, however, fall below the 0.7 cutoff, even with rounding considerations. The minimum patient population per facility for analysis may need to be increased for the PIP composite and global rating of kidney doctors. When ICR is examined with facilities of at least 60 patient respondents, the estimates are above the 0.7 cutoff. Therefore, when minimum patient respondent sample per facility is adjusted to provide the most robust estimates, the site-level reliability approaches or exceeds acceptability.

---

[5] Nunnally, J. (1978). *Psychometric methods.* New York, NY: McGraw-Hill.

[6] Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157.

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 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*) **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)*

Validity Evidence Based on Internal Structure: Confirmatory Factor Analysis

The composite structure established previously[7] was reevaluated using confirmatory factor analysis (CFA) on patient response data. CFA assesses the theoretical structure of the composites (e.g., user defined) against the underlying data structure. That is, CFA examines the relationships among the survey items to confirm alignment with the conceptual model structure, as defined by the research team or subject matter expert. More specifically, the CFA analysis evaluates the way the composites were originally defined and determines if those definitions still make sense in the data through fit statistics, which are presented in Table 2b1.3.1a for direct comparison with the analytic estimates. Analyses were conducted using the Mplus[8] analytic software because of the software's flexibility to handle multiple item formats including continuous, dichotomous, ordered polytomous (ordinal), and data that contain missing values.

Validity Evidence Based on Relationships With Conceptually Related Constructs: Composite and Global Rating Relationship Analysis

A mechanism to evaluate evidence for the validity of composite functioning is to determine how well the composites relate to summary items, such as the global ratings of doctors (Q8), dialysis center staff (Q32), and dialysis center (Q35). Therefore, correlations between each global rating item and all three of the composites were calculated. This analysis was conducted at the patient and the facility level and in the two specified coding schemes (survey format and top-box).

Validity Evidence Based on Internal Structure: Composite Intercorrelations

Composite correlations were examined to further evaluate that the designed item-composite structure is measuring constructs as intended. Large correlations (0.8 or greater), may indicate that the internal structure established through factor analysis is not a valid representation of the data. Therefore, patient respondent and facility level data were used in composite correlations in the survey format and top-box coding schemes.

Validity Evidence Based on Discrimination Among Composites: Item-Level Correlations With Composite Totals

The final validity investigation for evaluating ICH CAHPS composite functioning was to determine how items relate to their own composite total versus other composites. Therefore, the intent is to validate the composites through discrimination, specifically by calculating correlations between all items and all composites (NCC,

[7] Weidmer, B. A., Cleary, P. D., Keller, S., Evensen, C., Hurtado, M. P., Kosiak, B., Gallagher, P. M., Levine, R., & Hays, R. D. (2014). Development and evaluation of the CAHPS (Consumer Assessment of Healthcare Providers and Systems) Survey for in-center hemodialysis patients. *American Journal of Kidney Diseases, 64*(5), 753-760.

[8] Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's guide*. 8th Ed. Los Angeles, CA: Muthén & Muthén.

QDCCO and PIP). If an item's correlation with a different composite is greater than the correlation with its own, it is an indicator that composites may not be discriminately valid.

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

Confirmatory Factor Analysis

The ICH CAHPS composite structure was assessed with confirmatory factor model in both the survey format and top-box analysis. Fit statistics from the model are provided in the center column of **Tables 2b1.3.1a-b** with critical levels for determining good fit referenced in the last column.

Table 2b1.3.1a. ICH CAHPS confirmatory factor analysis fit statistics—Survey format analysis

Assessment criteria	Analytic value	Acceptable value
Root Mean Square Error of Approximation (RMSEA)	0.039	0.05 or less[9]
Comparative Fit Index (CFI)	0.983	0.90 at a minimum[10]
Tucker-Lewis index (TLI)	0.982	0.90 at a minimum historically, 0.95 indicates good fit

Table 2b1.3.1b. ICH CAHPS confirmatory factor analysis fit statistics—Top-box analysis

Assessment criteria	Analytic value	Acceptable value
Root Mean Square Error of Approximation (RMSEA)	0.038	0.05 or less[8] (Browne & Cudeck, 1993)
Comparative Fit Index (CFI)	0.980	0.90 at a minimum[9]
Tucker-Lewis index (TLI)	0.979	0.90 at a minimum historically, 0.95 indicates good fit

Composite and Global Rating Relationship Analysis

Tables 2b1.3.2a and 2b1.3.2b show composite correlations with each global rating item for the survey format coding (**Table 2b1.3.2a**) and the top-box coding (**Table 2b1.3.2b**) for data at the patient and facility levels. If the correlation with the global rating items is low, particularly between two that would be expected to have a strong relationship (e.g., Quality of Dialysis Center Care and Operations [QDCCO] and Global Rating of Dialysis Center Staff) then the composite may not be externally valid.

Table 2b1.3.2a. ICH CAHPS composites (survey format analysis) correlations with global ratings

Composite	Global rating of kidney doctors	Global rating of dialysis center staff	Global rating of dialysis center
Patient Level			
Nephrologists' Communication and Caring	0.77	0.45	0.44
Quality of Dialysis Center Care and Operations	0.48	0.77	0.73
Providing Information to Patients	0.34	0.44	0.40
Facility Level			
Nephrologists' Communication and Caring	0.86	0.51	0.48
Quality of Dialysis Center Care and Operations	0.47	0.86	0.82
Providing Information to Patients	0.37	0.55	0.53

[9] Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & K. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park: Sage.

[10] Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Table 2b1.3.2b. ICH CAHPS composites (top-box analysis) correlations with global ratings

Composite	Global rating of kidney doctors	Global rating of dialysis center staff	Global rating of dialysis center
Patient Level			
Nephrologists' Communication and Caring	0.70	0.44	0.43
Quality of Dialysis Center Care and Operations	0.44	0.70	0.66
Providing Information to Patients	0.34	0.44	0.40
Facility Level			
Nephrologists' Communication and Caring	0.78	0.47	0.45
Quality of Dialysis Center Care and Operations	0.45	0.79	0.76
Providing Information to Patients	0.36	0.52	0.49

Composite intercorrelations

Tables 2b1.3.3a and 2b1.3.3b show composite intercorrelations to determine if any composite scores lack distinction from each other. The correlations are moderate (0.4-0.5) but not strong enough to have concerns about the composite structure based on previous factor analysis findings.

Table 2b1.3.3a. ICH CAHPS correlations between composites (survey format analysis)

Composite	NCC	QDCCO	PIP
Patient Level			
Nephrologists' Communication and Caring	1		
Quality of Dialysis Center Care and Operations	0.57	1	
Providing Information to Patients	0.40	0.52	1
Facility Level			
Nephrologists' Communication and Caring	1		
Quality of Dialysis Center Care and Operations	0.54	1	
Providing Information to Patients	0.43	0.6	1

Table 2b1.3.3b. ICH CAHPS correlations between composites (top-box analysis)

Composite	NCC	QDCCO	PIP
Patient Level			
Nephrologists' Communication and Caring	1		
Quality of Dialysis Center Care and Operations	0.59	1	
Providing Information to Patients	0.38	0.45	1
Facility Level			
Nephrologists' Communication and Caring	1		
Quality of Dialysis Center Care and Operations	0.54	1	
Providing Information to Patients	0.38	0.54	1

Item-Level Correlations With Composite Totals

Table 2b1.3.4 shows all the ICH CAHPS items in the NCC, QDCCO and PIP composites, the item correlation to their own composite total (bolded), and the item's relationship to the other two composites.

Table 2b1.3.4. ICH CAHPS discriminant validity – correlation with own composite total versus other composites

	NCC correlation		QDCCO correlation		PIP Correlation	
	Survey Format	**Top-Box Coding**	**Survey Format**	**Top-Box Coding**	**Survey Format**	**Top-Box Coding**
Nephrologists' Communication and Caring (NCC)						
Q3. In the last 3 months, how often did your kidney doctors listen carefully to you?	**0.85**	**0.82**	0.46	0.45	0.33	0.28
Q4. In the last 3 months, how often did your kidney doctors explain things in a way that was easy for you to understand?	**0.86**	**0.82**	0.49	0.48	0.38	0.31
Q5. In the last 3 months, how often did your kidney doctors show respect for what you had to say?	**0.86**	**0.81**	0.50	0.48	0.34	0.29
Q6. In the last 3 months, how often did your kidney doctors spend enough time with you?	**0.86**	**0.79**	0.51	0.48	0.36	0.28
Q7. In the last 3 months, how often did you feel your kidney doctors really cared about you as a person?	**0.87**	**0.83**	0.53	0.51	0.36	0.30
Q9. Do your kidney doctors seem informed and up-to-date about the health care you receive from other doctors?	**0.60**	**0.57**	0.37	0.32	0.32	0.32
Quality of Dialysis Center Care and Operations (QDCCO)						
Q10. In the last 3 months, how often did the dialysis center staff listen carefully to you?	0.47	0.46	**0.82**	**0.79**	0.39	0.32
Q11. In the last 3 months, how often did the dialysis center staff explain things in a way that was easy for you to understand?	0.51	0.49	**0.81**	**0.77**	0.44	0.35
Q12. In the last 3 months, how often did the dialysis center staff show respect for what you had to say?	0.49	0.48	**0.84**	**0.80**	0.40	0.33

	NCC correlation		QDCCO correlation		PIP Correlation	
	Survey Format	Top-Box Coding	Survey Format	Top-Box Coding	Survey Format	Top-Box Coding
Quality of Dialysis Center Care and Operations (continued)						
Q13. In the last 3 months, how often did the dialysis center staff spend enough time with you?	0.49	0.48	**0.82**	**0.77**	0.42	0.33
Q14. In the last 3 months, how often did you feel the dialysis center staff really cared about you as a person?	0.48	0.47	**0.85**	**0.80**	0.41	0.34
Q15. In the last 3 months, how often did dialysis center staff make you as comfortable as possible during dialysis?	0.44	0.44	**0.81**	**0.78**	0.38	0.32
Q16. In the last 3 months, did dialysis center staff keep information about you and your health as private as possible from other patients?	0.25	0.21	**0.43**	**0.36**	0.26	0.26
Q17. In the last 3 months, did you feel comfortable asking the dialysis center staff everything you wanted about dialysis care?	0.33	0.28	**0.57**	**0.46**	0.36	0.36
Q21. In the last 3 months, how often did dialysis center staff insert your needles with as little pain as possible?	0.30	0.29	**0.57**	**0.56**	0.26	0.20
Q22. In the last 3 months, how often did dialysis center staff check you as closely as you wanted while you were on the dialysis machine?	0.44	0.44	**0.78**	**0.75**	0.40	0.31
Q24. In the last 3 months, how often was the dialysis center staff able to manage problems during your dialysis?	0.40	0.38	**0.72**	**0.68**	0.37	0.31
Q25. In the last 3 months, how often did dialysis center staff behave in a professional manner?	0.38	0.39	**0.75**	**0.74**	0.35	0.31

	NCC correlation		QDCCO correlation		PIP Correlation	
	Survey Format	Top-Box Coding	Survey Format	Top-Box Coding	Survey Format	Top-Box Coding
Quality of Dialysis Center Care and Operations (continued)						
Q26. In the last 3 months, did dialysis center staff talk to you about what you should eat and drink?	0.21	0.19	**0.35**	**0.33**	0.35	0.35
Q27. In the last 3 months, how often did dialysis center staff explain blood test results in a way that was easy to understand?	0.42	0.44	**0.66**	**0.65**	0.47	0.38
Q33. In the last 3 months, when you arrived on time, how often did you get put on the dialysis machine within 15 minutes of your appointment or shift time?	0.32	0.29	**0.60**	**0.54**	0.31	0.21
Q34. In the last 3 months, how often was the dialysis center as clean as it could be?	0.32	0.34	**0.61**	**0.61**	0.27	0.22
Q43. In the last 12 months, how often were you satisfied with the way they handled these problems?	0.41	0.37	**0.70**	**0.65**	0.41	0.30
Providing Information to Patients (PIP)						
Q19. The dialysis center staff can connect you to the dialysis machine through a graft, fistula, or catheter. Do you know how to take care of your graft, fistula or catheter?	0.18	0.16	0.22	0.17	**0.47**	**0.47**
Q28. As a patient you have certain rights. For example, you have the right to be treated with respect and the right to privacy. Did this dialysis center ever give you any written information about your rights as a patient?	0.24	0.21	0.35	0.28	**0.59**	**0.59**
Q29. Did dialysis center staff at this center ever review your rights as a patient with you?	0.30	0.28	0.43	0.39	**0.68**	**0.68**

	NCC correlation		QDCCO correlation		PIP Correlation	
	Survey Format	**Top-Box Coding**	**Survey Format**	**Top-Box Coding**	**Survey Format**	**Top-Box Coding**
Q30. Has dialysis center staff ever told you what to do if you experience a health problem at home?	0.28	0.26	0.39	0.34	**0.63**	**0.63**
Q31. Has any dialysis center staff ever told you how to get off the machine if there is an emergency at the center?	0.14	0.13	0.16	0.14	**0.43**	**0.43**
Q36. You can treat kidney disease with dialysis, kidney transplant or with dialysis at home. In the last 12 months, did your kidney doctors or dialysis center staff talk to you as much as you wanted about which treatment is right for you?	0.36	0.34	0.42	0.37	**0.69**	**0.69**
Q38. In the last 12 months, has a doctor or dialysis center staff explained to you why you are not eligible for a kidney transplant?	0.21	0.20	0.23	0.20	**0.58**	**0.58**
Providing Information to Patients (continued)						
Q39. Peritoneal dialysis is dialysis given through the belly and is usually done at home. In the last 12 months, did either your kidney doctors or dialysis center staff talk to you about peritoneal dialysis?	0.17	0.18	0.20	0.20	0.60	0.60
Q40. In the last 12 months, were you as involved as much as you wanted in choosing the treatment that is right for you?	0.30	0.27	0.34	0.29	0.62	0.62

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?*)

Confirmatory Factor Analysis

The ICH CAHPS composite structure was confirmed with confirmatory factor analysis in both coding instances (survey format and top-box). Fit statistics were well above critical levels for both analyses and we conclude that the validity evidence for internal structure is sound.

Composite and Global Rating Relationship Analysis

The ICH CAHPS correlational analysis between the composites (NCC, QDCCO, and PIP) and the global ratings that patients give to the Nephrologists, Center staff, and the Center itself were examined for construct validity. Given the large sample size in the 2017 data, all correlations are statistically significant, and most are of the moderate to strong range (0.4 or better). Most of the expected relationships (e.g., QDCCO with Global Rating of Dialysis Center Staff, NCC with Global Rating of Kidney Doctors) show correlations of 0.7 or better. Therefore, the ICH CAHPS composites show good evidence of validity concerning conceptually related construct.

Composite Intercorrelations

Correlations among the ICH CAHPS composites (NCC, QDCCO, and PIP) are moderate (0.4-0.5) but not strong enough to have concerns about composite structure integrity based on previous factor analysis findings. Therefore, the ICH CAHPS items measure three distinct constructs, as theorized by subject matter experts, established through factor analysis, and confirmed in the correlation analysis described.

Item-Level Correlations With Composite Totals

The assessment of validity evidence by discrimination indicates that whether in the survey format or top-box coding schemes, the items are well targeted within their composites. For example, the QDCCO items were found to correlate higher with the QDCCO total score than with either the NCC or PIP total scores. This pattern of stronger relationships with the item's specified composite was replicated with the NCC and PIP composites. Therefore, the ICH CAHPS items show good evidence of validity by discrimination.

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — skip to section 2b4

No subgroups are excluded from the data analysis. There are exclusions on the sampling frame, but this is so that the population of interest can be targeted—adults who have received an in-center hemodialysis treatment for at least 3 months. Therefore, individuals with any of the following criteria are not included in the study:

- receive home dialysis or peritoneal dialysis;
- are institutionalized (live in nursing home, in jail or prison);
- are under age 18 years of age; or
- have not received hemodialysis for 3 months or longer.

Q1 and Q2 of the survey instrument verifies patient status. Should they receive a different type of care, they are instructed to skip to Q45 at the end of the survey where basic demographic data are collected. Because these individuals are considered outside the scope of the study, we have marked the box in 2b2 as "no exclusions."

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)*

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)*

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)*

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 2b5.

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☒ **Other,** Patient-mix adjustment is used.

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.

The ICH CAHPS Survey mode experiment was conducted in 2014 using the same data collection methods currently being used in the national implementation of the survey. The sample frame for the mode experiment consisted of all patients who met survey eligibility criteria—patients *must* have been 18 years of age or older at the end of the 3-month sampling window and must have received in-center dialysis care from their current ICH facility for 3 months or longer. Those known to be deceased were excluded from the sample.

The goal of the sampling process for the mode experiment was to obtain approximately 1,570 completed surveys for each of the three approved data collection modes (mail-only, telephone-only, mixed mode of mail with phone follow-up). Given expected response rates for each mode, 4,800 hemodialysis patients were sampled for the mail-only mode, 3,733 for the telephone-only mode, and 3,360 were included in the mixed-mode sample. A total of 11,893 eligible patients were sampled from the ICH patient population. After the sample was selected, sampled patients were randomly assigned to one of the three data collection modes using the inverse of the estimated response rates. The number of respondents for each mode allowed a 5-percentage point difference to be detected with 80% power with an alpha-level of 0.05.

A total of 3,557 surveys were completed during the mode experiment, including 1,355 from sample patients in the mail-only mode, 994 for the telephone-only mode, and 1,208 for the mixed mode. Sample patients who reported during the data collection period that they receive home dialysis; those who were deceased, institutionalized, or receiving hospice care; and those who could not participate because the survey was not offered in their language were deemed to be ineligible for the survey. After adjusting for sampled patients found to be ineligible for the

survey, an overall response rate of 33.6% was achieved. The response rates for the mail-only and telephone-only modes were similar at 30.4% and 30.7%, respectively. The response rate for the mixed-mode was higher at 41.8%.

Data from the mode experiment were analyzed and included two types of variables: dependent variables that represented the patients' experiences with the care they received from ICH facilities and independent variables that represented the patient characteristics that may affect the dependent variables. The dependent variables included three variables calculated from individual ICH CAHPS Survey questions (the global ratings items) and 32 variables that are the multiple survey questions that comprise the composite measures. These dependent variables were:

- Global rating of nephrologist (calculated from survey Q8)
- Global rating of dialysis center staff (calculated from survey Q32)
- Global rating of dialysis center (calculated from survey Q35)
- Six questions that comprise the Nephrologists' communication and caring composite (Qs 3, 4, 5, 6, 7, and 9)
- Seventeen questions that comprise the Quality of dialysis center and operations composite (10, 11, 12, 13, 14, 15, 16, 17, 21, 22, 24, 25, 26, 27, 33, 34, and 43)
- Nine questions that comprise the Providing information to patients composite (calculated from survey Qs 19, 28, 29, 30, 31, 36, 38, 39, and 40)

For each dependent variable, two dichotomous variables were created: the most positive responses versus all other responses, referred to as the "top-box," and the least positive responses versus all other responses, referred to as the "bottom-box," for a total of 35 x 2 = 70 dependent variables. These 70 dependent variables were analyzed using Ordinary Least Squares regression models.

The results of the analysis of data from the ICH CAHPS Survey mode experiment showed significant differences in patients' ratings and assessment of their hemodialysis care based on survey mode and in responses to the survey items that are attributable to patient-mix characteristics. A total of 13 patient-mix characteristics and survey mode were found to be statistically significant in at least one of the regression models. The 14 adjusters (13 patient characteristics plus survey mode) include the following:

- Mode of survey administration
- Overall health
- Overall mental health
- Heart disease
- Deaf or serious difficulty hearing
- Blind or serious difficulty seeing
- Difficulty concentrating, remembering, or making decisions
- Difficulty dressing or bathing
- Age
- Sex

- Education
- Does the patient speak a language other than English at home
- Did someone help the patient complete this survey
- Total number of years on dialysis

During each ICH CAHPS public reporting period, CMS and its Coordination Team uses data from the two most recent semiannual ICH CAHPS Surveys to derive the 13 patient-mix adjustment factors using coefficients obtained from Ordinary Least Squares regression models for the top- and bottom-box scores for each of the three global ratings and the three composite measures. Patient-mix adjustment factors are calculated directly from these regression coefficients for each individual survey item by multiplying the coefficients by negative one (−1.0). The coefficient used to adjust for survey mode is based on the results of the mode experiment. CMS uses the coefficients to adjust the raw scores calculated on each measure from data collected in each semiannual survey. The ICH CAHPS scores that are publicly reported are the weighted average of the two most recent semiannual ICH CAHPS scores.

Calculating the Patient-Mix Adjusted Global Ratings and Composite Scores

Four sets of numbers are needed to calculate an ICH facility's adjusted score for the three individual global ratings (rating of nephrologist, the dialysis center staff, and dialysis center) and the individual survey questions included in each of the three composite measures. These are (1) the "raw score," or the ICH facility's mean on the respective ICH CAHPS outcome before adjustment (top- or bottom-box score for the global ratings and individual survey questions comprising the composites); (2) the national-level patient-mix adjustment factors shown in **Tables 2b3.1.1.a** and **2b3.1.1.b** (top- and bottom-box adjustment factors for the global ratings and individual survey questions comprising the composites); (3) the ICH facility's means on the patient-mix characteristics variables; and (4) the national mean on the patient-mix characteristics variables shown in **Table 2b3.1.1.d**.

The adjusted score for the ratings questions and a given individual survey question that is included in one of the three ICH CAHPS Survey composite measures is the sum of a series of products in the equation shown below, where each product multiplies the adjustment from **Table 2b3.1.1.a**. (top-box) and **Table 2b3.1.1.b**. (bottom-box) by the deviation of the ICH facility's mean on a given patient-mix characteristic from the national mean on that characteristic from **Table 2b3.1.1.d**.

$$y' = y + a1(h1 - m1) + a2(h2 - m2) + a3(h3 - m3) + \ldots + a28(h28 - m28) + a29*h29 + a30*h30$$

where

y'	is the facility's adjusted score (top- or bottom-box) for a ratings question or the individual ICH CAHPS question included in the composite.
y	is the facility's "raw score," or mean on the respective unadjusted top- or bottom-box ICH CAHPS ratings question or question included in the composite.
a1 to a28	are the national-level patient-mix characteristic adjustments, for the global ratings questions and individual questions that comprise the composites. **Tables 2b3.1.1.a and 2b3.1.1.b** show the adjustments for these patient-mix characteristics for the top-box and bottom-box scores, respectively. The

adjustments for the patient-mix characteristics in the tables are expressed as a proportion rather than as a percentage.

a29 to a30 are the national-level survey mode adjustments for the global ratings questions and the individual questions that comprise the composites. **Tables 2b3.1.1.a and 2b3.1.1.b** show the adjustments for survey mode for the top-box and bottom-box scores, respectively. The adjustment for survey mode in the tables are expressed as a proportion rather than as a percentage.

h1 to h28 are the facility's mean proportions of patients with each of the patient characteristics in the same row.

h29 to h30 are the facility's proportion for a given mode. This value will always be 0 or 1 because within a given facility all surveys are completed by either phone, mail, or mixed mode.

m1 to m28 are the national mean proportions of patients with each of the patient characteristics in **Table 2b3.1.1.d** across the facility's participating in ICH CAHPS.

The facility's patient-mix adjusted scores for the ratings questions or an individual survey question, as described in the formula above, are adjusted for differences between a facility's patient composition according to the ICH CAHPS patient-mix characteristics and the overall national composition of ICH patients on these same characteristics. This adjustment, which allows consumers to compare different ICH facilities based on the same overall patient composition, is made by subtracting the national mean—the "m's" in the equation above—for a given patient characteristic from an ICH facility's share of patients on this same patient characteristic—the "h's" in the equation above—and then multiplying the difference by the patient-mix adjustment factor—the "a's" in the equation above.

After each facility's patient-mix adjusted score is created for the ratings questions and individual survey questions, the facility-level composite scores are formed from the average of these facility-level adjusted scores for the individual survey questions that comprise a given composite. This creates the semiannual patient-mix facility-level ratings and composite scores. The two most recent semiannual patient-mix facility-level composite scores are then averaged to produce the current ICH CAHPS scores that are publicly reported.

For public reporting purposes, the final adjusted ICH CAHPS score is rounded to the nearest integer and expressed as a percentage (e.g., 70%). Note that middle-box scores are computed by subtracting the sum of patients who provided top- and bottom-box scores from 100.

Star Ratings

In 2018, the ICH CAHPS team reanalyzed the data from the 2014 mode experiment using linearized means as the dependent variables. This reanalysis was done to accommodate the newly reported linearized means and star ratings. The final results of this analysis determined that the original set of 13 patient-mix characteristics plus survey mode are also appropriate for the statistical adjustment of the linearized means and star ratings.

Star ratings are a supplement to the top- and bottom-box measures and make it easier for consumers to spotlight excellence in health care quality on the DFC. Star ratings are generated for each of the three publicly reported ICH CAHPS Survey global ratings (rating of the kidney doctors (nephrologists), dialysis center staff, and dialysis center) and three composite measures (kidney doctors' communication and caring, quality of dialysis center and operations, and

providing information). Additionally, an overall ICH CAHPS Survey summary star rating is calculated and shown on the DFC for each dialysis facility. The Survey summary star rating is a simple average of the six individual star ratings.

Methods for Calculating Star Ratings

There are two main steps in calculating star ratings: (a) constructing a linear mean for each global rating and composite, and (b) conducting a cluster analysis and grouping the linear means into five categories (i.e., the star ratings). This methodology is described in further detail below.

a. Construction and Adjustment of ICH CAHPS Linear Scores

The responses to the survey items used in each ICH CAHPS measure are converted to a 0–100 linear-scaled score in the following manner:

- For ICH CAHPS Survey global ratings (Survey items 8, 32, and 35):
  - Overall Rating "0" = 0; Overall Rating "1" = 10; Overall Rating "2" = 20; …; Overall Rating "10" = 100
- For ICH CAHPS Survey items 9, 16, 17, 19, 26, 28-31, 36, and 38-40:
  - "No" = 0; and "Yes" = 100
- For ICH CAHPS Survey items 3-7, 10-15, 21, 22, 24, 25, 27, 33, 34, and 43:
  - 1 = 0; 2 = 33 1/3; 3 = 66 2/3; and 4 = 100

The 0–100 linear-scaled ICH CAHPS scores are statistically adjusted for data collection mode and for patient-mix to account for the tendency of certain patient subgroups to respond more positively or negatively to the ICH CAHPS Survey based on data collection mode and specific patient characteristics. The steps directly parallel the process used for adjusting top- and bottom-box scores. The primary difference in this step is the independent variable. Whereas top- and bottom-box scores only allow for two values (0 or 100), the linear scores have a range of values between 0 and 100. The patient-mix adjustment factors for the 2017 ICH CAHPS Surveys are shown in **Table 2b3.1.1.c**.

Averages of ICH CAHPS linear scores across two survey periods are rounded to integer values using standard rounding rules, as follows:

- Let X represent the unrounded two-period average for an ICH CAHPS linear score.
- If X is less than [X.5], then round down to nearest whole integer.
- If X is equal to or greater than [X.5], then round up to nearest whole integer.

b. Conversion of Linear Scores Into ICH CAHPS Star Ratings

After the ICH CAHPS scores are linearized, adjusted, and rounded, we group the scores into one, two, three, four, or five whole stars (only whole stars will be assigned; partial stars will not be used) for each of the six ICH CAHPS measures by applying statistical methods that use relative distribution and clustering. We determine the star rating for each of the six ICH CAHPS measures by applying a clustering algorithm to the individual measure scores. Conceptually, the clustering algorithm identifies the "gaps" in the data and creates five categories (one for each star rating) such that scores of facilities in the same score category (star rating) are as similar as possible, and scores of facilities in different categories are as different as possible. The clustering algorithm that we use is the same one used by CMS to determine star ratings for most of the

Medicare Part C and Part D measures, the Home Health CAHPS (HHCAHPS) Survey, and the Hospital CAHPS Survey.

The goal of the clustering algorithm is to minimize the differences within each cluster and maximize the differences between each cluster. The variance in measure scores is separated into within- and between-cluster sum of squares components. The algorithm develops clusters that minimize the variance of measure scores within the clusters. More specifically, the clustering algorithm minimizes the within-cluster sum of squares for each of the star rating levels. Additional information about the clustering method is provided in Methodology for Producing Star Ratings for In-Center Hemodialysis CAHPS Survey Ratings and Composite Measures.

The cut points (boundaries) for star assignments are derived from the range of individual measures per cluster. The star levels associated with each cluster are determined by ordering the means of each cluster. The cut points for ICH CAHPS star ratings for the two 2017 ICH CAHPS Surveys are shown in **Table 2b3.1.1.e.** Cut points are recalculated for each reporting period.

Lastly, CMS will publish an ICH CAHPS Survey summary star rating, which is the average of all star ratings of the ICH CAHPS measures—the three global ratings and the three composite measures for each participating ICH facility. To calculate the summary star rating, we combine the star ratings for the six ICH CAHPS measures as a simple average. We apply the standard rounding rules described above to the six-measure average to arrive at the ICH CAHPS Survey summary star rating (1, 2, 3, 4, or 5 stars).

> Example. A CCN has the following individual Star Ratings—4, 3, 4, 4, 3, and 3. The simple average of these six ratings is (4 + 3 + 4 + 4 + 3 +3) = 21 ÷ 6 = 3.5. After rounding, their Summary Star Rating is 4 stars.

Table 2b3.1.1.a. "Top-Box" ICH CAHPS patient-mix adjustment factors (average for the 2017 Spring and 2017 Fall ICH CAHPS semiannual surveys) for the October 2018 public reporting period

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Survey Mode						
Mail Only	1.385	2.608	5.431	6.550	3.928	2.894
Phone Only	RC	RC	RC	RC	RC	RC
Mixed Mode	−3.086	−1.377	1.104	0.434	−0.041	1.293
Someone Helped Patient Complete Survey						
Yes	−2.588	−0.843	−0.505	−1.156	−1.686	−4.565
No	RC	RC	RC	RC	RC	RC

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Patient Speaks Language other than English at Home						
Yes	3.328	1.954	−0.596	6.197	3.092	2.012
No	RC	RC	RC	RC	RC	RC
Overall Health						
Excellent	−10.436	−14.015	−13.543	−7.597	−11.281	−2.916
Very Good	−4.296	−6.058	−6.18	−2.939	−4.317	−1.303
Good	RC	RC	RC	RC	RC	RC
Fair	2.492	1.743	2.118	2.366	2.101	1.011
Poor	4.07	3.62	3.908	3.69	2.772	2.572
Mental Health						
Excellent	−11.631	−10.509	−9.368	−10.177	−8.757	−2.7
Very Good	−4.993	−5.009	−4.331	−4.003	−3.353	−1.57
Good	RC	RC	RC	RC	RC	RC
Fair	3.468	3.333	3.146	3.656	2.966	2.165
Poor	7.725	9.497	9.598	8.841	6.437	7.434
Treated for Heart Disease or Problems						
Yes	−2.877	−2.221	−2.083	−2.501	−2.073	−2.194
No	RC	RC	RC	RC	RC	RC
Deaf or Difficulty Hearing						
Yes	1.334	0.67	0.592	1.866	0.839	1.582
No	RC	RC	RC	RC	RC	RC
Blind or Difficulty Seeing						
Yes	1.936	1.322	0.506	0.848	0.538	1.663
No	RC	RC	RC	RC	RC	RC
Difficulty Dressing or Bathing						
Yes	1.999	2.882	2.266	2.057	2.352	2.777
No	RC	RC	RC	RC	RC	RC
Age						
18–44	5.691	5.872	8.222	−0.51	1.702	−6.961
45–54	3.557	6.005	6.995	−0.579	2.329	−4.67
55–64	2.071	3.144	3.505	0.064	1.794	−2.223
65–74	RC	RC	RC	RC	RC	RC
75+	−0.77	−0.72	−2.138	0.791	−0.52	5.208

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Gender						
Male	4.588	2.495	1.923	1.357	−0.751	0.065
Female	RC	RC	RC	RC	RC	RC
Education						
8th Grade or Less	−4.055	−5.782	−6.903	−1.601	−3.479	1.473
Some High School	−2.658	−3.818	−4.584	−1.697	−2.935	0.579
High School	RC	RC	RC	RC	RC	RC
Some College	4.144	5.96	6.74	3.448	4.929	0.185
4-year Degree	5.726	8.18	9.405	4.591	6.305	0.635
More than 4-year college	5.62	9.755	11.476	5.07	7.985	1.237
Years on Dialysis						
1 Year	−0.798	−4.462	−5.082	−0.788	−4.505	1.385
2 Years	−0.276	−2.443	−2.986	−0.566	−2.12	0.905
3–4 Years	RC	RC	RC	RC	RC	RC
5–7 Years	−0.485	1.365	1.031	−0.316	1.241	−0.727
8+ Years	−1.745	1.615	0.717	−0.968	1.661	−1.464

RC = Reference Category

Table 2b3.1.1.b. "Bottom-box" ICH CAHPS patient-mix adjustment factors (average for the 2017 Spring and 2017 Fall ICH CAHPS semiannual surveys) for the October 2018 public reporting period

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Survey Mode						
Mail Only	−5.471	−3.445	−3.164	−0.244	0.496	−2.893
Phone Only	RC	RC	RC	RC	RC	RC
Mixed Mode	−1.093	−0.599	0.670	1.371	1.889	−1.292
Someone Helped Patient Complete Survey						
Yes	1.184	0.816	0.545	2.357	2.784	4.565
No	RC	RC	RC	RC	RC	RC
Patient Speaks Language other than English at Home						
Yes	−0.891	0.404	1.296	−3.78	−2.516	−2.012
No	RC	RC	RC	RC	RC	RC
Overall Health						
Excellent	3.17	4.245	4.017	2.621	4.07	2.916
Very Good	1.254	2.218	2.168	0.911	1.68	1.303
Good	RC	RC	RC	RC	RC	RC
Fair	−1.654	−0.87	−1.127	−1.319	−0.893	−1.011
Poor	−4.203	−3.562	−3.861	−3.85	−2.905	−2.572
Mental Health						
Excellent	3.451	1.998	1.505	3.886	2.536	2.7
Very Good	2.342	1.649	1.22	2.341	1.966	1.57
Good	RC	RC	RC	RC	RC	RC
Fair	−2.59	−2.374	−1.827	−3.117	−2.42	−2.165
Poor	−10.019	−9.435	−8.584	−10.584	−8.737	−7.434
Treated for Heart Disease or Problems						
Yes	1.892	1.388	1.219	2.104	1.698	2.194
No	RC	RC	RC	RC	RC	RC
Deaf or Difficulty Hearing						
Yes	−0.963	−0.592	−0.444	−1.573	−0.729	−1.582
No	RC	RC	RC	RC	RC	RC

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Blind or Difficulty Seeing						
Yes	−1.075	−0.789	−0.158	−2.092	−1.961	−1.663
No	RC	RC	RC	RC	RC	RC
Difficulty Dressing or Bathing						
Yes	−1.653	−2.142	−1.787	−2.4	−2.421	−2.777
No	RC	RC	RC	RC	RC	RC
Age						
18–44	−0.123	−3.024	−3.378	0.166	−2.197	6.961
45–54	−0.742	−3.186	−3.6	−0.358	−3.123	4.67
55–64	−0.639	−2.157	−1.966	−0.619	−2.347	2.223
65–74	RC	RC	RC	RC	RC	RC
75+	0.723	1.092	1.421	0.411	1.339	−5.208
Gender						
Male	−0.986	0.036	0.272	−0.967	0.487	−0.065
Female	RC	RC	RC	RC	RC	RC
Education						
8th Grade or Less	0.99	1.465	1.759	−2.033	−1.2	−1.473
Some High School	0.421	1.112	1.253	−1.129	−0.709	−0.579
High School	RC	RC	RC	RC	RC	RC
Some College	−1.65	−2.292	−2.599	−1.148	−1.192	−0.185
4−year Degree	−1.851	−2.009	−2.508	−1.333	−0.965	−0.635
More than 4-year college	−1.633	−2.539	−3.267	−1.712	−1.65	−1.237
Years on Dialysis						
1 Year	0.499	2.626	2.721	−0.115	2.655	−1.385
2 Years	0.142	1.406	1.395	−0.097	1.116	−0.905
3–4 Years	RC	RC	RC	RC	RC	RC
5–7 Years	0.502	−0.332	−0.501	0.499	−0.575	0.727
8+ Years	0.97	−0.103	−0.113	1.092	−0.538	1.464

RC = Reference Category

Table 2b3.1.1.c. Linear means ICH CAHPS patient-mix adjustment factors (average for the 2017 Spring and 2017 Fall ICH CAHPS semiannual surveys) for the October 2018 public reporting period

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Survey Mode						
Mail Only	2.252	2.058	2.825	2.851	1.362	2.893
Phone Only	RC	RC	RC	RC	RC	RC
Mixed Mode	−0.073	−0.094	0.479	−0.242	−0.648	1.292
Someone Helped Patient Complete Survey						
Yes	−1.086	−0.496	−0.247	−1.397	−1.777	−4.565
No	RC	RC	RC	RC	RC	RC
Patient Speaks Language other than English at Home						
Yes	1.026	0.292	−0.59	3.66	2.056	2.012
No	RC	RC	RC	RC	RC	RC
Overall Health						
Excellent	−3.897	−5.066	−5.036	−3.68	−5.428	−2.916
Very Good	−1.447	−2.053	−2.083	−1.42	−2.12	−1.303
Good	RC	RC	RC	RC	RC	RC
Fair	1.239	0.795	0.938	1.47	1.138	1.011
Poor	2.994	2.459	2.652	3.364	2.546	2.572
Mental Health						
Excellent	−3.779	−3.139	−2.735	−4.817	−3.7	−2.7
Very Good	−1.735	−1.449	−1.192	−2.263	−1.842	−1.57
Good	RC	RC	RC	RC	RC	RC
Fair	1.696	1.441	1.256	2.599	1.999	2.165
Poor	6.331	5.798	5.404	8.399	6.427	7.434
Treated for Heart Disease or Problems						
Yes	−1.425	−1.037	−0.959	−1.907	−1.517	−2.194
No	RC	RC	RC	RC	RC	RC
Deaf or Difficulty Hearing						
Yes	0.53	0.258	0.149	1.312	0.552	1.582
No	RC	RC	RC	RC	RC	RC

Patient-Mix Characteristic Patient-Mix Level	Rating of Kidney Doctors (Q8)	Rating of Dialysis Center Staff (Q35)	Rating of Dialysis Center (Q32)	Average of survey items comprising the Kidney Doctors Communication and Caring Composite	Average of survey items comprising the Quality of Dialysis Center and Operations Composite	Average of survey items comprising the Providing Information to Patients Composite
Blind or Difficulty Seeing						
Yes	0.826	0.41	0.005	1.216	0.961	1.663
No	RC	RC	RC	RC	RC	RC
Difficulty Dressing or Bathing						
Yes	1.19	1.406	1.155	1.761	1.82	2.777
No	RC	RC	RC	RC	RC	RC
Age						
18–44	1.522	2.516	3.335	−0.251	1.613	−6.961
45–54	1.319	2.727	3.103	−0.016	2.289	−4.67
55–64	0.801	1.562	1.659	0.277	1.668	−2.223
65–74	RC	RC	RC	RC	RC	RC
75+	−0.601	−0.731	−1.058	0.035	−0.728	5.208
Gender						
Male	1.542	0.693	0.453	0.897	−0.435	0.065
Female	RC	RC	RC	RC	RC	RC
Education						
8th Grade or Less	−1.179	−1.849	−2.482	0.466	−0.575	1.473
Some High School	−0.819	−1.32	−1.649	−0.123	−0.709	0.579
High School	RC	RC	RC	RC	RC	RC
Some College	1.863	2.439	2.703	1.802	2.252	0.185
4-year Degree	2.454	3.18	3.602	2.289	2.631	0.635
More than 4-year college	2.615	3.891	4.319	2.634	3.42	1.237
Years on Dialysis						
1 Year	−0.442	−2.111	−2.21	−0.229	−2.71	1.385
2 Years	−0.124	−1.201	−1.269	−0.104	−1.231	0.905
3–4 Years	RC	RC	RC	RC	RC	RC
5–7 Years	−0.271	0.385	0.387	−0.358	0.655	−0.727
8+ Years	−0.77	0.379	0.233	−0.904	0.731	−1.464

RC = Reference Category

Table 2b3.1.1.d. National means on patient-mix adjustment factors (average for the 2017 Spring and 2017 Fall ICH CAHPS semiannual surveys for the October 2018 Public reporting period)

Patient-Mix Characteristic Patient-Mix Level	Mean
Survey Mode	
Mail Only	0.037
Phone Only	0.012
Mixed Mode	0.951
Patient Assisted with Survey	
Yes	0.099
No	0.901
Patient Speaks Language Other than English at Home	
Yes	0.158
No	0.842
Overall Health	
Excellent	0.056
Very Good	0.160
Good	0.373
Fair	0.330
Poor	0.081
Mental Health	
Excellent	0.185
Very Good	0.264
Good	0.351
Fair	0.175
Poor	0.026
Treated for Heart Disease or Problems	
Yes	0.447
No	0.553
Deaf or Difficulty Hearing	
Yes	0.157
No	0.843
Blind or Difficulty Seeing	
Yes	0.208
No	0.792
Difficulty Dressing or Bathing	
Yes	0.186
No	0.814

Patient-Mix Characteristic Patient-Mix Level	Mean
Age	
18–44	0.064
45–54	0.128
55–64	0.251
65–74	0.294
75+	0.261
Gender	
Male	0.564
Female	0.436
Education	
8th Grade or Less	0.123
Some High School	0.144
High School	0.333
Some College	0.258
4-year Degree	0.073
More than 4-year college	0.069
Years on Dialysis	
1 Year	0.176
2 Years	0.190
3–4 Years	0.270
5–7 Years	0.202
8+ Years	0.162

Table 2b3.1.1.e. ICH CAHPS star rating cut points (average for the two 2017 ICH CAHPS semiannual surveys)

	1 Star	2 Stars	3 Stars	4 Stars	5 Stars
Rating of Nephrologist (Q8)	<78	≥78 to <82	≥82 to <86	≥86 to <92	≥92
Rating of Dialysis Care Staff (Q32)	<78	≥78 to <82	≥82 to <88	≥88 to <92	≥92
Rating of Dialysis Center (Q35)	<81	≥81 to <85	≥85 to <89	≥89 to <93	≥93
Communication and Caring Composite	<74	≥74 to <78	≥78 to <82	≥82 to <88	≥88
Quality and Operations Composite	<71	≥71 to <75	≥75 to <81	≥81 to <86	≥86
Providing Information Composite	<72	≥72 to <76	≥76 to <80	≥80 to <86	≥86

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.

Not applicable to ICH CAHPS.

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?*

As noted in **Section 2b3.1.1**, the ICH CAHPS Survey mode experiment conducted in 2014 determined the set of patient-mix adjusters used for statistical adjustments. In 2018, the ICH CAHPS team reanalyzed the same data for linearized mean outcomes and determined that the same set of patient-mix variables are still appropriate. The analysis methods used in both sets of analyses included descriptive statistics, correlation analysis, and multivariate regression analysis. The multivariate regression analysis models were used to assess the degree to which the outcome measures represented by the survey items used to create the statistically adjusted scores are affected by the mode and patient-mix characteristics.

Ordinary least squares multivariate regression models were estimated for the mode and patient-mix analysis. Each of the regression models was initially estimated using all of the independent variables identified as potential patient-mix variables and survey mode. Independent variables that were not statistically significant for any of the regression models were then dropped from the regression sequentially until the remaining independent variables were all statistically significant in at least one of the regression models. The regression models for all dependent variables were then finalized to include all of the independent variables that were statistically significant in at least one of the regression models. These were the independent variables identified for use as patient-mix adjusters for public reporting of the ICH CAHPS outcomes measures for individual ICH facilities.

The clinical factors in the ICH patient-mix adjustment model include patient's age, gender, self-reported overall health status, self-reported mental health status, heart disease, deaf or serious difficulty hearing, blind or serious difficulty seeing, difficulty concentrating, remembering or making decisions, difficulty dressing or bathing, and total number of years on dialysis.

The social risk factors in the ICH CAHPS patient-mix model include patient's self-reported education, if patient speaks a language other than English at home, and if someone helped the patient complete the survey.

We do not do any ordering when including risk factors in the patient-mix adjustment model; all patient-mix adjustment factors are entered simultaneously.

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?

The final results of the analysis to select the risk factors (patient-mix characteristics) were submitted to CMS as a Mode Experiment Report in August 2014. As noted in **Section 2b3.3a** the final set of patient-mix variables are all of the independent variables that were statistically

significant in at least one of the regression models. Similarly, for the reanalysis of the data for linearized means the same set of patient-mix variables were significant independent variables in at least one of the regression models.

Every quarter, we update the coefficients for the patient mix adjustments that are applicable to the data that are publicly reported on Dialysis Facility Compare. The most important variables impacting response tendencies for the ICHCAHPS survey are the survey mode and patient characteristics previously identified in the Mode Experiment. The patient-mix adjustment factors being used in the ICHCAHPS Survey are derived from coefficients obtained from Ordinary Least Squares regression analyses on each separate ICHCAHPS response item for the identified patient characteristics. The regression coefficients indicate the tendency of patients with particular characteristics to respond more positively or negatively to ICHCAHPS Survey questions.

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.**

We followed the CAHPS approach for patient-mix adjustment. The selection of factors in the ICH patient-mix adjustment model is based on evidence that these factors met CAHPS patient-mix adjustment criteria: exogeneity, significant variation between dialysis centers, and significant association with CAHPS outcomes within facilities. All patient-mix adjustment factors, including the social risk factors, are measured at the patient level; are exogenous and not caused by the facility; are varied to a substantial and statistically significant extent within facilities; and are associated with patient experience of care outcomes in the patient-mix adjustment model.

The facility's patient-mix adjusted scores for the ratings questions or an individual survey question, as described in the formula in **Section 2b3.1.1**, are adjusted for differences between a facility's patient composition according to the ICH CAHPS patient-mix characteristics and the overall national composition of ICH patients on these same characteristics. This adjustment, which allows consumers to compare different ICH facilities based on the same overall patient composition, is made by subtracting the national mean—the "m's" in the equation above—for a given patient characteristic from an ICH facility's share of patients on this same patient characteristic—the "h's" in the equation above—and then multiplying the difference by the patient-mix adjustment factor—the "a's" in the equation above. The following is an example of adjusting for patient-mix.

- If overall (nationally) 56% of survey respondents are male, but 58% of the respondents from an ICH facility are male, then the adjustment factors for this ICH facility are multiplied by the difference between the ICH facility's patient composition versus the overall national patient composition.
- The score for each of the ICH CAHPS ratings and composite measures for the ICH facility in this example is calculated as 58% minus 56%, or 2%. For the rating of the kidney doctor for this facility, the top-box adjustment factor for males is 4.646 (males were 4.646% less likely to report a "9" or "10" in the rating of their kidney doctors).

- To obtain the top-box rating of the kidney doctor for the ICH facility in this example, we multiply 4.646 times 2% to get 9.29%. In this example, the adjustment for gender for the top-box rating of the kidney doctor for this ICH facility is 9.29%.

As demonstrated in the formula and example above, whether the scores for a given facility are adjusted upward or downward for a given measure depends on the patient-mix adjustments and the patient-mix of that facility relative to the national average patient-mix. Furthermore, facilities that are at the extremes of social risk factors (patient-mix) that are associated with less positive response tendency receive substantial positive adjustments. Conversely, facilities that are at the extremes of social risk factors (patient-mix) that are associated with more positive response tendency received substantial negative adjustments.

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

The analysis methods used in the 2014 mode experiment and in the 2018 reanalysis of the 2014 mode experiment data included descriptive statistics, correlation analysis, and multivariate regression analysis. We first calculated descriptive statistics for all dependent and independent variables to check on the number of missing values for each variable, the sufficiency of sample sizes in each response category for the categorical variables, and the amount of variation for continuous variables. The descriptive statistics identified problems for two of the independent categorical variables with multiple response categories, age and education, that had some response categories with too few respondents included. As a result, several of the response categories were collapsed for those two variables, and the number of response categories reduced, to ensure a sufficient sample size for each of the remaining response categories.

We then conducted correlation analysis on the independent variables. Highly correlated independent variables can cause problems for estimating regression models when both of the correlated variables are included in the models. This analysis included calculating both Pearson correlation coefficients and variance inflation factor statistics. The correlation analysis indicated no major problems with including all of the independent variables in the regression models. Some correlations were found between pairs of independent variables, but they were not large enough to cause concern.

The multivariate regression analysis models were used to assess the degree to which the outcome measures represented by the dependent variables are affected by the mode and patient-mix characteristics represented by the independent variables. If the effects of the independent variables are statistically significant, then adjustments can be calculated for the ratings of ICH facilities for those independent variables that will affect their scores on the six ICH CAHPS outcome measures. The coefficient estimates from the multivariate regression models quantify the change in the dependent variables related to the individual characteristics represented by the independent variables.

Multivariate regression models were estimated for the mode and patient-mix analysis, including one regression model for the dependent variables. When estimating the multivariate regression models for the ICH CAHPS Survey mode experiment analysis, we used the individual patient as the unit of analysis. Ordinary Least Squares (OLS) was the method used in this analysis, following

the approach used in the Hospital CAHPS and Home Health CAHPS mode experiment analyses (CMS, 2008; Elliott et al., 2009; Ingber et al., 2010).

Generally, the linear forms of the multivariate regression models were as follows:

Dependent variable = sum of (coefficients*mode indicators) + sum of (coefficients*patient characteristic indicators)

For the categorical mode and patient-mix characteristics, such as age groups, one group was used as the reference category from the set of categories included in the regression model. That group is the reference to which the effects of the other categories for that variable were scaled, which simplifies the interpretation of the regression coefficients. Implicit in a regression model of this form is an assumption that the estimated patient characteristic coefficients do not vary with facility or mode. The modeling and interpretation can get very complex if there are such interactions. Usually the main effect terms have the largest influence, and few or no interaction terms are needed. As a result, we did not include interaction terms for the regression models used in this analysis.

We conducted the multivariate regression analysis using PROC GLM in SAS Version 9.3 software. We initially estimated each of the regression models using all of the identified potential patient-mix characteristics as independent variables. Independent variables that were not statistically significant for any of the regression models were then dropped from the regression models in four steps, in small increments of two or three variables per step, in sequential estimations of the regression models until the remaining independent variables were all statistically significant in at least one of the regression models. For categorical independent variables, PROC GLM calculates statistical significance for both the independent variable overall and for each of its response categories. The first three steps in this analysis dropped categorical independent variables that had no individual response categories statistically significant in any of the six regression models. The fourth step then dropped categorical independent variables where the variable overall was not statistically significant even if one individual response category was significant. Retaining independent variables significant in at least one of a series of multiple mode and patient-mix regression models for different dependent variables from a single CAHPS survey is an approach similar to the method used in the Hospital CAHPS mode experiment analysis (CMS, 2010; Elliott et al., 2009).

The final regression models for all the dependent variables were then finalized to each include all of the independent variables that were statistically significant in at least one of the regression models. These were the independent variables identified for future use as patient-mix adjusters for public reporting of the ICH CAHPS outcomes measures for individual ICH facilities. See **Section 2b3.1.1** for the final list of patient-mix variables and the statistical formula for creating the patient-mix adjusted scores.

As noted above, the ICH CAHPS team reanalyzed in 2018 the 2014 mode experiment data using linearized means. The analysis methods mirrored those described above and the ICH CAHPS team determined that the same set of patient-mix variables are appropriate for analyzing linearized means.

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

R-square values from regressions ranged from as high as 0.09 to as low as 0.04 for the original analysis using top-box scoring. For the linearized means the regressions ranged from 0.07 to 0.34.

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

Not applicable to ICH CAHPS.

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

Not applicable to ICH CAHPS.

2b3.9. Results of Risk Stratification Analysis:

Not applicable to ICH CAHPS.

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)*

As demonstrated in the formula found in **Section 2b3.1.1**, whether the scores for a given facility are adjusted upward or downward for a given measure depends on the patient-mix adjustments (the results of the regression coefficients) and the patient-mix of that facility relative to the national average patient-mix. Patient-mix adjustment factors are calculated directly from the regression coefficients by multiplying the coefficients by negative one (-1.0). For example, analyses of the data on which results that are being currently publicly reported showed that patients who rated themselves as having excellent overall health on the overall rating of their kidney doctor were 10.4% more likely to provide the most positive ("top-box") response (a rating or a 9 or a 10 for this measure) when compared to the reference group of patients with good health. Thus, the adjustment factor for excellent overall health is -10.4.

2b3.11. Optional Additional Testing for Risk Adjustment *(<u>not required</u>, 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)*

Not applicable to ICH CAHPS.

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)*

To determine if statistically significant and meaningful differences in the risk adjusted top-box scores can be detected we used a two-tailed t-test of means for proportions, with alpha equal to 0.05, to determine if an individual facility's risk adjusted score was statistically different than the overall unadjusted mean score. The first step in conducting this analysis was to compute the overall unadjusted national mean for each of the six measures. This was the mean of all the facilities' unadjusted means, in other words, a "mean of means." Each facility's unadjusted score was weighted equally. Next, we computed each facility's adjusted score on the six measures using the formula described in **Section 2b3.1.1**. Prior to calculating the patient risk factor facility level proportions (h's in the formula) and national level proportions (m's in the formula) we imputed missing values for the patient risk variables. We used hotdeck imputation (for discussion of this, see **Section 2b6.3**, the Item Level discussion) with facility as the imputation class variable. Finally, we calculated the standard errors for each facility's risk adjusted scores

and used a t-test to determine if each facility's risk adjusted mean was statistically different than the overall unadjusted mean, at the 95% confidence level, for each of the six measures. We conducted this analysis using the publicly reported Spring and Fall 2017 data. The results are shown in **Table 2b4.2.a.** If scores were suppressed for publicly reporting (not enough completes, didn't report in both survey periods, facility did not serve enough eligible patients) then they were not included in the analysis.

For the star ratings we present a summary table showing the number of CCNs with one, two, three, four, and five stars for the six measures. These results are found in **Table 2b4.2.b.**

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*)

Table 2b4.2.a. Facilities with patient-mix adjusted combined Spring and Fall 2017 top-box scores statistically different from the unadjusted overall average

Outcome	Number of Facilities Significantly Above the Average	Number of Facilities Significantly Below the Average	% Statistically Different from the Average out of 3302 Facilities
Rating of Nephrologist	270	565	25.3
Rating of Dialysis Care Staff	421	480	27.3
Rating of Dialysis Center	653	399	31.9
Communication and Caring Composite	300	192	14.9
Quality and Operations Composite	147	107	7.7
Providing Information Composite	194	50	7.4

Table 2b4.2.b. Distribution of star ratings, Spring and Fall 2017

	Star Ratings				
	1-Star	2-Stars	3-Stars	4-Stars	5-Stars
Rating of Nephrologist	332	533	882	1,306	249
Rating of Dialysis Care Staff	226	434	1,413	910	319
Rating of Dialysis Center	374	579	966	995	388
Communication and Caring Composite	348	476	750	1,230	498

	Star Ratings				
	1-Star	2-Stars	3-Stars	4-Stars	5-Stars
Quality and Operations Composite	157	379	1,258	1,116	392
Providing Information Composite	269	453	767	1,276	537

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?*)

The number of facilities with patient-mix adjusted scores above the unadjusted overall mean ranged from 194 to 653 out of a total of 3,302 facilities. Conversely, the number of facilities with patient risk adjusted score below the unadjusted overall mean ranged from 50 to 565 out of a total of 3,302 facilities. Together the percentage of facilities which, after the patient-mix adjustment model was applied, had scores significantly different than the overall unadjusted mean ranged from 7.4% for the Providing Information Composite to 31.9% for the Rating of Dialysis Center measure.

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.

Not applicable for ICH CAHPS.

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*)

Not applicable for ICH CAHPS.

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*)

Not applicable for ICH CAHPS.

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)*

Not applicable for ICH CAHPS.

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)*

Bias can be created through missing data if there are correlations between the missing data and the response patterns. ICH CAHPS has conducted two nonresponse bias analyses investigating differential nonresponse in the ICH CAHPS Survey by patient characteristics. The first nonresponse bias analysis was conducted using the 2014 mode experiment data and reported to CMS in the Final Mode Experiment Report. The second analysis was recently conducted using Spring 2017 data. For both analyses the ICH team analyzed and reviewed the potential for response bias at the unit and item levels to determine if further bias testing was necessary. Unit-level analysis uses logistic regression to determine whether differential nonresponse exists among demographic variables in addition to other variables on CrownWEB and CMS administrative records. Response rates at the item level were examined to look for potentially problematic questions with low response rates and changes over time.

The results we present below for the unit nonresponse use the Spring 2017 data and for the item nonresponse we present results from Spring 2015–Spring 2017.

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*)

Unit Nonresponse

By necessity, a nonresponse bias analysis must use data that are known for both respondents and nonrespondents. We first conducted a descriptive analysis of response rates by patient characteristics, shown in **Table 2b6.2.a.** Next, we conducted a logistic regression analysis modeling survey response propensity using the patient characteristics known for both respondents and nonrespondents. Using the results of the logistic model we calculated nonresponse-adjusted weights as the reciprocal of the predicted response propensity from the final logistical regression model.

Table 2b6.2.a. Distributions of respondents, sampled patients, and response rates by patient and facility characteristics for Spring 2017 Survey

Independent variable	Response category	Respondents		All sampled eligible patients		Response rate[1]
		Count	Percent	Count	Percent	
Was diabetes the primary cause of ESRD (CrownWeb)	1 = Yes	40,910	36.58	123,219	38.13	33.20
	2 = No	70,526	63.05	198,649	61.47	35.50
	Missing	415	0.37	1,317	0.41	31.51
Total years on dialysis (CrownWeb)	1 = 1 year	18,863	16.86	49,337	15.27	38.23
	2 = 2 years	21,972	19.64	58,651	18.15	37.46
	3 = 3–4 years	30,191	26.99	87,328	27.02	34.57
	4 = 5–7 years	22,696	20.29	69,358	21.46	32.72
	5 = 8+ years	18,129	16.21	58,511	18.10	30.98
	Missing	0	0	0	0	NA
LIS or Dual anytime (CME)	1 = Yes	43,790	39.15	142,066	43.96	30.82
	2 = No	57,261	51.19	141,785	43.87	40.36
	Missing	10,800	9.66	39,334	12.17	27.46

[1]*The response rate is calculated as the number of respondents divided by the number of sampled patients.*

Response Rates

The first step in the nonresponse bias analysis was to calculate the response rates for variables known for both respondents and nonrespondents. The patient characteristics we investigated included was diabetes the primary cause of ESRD, total years on dialysis, and if the patient was dual eligible or received a low-income subsidy at any time during the survey period. The frequency distribution of respondents, all patients sampled, and the response rates among these categories of patients are found in **Table 2b6.2.a**. This descriptive analysis of nonresponse shows that those less likely to respond were patients with diabetes as their primary cause of ESRD, those in dialysis longer, and those dually eligible for both Medicare or Medicaid or receiving a low-income subsidy.

Logistic Regression Modeling and Creating Nonresponse-Adjusted Weights

Next, we conducted the multiple logistic regression analysis to see which patient characteristics were statistically significant predictors (p-value <= .10) of response propensity. The characteristics included in the multiple logistic regression model are those that are known for responders and nonresponders and are the same found in **Table 2b6.2.a.** Characteristics that were statistically significant predictors should be included in the calculation of the nonresponse-adjusted weights. The results of the logistic regression analysis are found in **Table 2b6.2.b.** The patient characteristics diabetes as the cause of ESRD, total years on dialysis, and if dually eligible for both Medicare and Medicaid, or receiving a low-income subsidy were significant predictors of response propensity. We included these variables in the final logistic regression model and output each respondent's predicted response propensity. We calculated each respondent's nonresponse-adjusted weight as the reciprocal of their predicted response propensity.

Table 2b6.2.b. Results of the logistic regression analysis modeling response propensity for Spring 2017

Independent variable	P-value
Diabetes primary cause of ESRD	<0.0001
Total years on dialysis	<0.0001
Low-income subsidy or Dual anytime	<0.0001

Analysis of Nonresponse-Adjusted Weights

Finally, we conducted a correlation analysis between the nonresponse-adjusted weights and the residuals from the final patient-mix adjustment regression models. The correlations were calculated using the residuals from the top-box regression models and the residuals from the linearized mean regression models. Significant correlation coefficients would indicate that the nonresponse-adjusted weights added to the explanatory power to the patient-mix adjusted survey item top-box and linearized mean scores and improve these scores by adjusting for nonresponse. However, nonsignificant correlation coefficients would indicate that the mode and patient-mix adjusters are sufficient for providing patient-mix adjusted top-box and linearized mean scores and the nonresponse-adjusted weights do not add any explanatory power to the results and are not needed. For the top-box and linearized mean scores the correlations were insignificant (p-value > 0.05) for all of the survey items except for one survey item (Q40) that is used in creating the providing information to patients composite. Given this one result out of 35 we concluded that nonresponse-adjusted weights do not add any explanatory power to the results and are not needed.

As noted in **Section 2b6.1** a similar nonresponse bias analysis was conducted using the 2014 mode experiment data. The correlation coefficients between the residuals and the nonresponse adjusted weights using the mode experiment data were also all nonsignificant indicating nonresponse-adjusted weights do not add any explanatory power to the results and were not needed.

Item Nonresponse

In addition to unit-level nonresponse, we calculated and examined the response rates for each individual question. **Table 2b6.2.c.** summarizes the results of the item nonresponse for each survey period from Spring 2015 to Fall 2017. In particular, we are looking for nonresponse to survey items that exceed 10%, which could potentially indicate a problem and that further analysis is needed.

The table below shows that item nonresponse tends to be low. All items were below our threshold of 10%. Item nonresponse also seems to be stable across time with net differences being under 1%. The only pattern to note is that there is a general trend for item nonresponse to slightly increase as the survey goes on. However, this is not uncommon for surveys of this length.

Table 2b6.2.c. Item nonresponse by survey period

Question Item	Survey Period					
	F2017	S2017	F2016	S2016	F2015	S2015
Q1	1.4%	1.5%	1.9%	2.1%	2.1%	2.1%
Q2	1.0%	1.2%	1.2%	1.2%	1.2%	1.2%
Q3	1.3%	1.3%	1.1%	1.3%	1.3%	1.3%
Q4	0.9%	0.9%	0.9%	1.0%	0.9%	0.9%
Q5	0.9%	1.0%	0.9%	0.9%	0.9%	0.7%
Q6	1.1%	1.2%	1.1%	1.1%	1.1%	0.9%
Q7	1.7%	1.9%	1.2%	1.2%	1.1%	1.3%
Q8	2.1%	2.3%	1.6%	1.8%	1.8%	1.7%
Q9	5.1%	5.2%	4.7%	4.6%	4.6%	4.7%
Q10	1.5%	1.6%	0.7%	0.7%	0.7%	0.9%
Q11	1.4%	1.6%	0.8%	0.8%	0.7%	0.9%
Q12	1.4%	1.6%	0.8%	0.8%	0.7%	0.0%
Q13	1.5%	1.7%	0.9%	0.8%	0.8%	0.9%
Q14	1.4%	1.7%	0.8%	0.7%	0.7%	0.8%
Q15	1.3%	1.5%	0.7%	0.7%	0.6%	0.8%
Q16	2.6%	2.8%	2.5%	2.4%	2.5%	2.7%
Q17	1.8%	2.0%	1.2%	1.2%	1.1%	1.3%
Q18	2.5%	2.6%	1.9%	1.8%	1.9%	2.1%
Q19	2.6%	2.7%	1.9%	1.7%	1.8%	2.0%
Q20	3.1%	3.4%	3.0%	3.1%	3.1%	3.4%
Q21	3.3%	3.3%	2.8%	2.8%	2.8%	2.8%
Q22	1.8%	1.8%	1.0%	1.1%	1.1%	1.2%
Q23	4.8%	4.5%	5.5%	5.2%	5.1%	5.7%
Q24	4.2%	3.9%	4.5%	4.2%	4.2%	4.7%
Q25	3.1%	2.8%	3.3%	3.1%	3.1%	3.4%
Q26	2.9%	2.6%	3.0%	2.7%	2.8%	2.6%
Q27	3.0%	2.7%	3.2%	3.0%	3.0%	3.0%
Q28	4.8%	4.4%	4.7%	4.5%	4.6%	4.5%
Q29	5.0%	4.6%	5.0%	4.7%	4.8%	4.8%
Q30	3.9%	3.6%	4.2%	3.8%	3.9%	4.0%
Q31	3.9%	3.6%	4.3%	4.0%	3.9%	4.0%
Q32	3.8%	3.4%	4.0%	3.7%	3.6%	3.7%
Q33	3.8%	3.3%	4.0%	3.7%	3.6%	3.6%
Q34	3.7%	3.3%	3.9%	3.6%	3.5%	3.5%
Q35	3.9%	3.4%	4.1%	3.9%	3.9%	3.8%
Q36	4.9%	4.4%	5.2%	4.9%	4.7%	4.7%
Q37	6.4%	5.8%	7.3%	7.0%	6.9%	7.2%

Question Item	Survey Period					
	F2017	S2017	F2016	S2016	F2015	S2015
Q38	8.1%	7.6%	8.5%	8.0%	8.0%	8.1%
Q39	6.0%	5.5%	6.4%	6.0%	5.9%	5.7%
Q40	5.7%	5.4%	6.2%	5.8%	5.8%	5.6%
Q41	6.2%	5.8%	7.3%	6.8%	6.6%	6.6%
Q42	6.5%	6.2%	7.8%	7.3%	7.0%	7.0%
Q43	6.1%	5.8%	7.1%	6.6%	6.5%	6.4%
Q44	5.6%	5.2%	6.8%	6.3%	6.3%	6.2%
Q45	4.3%	3.9%	4.2%	4.0%	3.8%	3.3%
Q46	4.1%	3.8%	4.1%	3.8%	3.7%	3.1%
Q47	4.6%	4.2%	4.6%	4.2%	4.1%	3.6%
Q48	4.3%	4.0%	4.3%	3.9%	3.8%	3.3%
Q49	4.8%	4.5%	4.7%	4.4%	4.3%	3.8%
Q50	4.7%	4.3%	4.4%	4.1%	4.1%	3.5%
Q51	5.1%	4.6%	4.7%	4.3%	4.3%	3.7%
Q52	4.9%	4.6%	4.7%	4.3%	4.3%	3.8%
Q53	5.3%	4.8%	4.9%	4.6%	4.5%	4.0%
Q54	4.6%	4.3%	4.5%	4.2%	4.1%	3.6%
Q55	4.9%	4.6%	4.9%	4.5%	4.4%	3.9%
Q56	7.0%	6.1%	6.5%	6.3%	6.3%	5.7%
Q57	8.4%	6.5%	6.7%	6.5%	6.2%	6.4%

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*)

The purpose of this nonresponse analysis was to determine if patient characteristics significantly affecting nonresponse should also be used in creating patient-mix adjusted top-box and linearized mean scores by using nonresponse-adjusted weights. We found only one statistically significant correlation between the nonresponse weights and the residuals from the 35 regression models including mode and the final set of patient-mix factors. Therefore, we conclude that, when using our final patient-mix adjustment model, nonresponse-adjusted weights are not needed to further adjust the patient-mix adjusted top-box and linearized mean scores. Additionally, item nonresponse has been low on the ICH CAHPS survey. Furthermore, for the patient-mix variables we impute missing values using hot deck imputation. The process is to run the regression models to obtain the beta coefficients using raw data that for which missing values have not been imputed. Missing data are then imputed using hotdeck imputation with facility as the imputation class variable, and these imputed data are used for computing the facility-level scores. At this time, we do not feel that adjustments should be made at the item level.

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.

Other

If other: The ICH CAHPS survey collects information directly from ESRD patients receiving in-center hemodialysis via one of three modes of administration: mail only, telephone only and mixed mode with mail and telephone follow-up of non-respondents. Since this measures patient experiences, it would not be available from electronic sources. Proxies are not allowed because questions are only answerable by patients.

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**.

No data elements are in defined fields in electronic sources

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).

The ICH CAHPS survey collects information directly from ESRD patients receiving in-center hemodialysis using one of three modes of survey administration: mail only, telephone only, and mixed mode: mail and telephone follow-up of non-respondents. Since this measures patient experiences, it would not be available from electronic sources. Proxies are not allowed because questions are only answerable by patients. We do not intend to attempt to convert this to an eMeasure. We are not submitting a Feasibility Score Card for this reason.

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.

One difficulty in the "sampling" is that there are a large number of ICH facilities in the study (over 6,000) and the number of patients at these sites tend to be small. The median is approximately 50. Therefore, we have to conduct a census for nearly every facility. The challenge is obtaining enough completed surveys to be able to publicly report the results

We are currently working on the possibility of conducting the survey using the Web. CMS is testing web administration of other CAHPS surveys with a view to creating a protocol and guidelines for web administration of the survey.

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)*.

No fees, licensing, or other requirements to use any aspect of the measure as specified.

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)
	Public Reporting Dialysis Facility Compare https://www.medicare.gov/dialysisfacilitycompare/#search Payment Program ESRD Quality Incentive Program http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/ ESRD Quality Incentive Program http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/ Quality Improvement (external benchmarking to organizations) Dialysis Facility Compare https://www.medicare.gov/dialysisfacilitycompare/#search Quality Improvement (Internal to the specific organization) https://www.medicare.gov/dialysisfacilitycompare/#search Dialysis Facility Compare

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

a. Name and sponsor: ESRD Quality Incentive Program, CMS
Purpose: promote high-quality services in outpatient dialysis facilities treating patients with ESRD. The program links a portion of payment directly to facilities' performance on quality of care measures. These types of programs are known as "pay-for-performance" or "value-based purchasing" (VBP) programs.
Geographic area: United States
Accountable entities: There are approximately 6,000 dialysis facilities included in the survey. The numbers vary over time. For most facilities 100% of the patients were included in the survey.
Level of measurement and setting: Facility level, In-Center Hemodialysis Facilities
d-f Name and sponsor: Dialysis Facility Compare, CMS
Purpose: To provide information to the public and to providers about the qualify of care offered by In-Center Hemodialysis Facilities.
Geographic area: United States
Accountable entities: There are approximately 6,000 dialysis facilities included in the survey. The numbers vary over time.
Level of measurement and setting: Facility level, In-Center Hemodialysis Facilities

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?*)

NA

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.)

NA

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.

ICH-CAHPS Survey data is publicly reported on Dialysis Facility Compare on the Medicare.gov website. All six measures are reported twice a year. The Compare web site includes additional information about the data and its interpretation. Medicare-certified ICH facilities that served 30 or more survey-eligible ICH patients in the preceding calendar year are required to contract with an approved ICH CAHPS Survey vendor and have that vendor administer the ICH CAHPS Survey and submit data from the semiannual surveys to CMS.

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.

Results are available on the web at: https://www.medicare.gov/dialysisfacilitycompare/#search. They are updated twice year. Specific details about the data are provided at: https://www.medicare.gov/dialysisfacilitycompare/#data/about-data

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.

Focus groups with in-center hemodialysis patients were conducted for CMS in February 2016 (Baltimore) and April 2016 (San Antonio).

Feedback was also collected from telephone conference calls and in-person meetings with provider groups throughout the year.

4a2.2.2. Summarize the feedback obtained from those being measured.

The focus group report concluded that: "Current dialysis patients, in particular, emphasized that the survey questions and reported multi-item measures captured exactly the kind of information that was important to know and that they would look for about dialysis centers."

Feedback from informal meetings with patient groups reflect high interest in the survey. Suggestions for improvement include using the web to collect survey data, shortening the questionnaire and conducting the survey annually instead of twice a year.

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.

As part of the Patients Over Paperwork program, we are currently considering options regarding the frequency of administration of the survey. We are analyzing data to determine how shortening the questionnaire will impact the measures. We are also looking into the possibility of electronic

administration of several CAHPS surveys, including ICH CAHPS. No decisions have been made. No modifications have been made at this time.

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.

As noted in the tables for 1b.2 (Appendix Tables 3 and 4),the average score for the multi-item measures increased over time (nephrologists' communication and caring, quality of dialysis center care and operations, and providing information to patients) while the average scores for the three ratings questions (rating of the nephrologist, rating of the dialysis center staff, and rating of the dialysis facility) dropped from 2015 to 2016 and then increased in 2017.

The tables in 1b.2 present the number of CCNs included in the performance results as well as the distribution of the responding patients.

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.

Most CAHPS surveys, such as Hospital CAHPS and Home Health CAHPS, do not survey a chronic population. The only unexpected finding for ICH CAHPS is that patients have complained about having to answer the same survey twice a year. We have had feedback suggesting shortening the questionnaire. We are looking into the implications of doing this for Dialysis Facility Compare and for the QIP program.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

NA

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?

No

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

Because there are no competing measures differences, rationale, impact of interpretability and data collection burden do not exist.

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.)

Not applicable.

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.

Attachment **Attachment:** ICH_CAHPS_Appendix_Doc_V2_PSG.docx

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: Centers for Medicare & Medicaid Services

Co.4 Point of Contact: Debra, Dean-Whittaker, debra.dean-whittaker@cms.hhs.gov, 301-944-4049-

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.

The survey was developed starting in 2004. Many things have changed in the ensuring 15 years. The following list includes both the original developers and, where appropriate, names of people currently in important positions regarding ICH CAHPS and the overall CAHPS program.

AHRQ:

Charles Darby, Christine Crofton, AHRQ representatives and original developers of ICH CAHPS (Darby is deceased, Crofton retired)

Caren Ginsberg, AHRQ, Current director AHRQ CAHPS programs

CMS

Elizabeth Goldstein, CMS, developer of ICH CAHPS

YALE team for CAHPS Consortium

Paul Cleary, Yale, co-principal investigator and developer of ICH CAHPS

Lise Rybowski, Severn Group, contributed to development of ICH CAHPS

Dale Shaller, consultant, contributed to development of ICH CAHPS

RAND team for CAHPS consortium

Ron Hays, RAND, co-principal investigator and developer of ICH CAHPS

Julie Brown, RAND contributed to development of ICH CAHPS

Marc Elliott, RAND chief statistician for ICH CAHPS.

Westat CAHPS team

Joann Sorra, Westat, Current research director CAHPS programs

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2007

Ad.3 Month and Year of most recent revision: 01, 2015

Ad.4 What is your frequency for review/update of this measure?

Ad.5 When is the next scheduled review/update for this measure? 04, 2019

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments:

S.6. Denominator Statement

Table 1. Core ICH CAHPS Survey Questions Applicable to All Sample Patients

Question Number	Question Text
Q1	Where do you get your dialysis treatments?
Q2	How long have you been getting dialysis at [SAMPLE FACILITY NAME]?
Q3	In the last 3 months, how often did your kidney doctors listen carefully to you?
Q4	In the last 3 months, how often did your kidney doctors explain things in a way that was easy for you to understand?
Q5	In the last 3 months, how often did your kidney doctors show respect for what you had to say?
Q6	In the last 3 months, how often did your kidney doctors spend enough time with you?
Q7	In the last 3 months, how often did you feel your kidney doctors really cared about you as a person?
Q8	Using any number from 0 to 10, where 0 is the worst kidney doctors possible and 10 is the best kidney doctors possible, what number would you use to rate the kidney doctors you have now?
Q9	Do your kidney doctors seem informed and up to date about the health care you receive from other doctors?
Q10	In the last 3 months, how often did the dialysis center staff listen carefully to you?
Q11	In the last 3 months, how often did the dialysis center staff explain things in a way that was easy for you to understand?
Q12	In the last 3 months, how often did the dialysis center staff show respect for what you had to say?
Q13	In the last 3 months, how often did the dialysis center staff spend enough time with you?
Q14	In the last 3 months, how often did you feel the dialysis center staff really cared about you as a person?
Q15	In the last 3 months, how often did dialysis center staff make you as comfortable as possible during dialysis?
Q16	In the last 3 months, did dialysis center staff keep information about you and your health as private as possible from other patients?
Q17	In the last 3 months, did you feel comfortable asking the dialysis center staff everything you wanted about dialysis care?

Question Number	Question Text
Q18	In the last 3 months, has anyone on the dialysis center staff asked you about how your kidney disease affects other parts of your life?
Q19	The dialysis center staff can connect you to the dialysis machine through a graft, fistula, or catheter. Do you know how to take care of your graft, fistula, or catheter?
Q20	In the last 3 months, which one did they use most often to connect you to the dialysis machine?
Q22	In the last 3 months, how often did dialysis center staff check you as closely as you wanted while you were on the dialysis machine?
Q23	In the last 3 months, did any problems occur during your dialysis?
Q25	In the last 3 months, how often did dialysis center staff behave in a professional manner?
Q26	In the last 3 months, did dialysis center staff talk to you about what you should eat and drink?
Q27	In the last 3 months, how often did dialysis center staff explain blood test results in a way that was easy to understand?
Q28	As a patient you have certain rights. For example, you have the right to be treated with respect and the right to privacy. Did this dialysis center ever give you any written information about your rights as a patient?
Q29	Did dialysis center staff at this center ever review your rights as a patient with you?
Q30	Have dialysis center staff ever told you what to do if you experience a health problem at home?
Q31	Have any dialysis center staff ever told you how to get off the machine if there is an emergency at the center?
Q32	Using any number from 0 to 10, where 0 is the worst dialysis center staff possible and 10 is the best dialysis center staff possible, what number would you use to rate your dialysis center staff?
Q33	In the last 3 months, when you arrived on time, how often did you get put on the dialysis machine within 15 minutes of your appointment or shift time?
Q34	In the last 3 months, how often was the dialysis center as clean as it could be?
Q35	Using any number from 0 to 10, where 0 is the worst dialysis center possible and 10 is the best dialysis center possible, what number would you use to rate this dialysis center?
Q36	You can treat kidney disease with dialysis at a center, with a kidney transplant, or with dialysis at home. In the last 12 months, did your kidney doctors or dialysis center staff talk to you as much as you wanted about which treatment is right for you?
Q37	Are you eligible for a kidney transplant?
Q39	Peritoneal dialysis is dialysis given through the belly and is usually done at home. In the last 12 months, did either your kidney doctors or dialysis center staff talk to you about peritoneal dialysis?
Q40	In the last 12 months, were you as involved as much as you wanted in choosing the treatment for kidney disease that is right for you?
Q41	In the last 12 months, were you ever unhappy with the care you received at the dialysis center or from your kidney doctors?

S.16. Survey/Patient-reported data

Table 2. Survey Response Rates Spring 2015 – Spring 2018

		No. Sampled	Eligible	No. Completed Surveys	RR
Overall	2015 Spring	363,181	337,316	114,847	34.0%
	2015 Fall	356,721	324,139	103,808	32.0%
	2016 Spring	363,670	339,092	107,582	31.7%
	2016 Fall	347,879	323,386	100,184	31.0%
	2017 Spring	348,024	323,185	111,851	34.6%
	2017 Fall	351,700	321,818	105,120	32.7%
	2018 Spring	332,183	307,078	98,611	32.1%
	All Survey Periods	2,792,851	2,581,604	855,938	33.2%
By Mode	Mail Only	163,043	156,632	45,092	28.8%
	Phone Only	67,809	61,486	13,979	22.7%
	Mixed-Mode	2,561,999	2,363,486	796,867	33.7%
	All Modes	2,792,851	2,581,604	855,938	33.2%

1b2. Performance Scores

Table 3. Summary Statistics for Top-Box Measures

	Spring 2015/Fall 2015					
	TOP BOX Percent of Patients—Nephrologists' Communication and Caring	TOP BOX Percent of Patients—Quality of Dialysis Center Care and Operations	TOP BOX Percent of Patients—Providing Information to Patients	TOP BOX Percent of Patients—Rating of the Nephrologist	TOP BOX Percent of Patients—Rating of the Dialysis Center Staff	TOP BOX Percent of Patients—Rating of the Dialysis Facility
Number of CCN Publicly Reported	3,363	3,363	3,363	3,363	3,363	3,363
Mean	65.6	61.0	78.2	61.8	64.6	61.9
Std dev	9.2	7.8	5.9	11.8	12.8	12.5
max	94	86	93	97	99	95
min	21	37	55	13	14	18
IQR	12	10	7	16	18	17
1st decile	54	51	71	47	48	46
2nd decile	58	54	73	52	54	52
3rd decile	61	57	76	56	58	55
4th decile	63	59	77	59	62	59
5th decile	66	61	79	62	65	62
6th decile	68	63	80	65	69	66
7th decile	71	65	82	68	72	69
8th decile	73	68	83	72	76	73
9th decile	77	71	86	77	81	77

(continued)

Table 3. Summary Statistics for Top-Box Measures (continued)

	Spring 2016/Fall 2016					
	TOP BOX Percent of Patients—Nephrologists' Communication and Caring	TOP BOX Percent of Patients—Quality of Dialysis Center Care and Operations	TOP BOX Percent of Patients—Providing Information to Patients	TOP BOX Percent of Patients—Rating of the Nephrologist	TOP BOX Percent of Patients—Rating of the Dialysis Center Staff	TOP BOX Percent of Patients—Rating of the Dialysis Facility
Number of CCN Publicly Reported	3,159	3,159	3,159	3,159	3,159	3,159
Mean	66.6	61.3	79.5	59.4	66.2	61.0
Std dev	9.2	7.8	5.8	11.7	12.9	12.4
max	92	86	94	92	100	97
min	20	34	56	12	13	17
IQR	12	10	8	15	17	17
1st decile	55	51	72	44	49	44
2nd decile	59	55	75	49	55	50
3rd decile	62	57	77	53	60	54
4th decile	65	59	78	57	64	58
5th decile	67	61	80	60	67	62
6th decile	69	63	81	63	70	65
7th decile	71	65	83	66	74	68
8th decile	74	68	84	69	77	72
9th decile	78	72	87	75	82	77

(continued)

Table 3. Summary Statistics for Top-Box Measures (continued)

	Spring 2017/Fall 2017					
	TOP BOX Percent of Patients—Nephrologists' Communication and Caring	TOP BOX Percent of Patients—Quality of Dialysis Center Care and Operations	TOP BOX Percent of Patients—Providing Information to Patients	TOP BOX Percent of Patients—Rating of the Nephrologist	TOP BOX Percent of Patients—Rating of the Dialysis Center Staff	TOP BOX Percent of Patients—Rating of the Dialysis Facility
Number of CCN Publicly Reported	3,299	3,299	3,299	3,299	3,299	3,299
Mean	67.4	62.3	79.9	60.0	67.1	62.2
Std dev	8.9	7.7	5.8	11.6	12.5	12.2
max	95	89	97	94	100	96
min	26	34	59	18	25	23
IQR	12	10	8	16	17	17
1st decile	56	52	72	45	50	46
2nd decile	60	56	75	50	57	52
3rd decile	63	58	77	54	61	56
4th decile	66	60	79	57	65	60
5th decile	68	62	80	60	68	63
6th decile	70	64	82	63	71	66
7th decile	73	66	83	66	75	69
8th decile	75	69	85	70	78	73
9th decile	79	72	87	75	82	77

1b2. Performance Scores – continued

Table 4. Summary of Respondent Characteristics

Respondent Characteristics		Distribution of Respondents											
		Spring 2015		Fall 2015		Spring 2016		Fall 2016		Spring 2017		Fall 2017	
Total Respondents		**114,847**	**100.0%**	**103,808**	**100.0%**	**107,582**	**100.0%**	**100,184**	**100.0%**	**111,851**	**100.0%**	**105,120**	**100.0%**
Race	**AMINonly**	2,191	1.9%	1,963	1.9%	2,008	1.9%	1,898	1.9%	2,180	1.9%	1,961	1.9%
	Asianonly	2,442	2.1%	2,162	2.1%	2,383	2.2%	2,228	2.2%	2,507	2.2%	2,397	2.3%
	Blackonly	35,088	30.6%	32,005	30.8%	32,581	30.3%	29,799	29.7%	35,219	31.5%	31,764	30.2%
	Filipinoonly	1,811	1.6%	1,638	1.6%	1,786	1.7%	1,753	1.7%	1,919	1.7%	1,752	1.7%
	Mixed	2,659	2.3%	2,418	2.3%	2,539	2.4%	2,504	2.5%	3,905	3.5%	4,471	4.3%
	OtherPacIsl	1,143	1.0%	1,063	1.0%	1,134	1.1%	1,116	1.1%	1,266	1.1%	1,133	1.1%
	Unknown	16,966	14.8%	15,740	15.2%	16,677	15.5%	15,659	15.6%	15,672	14.0%	14,858	14.1%
	Whiteonly	52,547	45.8%	46,819	45.1%	48,474	45.1%	45,227	45.1%	49,183	44.0%	46,784	44.5%
	Missing	14,480	12.6%	12,770	12.3%	13,509	12.6%	13,073	13.0%	13,512	12.1%	15,668	14.9%
Hispanic	**Hispanic**	18,812	16.4%	17,556	16.9%	18,242	17.0%	17,116	17.1%	20,013	17.9%	18,841	17.9%
	NonHispanic	81,555	71.0%	73,482	70.8%	75,831	70.5%	69,995	69.9%	78,326	70.0%	70,611	67.2%
Overall Health Q45	**Missing**	3,795	3.3%	3,928	3.8%	4,310	4.0%	4,231	4.2%	4,405	3.9%	4,490	4.3%
	Excellent	6,070	5.3%	5,751	5.5%	5,776	5.4%	5,477	5.5%	6,050	5.4%	5,710	5.4%
	Very Good	18,299	15.9%	16,787	16.2%	17,017	15.8%	15,481	15.5%	16,956	15.2%	16,244	15.5%
	Good	41,522	36.2%	37,076	35.7%	38,513	35.8%	35,853	35.8%	39,821	35.6%	37,762	35.9%
	Fair	35,978	31.3%	32,214	31.0%	33,569	31.2%	31,420	31.4%	35,804	32.0%	32,800	31.2%
	Poor	9,183	8.0%	8,052	7.8%	8,397	7.8%	7,722	7.7%	8,815	7.9%	8,114	7.7%
Overall Mental Q46	**Missing**	3,615	3.1%	3,799	3.7%	4,128	3.8%	4,127	4.1%	4,222	3.8%	4,301	4.1%
	Excellent	19,949	17.4%	18,121	17.5%	18,955	17.6%	17,436	17.4%	19,737	17.6%	18,710	17.8%
	Very Good	30,066	26.2%	26,999	26.0%	27,598	25.7%	25,689	25.6%	28,292	25.3%	26,670	25.4%
	Good	38,882	33.9%	34,708	33.4%	35,956	33.4%	33,457	33.4%	37,684	33.7%	35,423	33.7%
	Fair	19,377	16.9%	17,579	16.9%	18,273	17.0%	17,063	17.0%	18,988	17.0%	17,490	16.6%
	Poor	2,958	2.6%	2,602	2.5%	2,672	2.5%	2,412	2.4%	2,928	2.6%	2,526	2.4%

Respondent Characteristics		Distribution of Respondents											
		Spring 2015		Fall 2015		Spring 2016		Fall 2016		Spring 2017		Fall 2017	
Treated for diabetes or high blood sugar Q48	**Missing**	3,780	3.3%	3,918	3.8%	4,213	3.9%	4,275	4.3%	4,520	4.0%	4,502	4.3%
	Yes	56,577	49.3%	52,008	50.1%	52,351	48.7%	49,663	49.6%	55,627	49.7%	52,482	49.9%
	No	54,490	47.4%	47,882	46.1%	51,018	47.4%	46,246	46.2%	51,704	46.2%	48,136	45.8%
Treated for heart disease or heart problems Q49	**Missing**	4,399	3.8%	4,430	4.3%	4,692	4.4%	4,695	4.7%	5,080	4.5%	5,072	4.8%
	Yes	48,812	42.5%	44,002	42.4%	45,409	42.2%	42,535	42.5%	47,670	42.6%	44,902	42.7%
	No	61,636	53.7%	55,376	53.3%	57,481	53.4%	52,954	52.9%	59,101	52.8%	55,146	52.5%
Deaf or difficulty hearing Q50	**Missing**	4,074	3.5%	4,216	4.1%	4,427	4.1%	4,435	4.4%	4,809	4.3%	4,945	4.7%
	Yes	17,442	15.2%	15,883	15.3%	16,307	15.2%	15,460	15.4%	16,930	15.1%	15,619	14.9%
	No	93,331	81.3%	83,709	80.6%	86,848	80.7%	80,289	80.1%	90,112	80.6%	84,556	80.4%
Blind or difficulty seeing Q51	**Missing**	4,286	3.7%	4,477	4.3%	4,661	4.3%	4,667	4.7%	5,162	4.6%	5,331	5.1%
	Yes	22,508	19.6%	20,598	19.8%	20,415	19.0%	19,342	19.3%	22,322	20.0%	20,548	19.5%
	No	88,053	76.7%	78,733	75.8%	82,506	76.7%	76,175	76.0%	84,367	75.4%	79,241	75.4%
Difficulty remembering or making decisions Q52	**Missing**	4,343	3.8%	4,511	4.3%	4,627	4.3%	4,702	4.7%	5,118	4.6%	5,129	4.9%
	Yes	17,831	15.5%	15,971	15.4%	16,352	15.2%	15,269	15.2%	17,304	15.5%	15,783	15.0%
	No	92,673	80.7%	83,326	80.3%	86,603	80.5%	80,213	80.1%	89,429	80.0%	84,208	80.1%
Difficulty dressing or bathing Q54	**Missing**	4,174	3.6%	4,235	4.1%	4,487	4.2%	4,555	4.5%	4,783	4.3%	4,837	4.6%
	Yes	21,199	18.5%	18,847	18.2%	19,044	17.7%	17,925	17.9%	20,048	17.9%	18,674	17.8%
	No	89,474	77.9%	80,726	77.8%	84,051	78.1%	77,704	77.6%	87,020	77.8%	81,609	77.6%
5 Level Age Variable	**18–44**	8,256	7.2%	6,871	6.6%	7,267	6.8%	6,014	6.0%	7,272	6.5%	6,698	6.4%
	45–54	15,842	13.8%	13,517	13.0%	14,383	13.4%	12,641	12.6%	14,548	13.0%	13,272	12.6%
	55–64	29,502	25.7%	26,511	25.5%	27,372	25.4%	25,057	25.0%	28,301	25.3%	26,265	25.0%
	65–74	32,062	27.9%	29,953	28.9%	31,051	28.9%	29,575	29.5%	32,826	29.3%	31,055	29.5%
	75+	29,185	25.4%	26,956	26.0%	27,509	25.6%	26,897	26.8%	28,904	25.8%	27,830	26.5%
Gender	**Male**	63,789	55.5%	58,287	56.1%	60,465	56.2%	56,543	56.4%	62,893	56.2%	59,509	56.6%
	Female	51,058	44.5%	45,521	43.9%	47,117	43.8%	43,641	43.6%	48,958	43.8%	45,611	43.4%

Respondent Characteristics		Distribution of Respondents											
		Spring 2015		Fall 2015		Spring 2016		Fall 2016		Spring 2017		Fall 2017	
6 Level Education Variable	**Missing**	6,585	5.7%	6,523	6.3%	6,816	6.3%	6,505	6.5%	6,773	6.1%	7,322	7.0%
	8th Grade or Less	12,234	10.7%	12,550	12.1%	11,627	10.8%	11,135	11.1%	12,608	11.3%	12,117	11.5%
	Some High School	15,723	13.7%	15,919	15.3%	14,315	13.3%	13,525	13.5%	15,327	13.7%	13,959	13.3%
	High School	36,696	32.0%	32,167	31.0%	33,841	31.5%	31,514	31.5%	35,186	31.5%	32,501	30.9%
	Some College	27,820	24.2%	23,211	22.4%	26,286	24.4%	23,889	23.8%	27,173	24.3%	25,146	23.9%
	4-yr Degree	8,167	7.1%	7,097	6.8%	7,466	6.9%	6,936	6.9%	7,579	6.8%	7,197	6.8%
	More than 4-yr college	7,622	6.6%	6,341	6.1%	7,231	6.7%	6,680	6.7%	7,205	6.4%	6,878	6.5%
Language mainly spoken at home Q57	**Missing**	7,321	6.4%	6,440	6.2%	7,015	6.5%	6,753	6.7%	7,283	6.5%	8,796	8.4%
	Other Language	14,209	12.4%	14,074	13.6%	14,312	13.3%	13,903	13.9%	15,861	14.2%	15,139	14.4%
	English	93,317	81.3%	83,294	80.2%	86,255	80.2%	79,528	79.4%	88,707	79.3%	81,185	77.2%
5 Level Years on Dialysis	**Missing**	10	0.0%	—	0.0%	—	0.0%	—	0.0%	—	0.0%	—	0.0%
	1 year	23,074	20.1%	19,040	18.3%	20,192	18.8%	17,176	17.1%	18,863	16.9%	19,252	18.3%
	2 years	20,275	17.7%	19,740	19.0%	20,259	18.8%	19,803	19.8%	21,972	19.6%	19,205	18.3%
	3–4 years	28,812	25.1%	27,342	26.3%	27,510	25.6%	26,948	26.9%	30,191	27.0%	28,460	27.1%
	5–7 years	23,549	20.5%	22,115	21.3%	22,629	21.0%	20,586	20.5%	22,696	20.3%	21,098	20.1%
	8+ years	19,127	16.7%	15,571	15.0%	16,992	15.8%	15,671	15.6%	18,129	16.2%	17,105	16.3%
Use fistula to connect to dialysis machine?	**Missing**	3,940	3.4%	3,262	3.1%	3,386	3.1%	3,039	3.0%	3,790	3.4%	3,298	3.1%
	Yes	72,641	63.3%	66,860	64.4%	69,664	64.8%	65,004	64.9%	71,352	63.8%	67,502	64.2%
	No	38,266	33.3%	33,686	32.5%	34,532	32.1%	32,141	32.1%	36,709	32.8%	34,320	32.6%