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 #: 0006

Corresponding Measures:

De.2. Measure Title: Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey, Version 5.0 (Medicaid and Commercial)

Co.1.1. Measure Steward: Agency for Healthcare Research and Quality

De.3. Brief Description of Measure: The CAHPS Health Plan Survey is a survey that asks health plan enrollees to report about their care and health plan experiences as well as the quality of care received from physicians. HP-CAHPS Version 4.0 was endorsed by NQF in July 2007 (NQF #0006) and Version 5.0 received maintenance endorsement in January 2015. The survey is part of the CAHPS family of patient experience surveys and is available in the public domain at https://www.ahrq.gov/cahps/surveys-guidance/hp/index.html

The survey is designed to be administered to includes individuals (18 and older for the Adult version; parents or guardians of children aged 0-17 for the Child version) who have been enrolled in a health plan for a specified period (6 months or longer for Medicaid version, 12 months or longer for Commercial version) with no more than one 30-day break in enrollment.

The CAHPS Adult Health Plan Survey has 39 items, and the CAHPS Child Health Plan Survey has 41 items. Ten of the adult survey items and 11 of the child survey items are used to form 4 composite measures. Each survey also has 4 single-item rating measures. The aspect of quality assessed by each measure is described below:

Measure 1: Getting Needed Care (2 items)

Measure 2: Getting Care Quickly (2 items)

Measure 3: How Well Doctors Communicate (4 items in Adult survey & 5 items in Child survey)

Measure 4: Health Plan Information and Customer Service (2 items)

Measure 5: How People Rated Their Personal Doctor (1 item)

Measure 6: How People Rated Their Specialist (1 item)

Measure 7: How People Rated Their Health Care (1 item)

Measure 8: How People Rated Their Health Plan (1 item)

1b.1. Developer Rationale: The CAHPS Health Plan survey focuses on patient-centered care, which is a key element of health care quality (IOM, 2001). It measures aspects of health plan and health care quality that are

important to consumers and for which consumers are the best or only source of information (Cleary, Edgman-Levitan, 1997; Cleary, 2016). Use of this measure will benefit both patients and health plans:

1) Patients can use information from the measures to help make more informed choices about which health plan to use.
2) Health Plans and their providers can use data from the surveys for quality improvement initiatives and incentives.
3) Researchers can use data files from the surveys to help answer important health services research questions.

Citations:

Cleary PD, Edgman-Levitan S. (1997). Health care quality. Incorporating consumer perspectives. JAMA. 1997 Nov 19;278(19):1608-12.

Cleary PD. Evolving concepts of patient-centered care and the assessment of patient care experiences; optimism and opposition. J Health Pol, Policy & Law, 2016, 41 (4): 675-696.

Institute of Medicine. "Crossing the Quality Chasm: A New Health System for the 21st Century". March 1, 2001. Accessible at http://iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx.

S.4. Numerator Statement: We recommend that CAHPS Health Plan Survey items and composites be calculated using a top box scoring method. The top box score refers to the percentage of patients whose responses indicated that they "always" received the desired care or service for a given measure. The top box numerator for each of the four Overall Ratings items is the number of respondents who answered 9 or 10 for the item; with a 10 indicating the "Best possible."

S.6. Denominator Statement: The eligible population for the survey includes all individuals who have been enrolled in a health plan for at least 6 (Medicaid) or 12 (Commercial) months with no more than one 30-day break in enrollment. Denominators will vary by item and composite.

S.8. Denominator Exclusions: Individuals are excluded from the survey target population if:

1) They were not continuously enrolled in the health plan (excepting an allowable enrollment lapse of less than 30 days).
2) Their primary health coverage was not through the plan.
3) Another member of his or her household had already been sampled.
4) They had been institutionalized (put in the care of a specialized institution) or are deceased.

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

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Jul 01, 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? Not applicable

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+ and parents/caregivers of patients ages 0-17 who are care recipients through Medicare and Commerical health plans.
- Developers offer a depiction of a logic model connecting patient reported experience of care with structures, clinical quality, patient behavior and outcomes.
- Developers offer evidence of meaningfulness and value:
    - Studies that indicate patients more likely to change physicians based on quality of relationships
    - Studies indicating variance of importance of CAHPS domains across racial and ethnic subgroups
    - Studies indicating importance of provider communication varying by provider type, but consistency for respectful treatment
- The developers provide a literature review of studies that support how changes in the health care system can effect their patient-reported outcome, and how that outcome can impact more distal outcomes.
    - Associations between financial strength of health paln and favorable CAHPS scores
    - Improving infrastructure supporting care suggested to improve CAHPS
    - Improvement in patient safety culture
    - Changes in contracting with providers
- Developer also offers examples of the connection between patient reported experience and positive health outcomes

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:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *If derived from patient report, does the target population value the measured outcome and finds it meaningful?*

Guidance from the Evidence Algorithm

Measure assesses performance on a patient-reported outcome (Box 1) → Empirical data suggests an action providers can take that is associated with greater improvement on the measure (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 reported summary statistics from the HP-CAHPS database for surveys administered September 2018 – June 2017
- Means and standard deviations for the 8 domains of the Adult HP-CAHPS ranged from 0.53 – 0.74 and 0.04 – 0.07 respectively
    - For the Getting needed care composite, the mean top box score/proportion = 0.55 (or 55%); standard deviation (std) = 0.05, minimum (min) = 0.37, median = 0.56, and max = 0.66 (or 66%).
    - For the Getting care quickly composite, the mean top box score/proportion = 0.59 (or 59%); standard deviation (std) = 0.05, minimum (min) = 0.43, median = 0.60, and max = 0.71 (or 71%).
    - For the Communication composite, the mean top box score = 0.74; std = 0.04, min = 0.61, median = 0.75, and max = 0.88 (or 88%).
    - For the Health plan information and customer service composite, the mean top box score = 0.67; std = 0.05, min = 0.53, median = 0.68, and max = 0.80 (or 80%).
    - For the global rating of all health care, the mean top box score = 0.53; std = 0.06, min = 0.33, median = 0.54, and max = 0.75 (or 75%).
    - For the global rating of personal doctor, the mean top box score = 0.65; std = 0.05, min = 0.50, median = 0.66, and max = 0.88 (or 88%).
    - For the global rating of specialist, the mean top box score = 0.66; std = 0.05, min = 0.48, median = 0.66, and max = 0.82 (or 82%).
    - For the global rating of health plan, the mean top box score = 0.57; std = 0.07, min = 0.37, median = 0.58, and max = 0.74 (or 74%).
- Means and standard deviations for the 8 domains of the Child HP-CAHPS ranged from 0.60 – 0.78 and 0.04 – 0.08 respectively
    - For the Getting needed care composite, the mean top box score/proportion = 0.60 (or 60%); standard deviation (std) = 0.06, minimum (min) = 0.42, median = 0.60, and max = 0.72 (or 72%).
    - For the Getting care quickly composite, the mean top box score/proportion = 0.73 (or 73%); standard deviation (std) = 0.06, minimum (min) = 0.55, median = 0.73, and max = 0.86 (or 86%).

- For the Communication composite, the mean top box score = 0.78; std = 0.04, min = 0.66, median = 0.78, and max = 0.90 (or 90%).
- For the Health plan information and customer service composite, the mean top box score = 0.66; std = 0.08, min = 0.0, median = 0.67, and max = 0.84 (or 84%).
- For the global rating of all health care, the mean top box score = 0.67; std = 0.05, min = 0.50, median = 0.68, and max = 0.79 (or 79%).
- For the global rating of personal doctor, the mean top box score = 0.75; std = 0.04, min = 0.60, median = 0.75, and max = 0.84 (or 84%).
- For the global rating of specialist, the mean top box score = 0.71; std = 0.08, min = 0.0, median = 0.72, and max = 0.86 (or 86%).
- For the global rating of health plan, the mean top box score = 0.68; std = 0.07, min = 0.44, median = 0.69, and max = 0.83 (or 83%).

- This data suggests a moderate opportunity for improvement for health plans.

Disparities

- Performance data between White race and non-White race suggest marginal differences between some domains, with Whites reporting better experiences than non-Whites by 1-2% in 6/8 domains and reporting the same in the remaining two.
- Comparable findings for gender differences; generally similar or same results.

Questions for the Committee:

- Are performance gaps justifying this measure evident, or does it appear performance is pretty high and narrow between providers?

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.*

- model clear to PRO; updated info provided; pass
- Good start in identifying patient feelings about provider/plans. May need to keep track of evolving priorities for what matters to patients . Collins (2017) did note that domain importance did vary across subgroups. Evidence is a 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?*

- performance data provided; disparities disucssed; more information on performance gaps with opportunities to improve would be helpful; moderate
- Mean scores for adults and children suggest a MODERATE opportunity for improvement. Disparities are comparable for gender and marginal for race

Criteria 2: Scientific Acceptability of Measure Properties

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

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

2c. For composite measures: empirical analysis support composite approach

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.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

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 and discussed on the call. A summary of the measure and the Panel discussion is provided below.

- **Ratings for reliability:** 1 high, 4 moderate, 0 low and 1 insufficient → Measure passes with a moderate reliability
    - Testing included score-level and data element testing
    - SMP members noted that data element and score-level testing was conducted via "Cronbach's alpha, ICC and plan-level reliability (signal-to-noise). Cronbach's alphas tended to be below 0.70 threshold, largely because of 2-item scales. The coefficients are high enough to suggest they will perform reasonably well. Regarding between vs within plan variance, I generally considered all ICCs to be problematic (all below 0.05), as they indicate that clinicians and sites may not be able to be differentiated. Plan-level reliabilities (mis-stated as site-related) were generally good and suggested by the submitters as surrogates for review of plans." Another panelist added, "The methods used for reliability testing were generally acceptable, using standard and well-accepted methods, at both data element and measure score levels," and "The composite measures have generally good internal consistency. Signal to noise (i.e., plan-level reliability values) were mixed and one of the measures had less than optimal estimates ('Global rating of specialist' had a STN of 0.60 and 0.59 for adult and child populations,

respectively). These lower values are acceptable, however, as a whole, the instrument should receive a moderate rating."

- Results were summarized as "Cronbach's alpha (data element level testing) coefficients were modest (alpha ranged from 0.61 - 0.89) but acceptable for the adult health plan population and weaker for the child health plan population (alpha ranged from 0.57 - 0.85). The plan level STN analysis revealed acceptable plan level reliability (STN ratios ranged from 0.67 to 0.88 for adults and 0.59 to 0.92 for children's data)."
- For risk adjustment, "The developer needs to clearly specify if these are risk adjusted measures. Both reliability and validity testing should be consistent with how the measures are specified," and "The developer checked the "statistical model" in the risk adjustment section, however, both reliability and validity tests seemed to be based on unadjusted results."
- Note: Cronbach's alphas were reported for data-element testing domains with more than one item. Measure developer did not perform data-element level reliability testing for the single-item measures.

  Measure 5: How People Rated Their Personal Doctor (1 item)
  Measure 6: How People Rated Their Specialist (1 item)
  Measure 7: How People Rated Their Health Care (1 item)
  Measure 8: How People Rated Their Health Plan (1 item)

- **Ratings for validity:** 1 high, 4 moderate, 1 low and 0 insufficient → Measure passes with a moderate reliability
  - Testing included score-level and data element testing
  - SMP members characterized the testing as follows: "Intercorrelations of composites and global scores, and item-to-total were a reasonable start. In the submission, correlation guidelines proposed are: 0.10 is a small correlation; 0.30 is a medium correlation; and 0.50 is a large correlation. I don't [think] this aligns with our emerging standard which would push these numbers up. "
  - Summarizing the results, SMP noted: "Overall item and composite correlations with global satisfaction measures were positive and moderately strong as one would expect. More encouraging was the fact that the top box methodology used in these measures showed a consistent and marked pattern of increasing the relationship. The results supported the construct validity for the measures."
  - One critique of the approach from the SMP was that "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 clinic 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."

Questions for the Committee regarding reliability:

- 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**

Measure Number: 0006

Measure Title: Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey, Version 5.0 (Medicaid/ and Commercial)

Measure Title: CAHPS Health Plan Survey, Version 5.0

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** ☐ **Enrollment 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#1: No concerns

   MP#4: None.

   MP#3: The developer needs to clearly specify if these are risk adjusted measures. Both reliability and validity testing should be consistent with how the measures are specified.

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** **N/A** NA

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

   MP#2:

   a. **Methods appear to be appropriate.** Cronbach's alpha, ICC and plan-level reliability (signal-to-noise). Cronbach's alphas tended to be below 0.70 threshold, largely because of 2-item scales. The coefficients are high enough to suggest they will perform reasonably well. Regarding between vs within plan variance, I generally considered all ICCs to be problematic (all below 0.05), as they indicate that clinicians and sites may not be able to be differentiated. Plan-level reliabilities (mis-stated as site-related) were generally good, and suggested by the submitters as surrogates for review of plans.

   Submission document: Testing attachment, section 2a2.2

   MP#6: ICC was appropriate analysis.

   MP#4: Testing methods are appropriate

   MP#3: The method used for assessing measure score reliability was appropriate, reporting both ICC and plan-level reliability.

   Cronbach's alphas were calculated to assess the internal consistency of multi-item scales. However, the measures as specified were not based on the scale scores. First, instead of using the original response categories for each item, responses were dichotomized using top-box scaling procedure. Second, a top-box response rate was calculated for each item each health plan, then rates for multiple items within a scale were averaged to arrive at an overall composite score. It seems that reported Cronbach's alphas were calculated at the respondent level while the measures as specified were composited at the health plan level.

   MP#1: The developer submitted an analysis of internal consistency of the items in the plan HCAHPS instrument within each composite using Cronbach's alpha, an ICC to evaluate plan-level effects on scores and plan-level reliability scores via a signal-to-noise approach. These approaches are appropriate. The psychometric properties of the instrument have been published previously.

   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#2:

   a. **See above**

   Submission document: Testing attachment, section 2a2.3

   MP#6: ICC demonstrated adequate reliability. Submitors noted and addressed the lower ratings for the specialist. Explanation was acceptable.

   MP#4: Testing sample is adequate for generalization.

   MP#3: The results of measure score reliability were in general good, particularly when the number of respondents per plan is above 300. This should be taken into account when implementing the measures.

   At the respondent level, internal consistency estimates were very good for "how well doctors communicate" composite, above 0.8 for both adult and child surveys. Cronbach's alphas were lower for the remaining three composites. More importantly, given how the measures are specified, Cronbach's alphas as calculated do not provide the information needed.

   MP#1: Cronbach's alpha (data element level testing) coefficients were modest (alpha ranged from 0.61 - 0.89) but acceptable for the adult health plan population and weaker for the child health plan population

(alpha ranged from 0.57 - 0.85). The plan level STN analysis revealed acceptable plan level reliability (STN ratios ranged from 0.67 to 0.88 for adults and 0.59 to 0.92 for children's data).

MP#5: The results of reliability testing were acceptable, although the results presented for measure score reliability were somewhat confusing. One test, referred to as ICC, presented results on a scale that is not the usual one for ICC. Results of this method of testing generally fell below a declared threshold of .05 needed to identify meaningful differences among practices. On the basis of that one test, the results at the measure score level would seem to be unreliable. But, another test of "reliability" was presented indicating that results at the clinic level were sufficiently reliable. The first test apparently indicates the percent of overall variance accounted for by true differences in performance rather than measurement error, with findings for most measures in the set falling below a 5% threshold. If this interpretation is correct, then less than 5% of the variance in observed scores is accounted for by some meaningful underlying performance dimension. With sufficient sample sizes, though, even this very weak "signal" can be reliable.

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** See comments under 7.☐☒ **No** (data element reliability not evaluated for all measures, just the composites, however, score level reliability was appropriately tested and acceptable)☐

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

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

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

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

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

    ☒☐ **Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> 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#6: Submitors discussed lower results of the specialist ratings. They do note that the health plan level reliability is adequate and that is what it is designed to measure.

    MP#1: The composite measures have generally good internal consistency. Signal to noise (i.e., plan-level reliability values) were mixed and one of the measures had less than optimal estimates ('Global rating of specialist' had a STN of 0.60 and 0.59 for adult and child populations, respectively). These lower values are acceptable, however, as a whole, the instrument should receive a moderate rating.

    MP#3: NQF requires reliability testing for both data element and measure score for instrument-based measures. Reported internal consistency testing is not consistent with how the measures are specified.

MP#4: Evaluated at the plan-level with good plan-level reliability when completed by 300 or more respondents per plan.

MP#2:

a. **Bordering on low in my opinion, based on ICCs and likelihood that plan differences cannot be measured, even if elements are reliable, and the elements are on the borderline.**

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#6: No exclusions

MP#2:

a. No concerns

Submission document: Testing attachment, section 2b2.

MP#4: No exclusions per steward

MP#3: No concern.

MP#1: N/A

MP#5: None

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

MP#2:

a. **See above. Differentiating plans hampered by reliability data (ICCs)**

Submission document: Testing attachment, section 2b4.

MP#4: Must have a sufficient sample size e.g., 300 completes to identify meaningful differences in performance.

MP#3: No concern.

MP#1: None

MP#5: The developers find that a substantial number of practices 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.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

MP#2:

a. **No concern**

Submission document: Testing attachment, section 2b5.

MP#4: NA

MP#3: No concern.

MP#5: N/A

15. **Please describe any concerns you have regarding missing data.**

MP#2:

a. **Not sure if accepting data if only 51% of respondents answer is sufficiently representative.**

Submission document: Testing attachment, section 2b6.

MP#4: No concerns.

MP#3: No concern

MP#5: None.

16. **Risk Adjustment**

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

MP#3: The developer checked the "statistical model" in the risk adjustment section, however, both reliability and validity tests seemed to be based on unadjusted results.

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

☒☐ Yes ☐ No ☐☒ Not applicable

MP#3: It is left up to the users to decide how to implement it. It says, "the variables chosen for adjustment and the steps for calculations of risk-adjusted scores are user-defined."

16c. **Social risk adjustment:**

MP#2: Assuming education and language are "social" risk factors

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:** Not applicable**.**

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

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#6: Submitors note that CAHPS does not require risk adjustment but provided risk adjustment analysis that could be applied to the survey results.

MP#4: The CAHPS results are not required to be risk adjusted by users other than those that report measure results to the public e.g., CMS.

MP#3: Risk adjustment is not required for the measures.

MP#1: Approach is modest but acceptable.

MP#2: CAHPS results are not required to be risk adjusted by users. However, users of the survey, including public reporting entities such as CMS or NCQA, may voluntarily decide that they want to adjust the data to account for patient case-mix differences if comparing plans. The recommended and provided software applies linear regression models and the current CAHPS Analysis Program (described in section 2b3.1.1) suggests adjusting for general health status, mental health status, age, and education. These data are available, along with language (Spanish/English)

MP#5: The approach is generally acceptable, although the discussion does not go into any significant detail about attempts to include race/ethnicity or SES in the adjustment models**.** The data set used for analysis came from only Medicaid health plans, so the ability to test and possibly adjust for a range of social and economic factors would be limited by the range of such factors found in this data set. Education is included in the adjustment model. Race/ethnicity and SES are not. It should also be noted that risk adjustment can be done using software available on the CAHPS web site, but adjustment is not automatic or mandatory in the CAHPS process itself. Users (e.g., payors) may or may not require risk adjustment. Although the

developers show that the risk adjustment they have available has relatively little effect on scores, an endorsement process at NQF should probably recommend endorsement on the condition that risk adjustment is used.

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#6: No concerns with analysis

    MP#2:

    a. **Intercorrelations of composites and global scores, and item-to-total were a reasonable start. In the submission, correlation guidelines proposed are: 0.10 is a small correlation; 0.30 is a medium correlation; and 0.50 is a large correlation. I don't this aligns with our emerging standard which would push these numbers up.**

    Submission document: Testing attachment, section 2b2.2

    MP#4: Used Spearman Rho Correlation coefficient to determine associations within the instrument.

    MP#3: Both item-level top-box scores and composite scores were correlated with the global ratings of provider at the individual and health plan level. Validity of the global ratings were assumed.

    MP#1: The developer correlated each measure's (individual and composite) top box score with the global rating of doctor, specialist, healthcare and health plan. The developers focused on construct validity using confirmatory factor analysis and correlational analyses. While this approach is unorthodox for NQF validity testing, it is reasonable to perform this approach for patient experience measures.

    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**

    MP#6: Adequate for comparing at the health plan level

    MP#2:

    a. **Generally, the data provided supports plan-level comparisons on the global and composite measures as proposed. There are some exceptions: Global scores are all more-or-less related to, but also distinct from one another, for both adults and children. Similarly, correlations of global measures were all moderate (suggesting uniqueness), with exception of 'getting care quickly' and 'getting needed care' which appear to be redundant measures (correlating 0.80 and 0.81, for adults and children, respectively.**

Submission document: Testing attachment, section 2b2.3

MP#4: Used Medicaid data exclusively with an adequate sample size that may be generalized to the widespread population.

MP#1: Overall item and composite correlations with global satisfaction measures were positive and moderately strong as one would expect. More encouraging was the fact that the top box methodology used in these measures showed a consistent and marked pattern of increasing the relationship. The results supported the construct validity for the measures.

MP#3: For adult survey, moderate correlations were reported between both individual items and composite scores and the global measures at the health plan level. The results were not as good for child survey, in particular, the results for "how well doctors communicate" and "getting care quickly" were poor.

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 <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> 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 <u>is required</u>; 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#6: No concerns, submitters provided adequate validity testing between health plans.

    MP#3: Correlations with the global ratings ranged from low to moderate.

    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 clinic 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.

MP#1: Construct validity is weaker than other forms of validity, but the findings presented support continued use of these measures.

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: Composite measure demonstrated expected association of components, and consistency of direction

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#1: Standing committee should provide input if there are measures that would serve as reasonable gold-standards for these measures.

    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.

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?*

- detailed analysis included; no concerns identified
- Risk adjustment is not required. Measure reliability is MODERATE. Between vs in plan variance suggests some difficulty in differentiating paln

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

- SMP - moderate; no further questions; results clear
- Reliability is MODERATE. No concerns

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

- SMP - moderate; no further questions
- Testing results are valid. Rated MODERATE

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?*

- no questions
- No threats seen

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 clear; risk model well-explained; SDOH described; results acceptable
- Risk adjustment not required

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.

- Developers describe several modalities of survey administration available and present some research associated with response rate for mixed modalities.
- Developers note that the instrument is free to use.
- Developers note that health plans are required to use a CMS approved vendor to administer the surveys but does not provide a cost assessment associated with retaining such a vendor.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?
- If an eCQM, does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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

RATIONALE:

- 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?*

- same concerns as with HH-CAHPS - no description of entity costs or admin burdens to administer survey; low
- Operationally feasible. The fee for a vendor as a burden was identified though I don't think that an expense for a PRO-PM is a burden but rather a necessity. Feasibility is MODERATE

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

- NCQA Health Plan Rankings—Public reporting
- CMS Medicare Advantage Plan Finder—Public reporting
- Qualified Health Plan Quality Ratings—Public reporting
- OPM's Federal Employee Health Benefits Performance Assessment—Payment program
- NCQA Health Plan Accreditation—Regulatory and accreditation
- CAHPS Database—Quality improvement (external benchmarking)
- Health Share of Oregon—Quality improvement (internal)

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

- Measure developer provides review of feedback for both development and implementation
- Development

- Literature review and review of existing measures
- Development and consultation with technical expert panels
- Focus groups with consumers
- Cognitive testing
- Field testing
- Public comment
- On-going collaboration and harmonization with key partners and stakeholders
- Input from the NCQA Task Force and review and approval by the NCQA Committee on Performance Measurement

- Implementation
  - User feedback from the AHRQ CAHPS Database
  - CAHPS Database team does yearly outreach to survey vendors to gather feedback for improvement
  - Develop states they are unaware of any problems experienced by health plans and does not offer any examples of feedback from health plans; this is also true of respondents
  - Developer

Additional Feedback:

***Questions for the Committee**:*

- Is there anything that the Committee wishes to discuss related to the current use of the measure?

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

- Results show modest but fairly consistent year over year improvement.
- From 2007 to 2017, the top box score for all four composites and the provider ratings increased for the Adult Medicaid population.
- From 2007-2017, for the Child Medicaid population, the top box composite scores improved for only two of the composites: Getting needed care and Health plan information and customer service, yet there was improvement in the provider ratings.
- For a chart depiction of improvement see pages 8-11 of Chartbook, found at https://www.ahrq.gov/sites/default/files/publications2/files/cahps-database-2017-hp-chartbook_0.pdf

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

- Developer report that transparency of CAHPS scores to support consumer choice have influenced improvement in quality of care for HCAHPS, though has no evidence for HP-CAHPS to the same effect.

Potential harms

- Developer does not offer any examples of potential harms

Additional Feedback:

Questions for the Committee:

- Are the trending data sufficient to demonstrate the quality metric is yielding improvement?

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?*

- evidence of improvement over time included; better understanding of factors that contributed would be of value; moderate
- 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.*

- usability described; benefits > any possible harm; moderate
- Improvements are modest over time but improving. MODERATE

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?*

- s/a described in HH-CAHPS
- None seen

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

HP_CAHPS_evidence_040919.docx,HP_CAHPS_evidence_041819_508.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.

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0006

Measure Title: CAHPS Health Plan Survey, Version 5.0

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

Date of Submission: 4/9/2019

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

Outcome

☐ Outcome:

☒Patient-reported outcome (PRO): Patient experience with care

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:

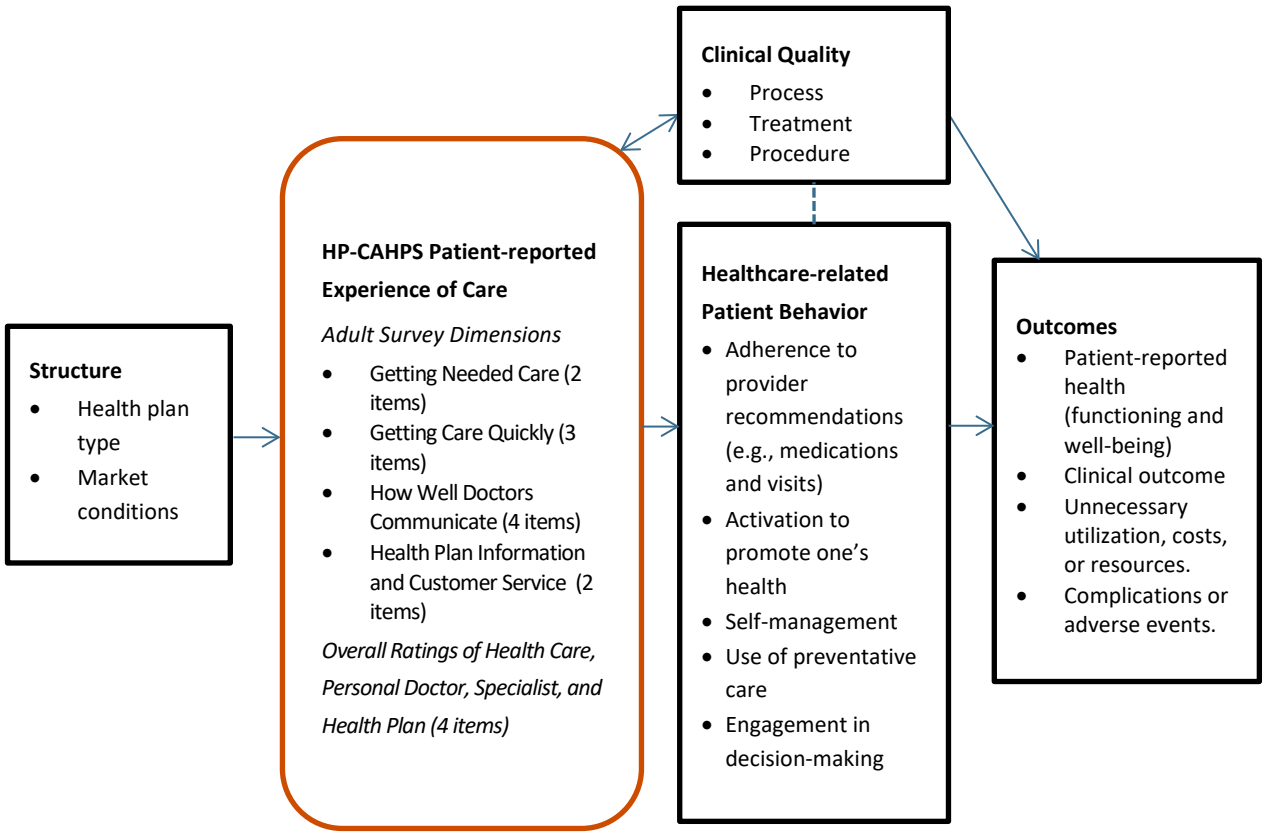
☐ Structure:

☐ 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 model below shows the CAHPS measures, the healthcare factors that influence CAHPS dimensions, and the relationship between HP-CAHPS and patient behaviors and clinical outcomes. Patient care experiences are

influenced by structural features of health care; in turn, patient perceptions of high quality care improve patient adherence and activation, leading to improved clinical 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 CAHPS surveys measure aspects of patient-centered care that complement clinical process and outcome measures in consumer choice, quality improvement, public reporting, and pay-for-performance programs (Anhang Price et al, 2014). Published research indicates that individuals use information from patient experience measures to *make decisions* about their healthcare providers and plans. One study found that seeing publicly reported quality information was a determinant of choosing higher quality-rated health plans, although the weight given to quality information also depended on other features, such as cost and provider choice (Faber et al., 2009). A study of low-income parents in New York State found that parents choose SCHIP managed care plans with higher CAHPS scores for their newly enrolled children (Liu et al., 2009). Additionally, a study of physician choice found that patients choosing a new primary care physician valued other patients' reports of interpersonal quality and overall recommendations (Fanjiang et al., 2007).

Patient experiences with health plans are also linked to their *persistence in the plans*. For example, one study found that the mean voluntary disenrollment rate among Medicare managed care enrollees is four times higher for plans in the lowest 10 percent of overall CAHPS Health Plan survey ratings than for those in the highest 10 percent (Lied et al, 2003). At the provider level, patients who reported the poorest-quality relationships with their physicians are three times more likely to voluntarily leave the physicians' practice than patients with the highest-quality relationships (Safran et al., 2001).

Racial and ethnic patient subgroups may value various aspects of the care experience differently. CAHPS surveys have been used to measure these differences. For example, Collins et al (2017) found that the CAHPS domains with the most importance to respondents varied across subgroups. These researchers conclude that

tailoring quality improvement programs to the factors most important to the racial, ethnic, and language mix of the patient population of the health plan may more efficiently reduce disparities and improve quality.

Citations:

Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, Edgman-Levitan S, Cleary PD. (2014) Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev 71 (5):522-54.

Collins RL, Haas A, Haviland AM, Elliott MN. (2017). What Matters Most to Whom: Racial, Ethnic, and Language Differences in the Health Care Experiences Most Important to Patients. *Med Care.* Nov;55(11):940-947.

Faber, M, Bosch, M., Wollersheim, H, Leatherman, S, and Grol, R. (2009). Public reporting in health care: how do consumers use quality-of-care information? A systematic review. *Med Care* 47(1):1-8.

Fanjiang, G, von Glahn, T, Chang, H, Rogers, W, and Safran, D. (2007). Providing patients web-based data to inform physician choice: if you build it, will they come? *J Gen Intern Med* 22(10): 1463-1466.

Lied TR, Sheingold SH, Landon BE, Shaul JA, Cleary PD. Beneficiary reported experience and reported voluntary disenrollment in Medicare managed care. *Health Care Finance Rev.* 2003;25(1):55–66.

Liu, H, Phelps, C, Veazie, P, Dick, A, Klein, J, Shone, L, Noyes, K, and Szilagyi, P. (2009). Managed care quality of care and plan choice in New York SCHIP. Health Serv Res 44(3): 843-861.

Safran DG, Montgomery JE, Chang H, Murphy J, Rogers WH. (2001). Switching doctors: predictors of voluntary disenrollment from a primary physician's practice. *J Fam Practice.* 50(2):130–6.

****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.

INTRODUCTION

CAHPS surveys measure key aspects of patient experiences, such as provider communication and ease of access, that are consistent with patient-centered care. CAHPS focuses on aspects of care that consumers have identified as important and for which patients are the best or only source of information. CAHPS scores are actionable measures that can help clinicians and health plans target develop and assess interventions that could improve the quality of patient-centered care. Below we review the literature on the determinants of patient care experiences measured by CAHPS and their associations with other indicators of health care quality.

REVIEW OF THE EVIDENCE

Prior research has identified several features of healthcare delivery structure, including plan characteristics and market-level characteristics that are associated with patient experiences. Two major systematic reviews have examined the relationships among patient experience, clinical processes, and patient outcomes. A systematic review performed by researchers in the U.K. found that patient experience is favorably associated with adherence to recommended medications and treatments, preventive care such as screenings and immunizations, patient-reported health outcomes, clinical outcomes, reduced hospitalizations and primary care visits, and reduced adverse events (Doyle et al., 2013). Anhang Price et al. (2014) reviewed evidence on the association between patient experiences and other measures of health care quality in the U.S. They similarly found that better patient care experiences are associated with higher levels of adherence to recommended prevention and treatment processes, better clinical outcomes, and less health care utilization.

RATIONALE

Structure

Health plan type and market characteristics have been found to predict patient experience in several studies. Among managed care organizations (MCOs), for example, Medicaid enrollees had significantly less favorable CAHPS scores than commercial plan enrollees (Elliot, Farley, et al., 2005). The financial strength of health plans, as measured by their fiscal margins, was associated with more favorable CAHPS scores (Beauvais et al., 2007). Market-level factors such as HMO competition and penetration did not appear to affect patient experience (Scanlon, Swaminathan, et al., 2008), but Medicare beneficiaries in "higher-intensity" healthcare markets reported more problems getting care quickly than in markets with less healthcare consumption (Mittler, Landon, et al., 2010). For health plans, prior work highlights multi-level impacts on patient experience, which can occur at the system, care site, or physician level – with physicians accounting for the largest proportion of explainable variance (Rodriguez, Scoggins, at al., 2009). Improving the infrastructure supporting certain aspects of care may have broad effects because system changes can influence multiple outcomes (Cleary, 2016).

Health Plan Influence on the Patient Experience

For the health plan beneficiary, the patient experience is an outcome that is influenced by health plan and provider processes of care and service. To improve the patient experience, health plans use HP-CAHPS results to determine where improvement is needed and to choose strategies for improvement. For example, Harvard Pilgrim Health Care analyzes members' administrative/service complaints and feedback and the CAHPS results of members who have filed complaints, as well as service metrics like call abandonment rates and average speed-to-answer (Harvard Pilgrim Health Care Quality Program Description, 2017). Health plans can also assist their contract providers in the education and application of quality improvement techniques such as Lean. Lean enables organization to identify and make positive changes to processes and workflows (AHRQ CAHPS Ambulatory Care Improvement Guide, 2018). Lastly, health plans can offer support groups and self-care programs that can increase patients' knowledge about their disease, which in turn can improve patient-provider communication and patient adherence to a prescribed treatment. Studies of support groups formed for chronic arthritis, heart disease, stroke, and lung disease have shown that such groups have beneficial effects on mental and physical health as well as social functioning (Lorig et al, 1999).

Health-related Patient Behavior

One composite of the HP-CAHPS survey assesses patients' perceptions of how well providers communicate with them. Better patient-provider communication promotes healthcare-related patient behaviors (Fuertes, Boylan et al. 2009). A 2009 meta-analysis of 127 studies assessing the link between patient treatment adherence and physician-patient communication found a 19% higher risk of non-adherence among patients whose physician communicated poorly (Zolnierek and Dimatteo 2009). Doyle's (2013) meta-analysis showed positive associations between the quality of clinician-patient communications and adherence to medical treatment in 125 of 127 studies analyzed. Studies using the CAHPS measure have found that better provider communication is positively associated with adherence to hypoglycemic medications among diabetics (Ratanawongsa, Karter et al. 2013), adherence to tamoxifen among breast cancer patients (Liu, Malin et al. 2013), and higher rates of colorectal cancer screening among adults in the US (Carcaise-Edinboro and Bradley 2008).

Quality Improvement / Interventions

Researchers have used the CAHPS survey to learn if accountable care organization (ACO) incentives to limit health care use and improve quality may enhance or hurt patients' experiences with care. More specifically, using CAHPS survey data covering 3 years before and 1 year after the start of Medicare ACO contracts in 2012 as well as linked Medicare claims, McWilliams et al. (2014) compared patients' experiences in a group of 32,334 fee-for-service beneficiaries attributed to ACOs (ACO group) with those in a group of 251,593 beneficiaries attributed to other providers (control group), before and after the start of ACO contracts. They found that, in the first year, ACO contracts were associated with meaningful improvements in some measures of patients' experience and with unchanged performance in others. Lastly, Sorra et al found that hospitals with more positive perceptions of patient safety culture tend to have more positive CAHPS scores from their

patients. This finding suggests that improvements in patient safety culture may lead to improved patient experience with care.

Clinical Quality

Measuring patient experiences can support the goal of providing high quality, patient-centered care. Schneider and colleagues (2001) found that two HP-CAHPS composites were associated with several Healthcare Effectiveness Data and Information Set (HEDIS) clinical process measures among Medicare health plan enrollees. They found that experiences obtaining needed care and getting information and customer service from health plans were associated with mammography, eye examinations for diabetics, receipt of beta-blockers following myocardial infarction, LDL cholesterol testing following an acute cardiovascular event, and follow-up within 30 days following a hospitalization for mental illness (Schneider et al, 2001).

Outcomes

Doyle's (2013) meta-analysis of 40 evidence studies using outcome measures found 29 studies that reported positive associations between patient experience and clinical outcomes, 11 with no associations, and none with negative associations. The lack of more evidence may be due to associations between a patient's illness level, their level of care, and their likelihood for a poor outcome such as mortality, morbidity, or a readmission. Often, such associations have more than one plausible direction of causality. For example, clinicians may be especially attentive to the needs of sicker patients (Kahn et al., 2007) and patients near the end of life (Elliott, Haviland et al. 2013).

Research suggests an association between better patient experiences and lower healthcare utilization. Children with asthma were less likely to visit the emergency department, make urgent office visits, or be hospitalized if their physicians had reviewed a long-term therapeutic plan with their parents (Clark, Cabana et al. 2008). Among African Americans with Type 2 diabetes, those who reported that doctors or nurses usually listened carefully or spent enough time with them were significantly less likely to visit the emergency department in the 12 months following completion of a patient experience survey (Gary, Maiese et al. 2005). Fenton et al. found that patients who rated their providers most highly had lower odds of visiting the emergency department but higher odds of being admitted to the hospital the following year (Fenton, Jerant et al. 2012). Children whose parents report longer waits for primary care visits were more likely to visit the emergency department for non-urgent reasons than those who report waiting for less time (Brousseau, Bergholte et al. 2004).

Bibliography

AHRQ CAHPS Ambulatory Care Improvement Guide: Practical Strategies for Improving Patient Experience. (Feb 2018). Available at http://www.ahrq.gov/cahps/quality-improvement/improvement-guide/improvement-guide.html

Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, Edgman-Levitan S, Cleary PD. (2014) Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev* 71 (5):522-54.

Beauvais, B, Wells, R., Vasey, J., and DelliFraine, J. (2007). Does money really matter? The effects of fiscal margin on quality of care in military treatment facilities. *Hosp. Top.* 85(3): 2-15.

Brousseau, D. C., J. Bergholte, et al. (2004). The effect of prior interactions with a primary care provider on nonurgent pediatric emergency department use. *Archives of Pediatrics & Adolescent Medicine*. 158(1): 78-82.

Carcaise-Edinboro, P. and C. J. Bradley (2008). Influence of patient-provider communication on colorectal cancer screening. *Medical Care*. 46(7): 738-745.

Clark, N. M., M. D. Cabana, et al. (2008). The clinician-patient partnership paradigm: Outcomes associated with physician communication behavior. *Clinical Pediatrics.* 47(1): 49-57.

Cleary PD. (2016) Evolving concepts of patient-centered care and the assessment of patient care experiences; optimism and opposition. *J Health Pol, Policy & Law* 41 (4): 675-696.

Doyle, C., L. Lennox, et al. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ open* 3(1). http://bmjopen.bmj.com/content/3/1/e001570.full

Elliott, M.N., Farley, D., Hambarsoomians, K., and Hays, R.D. (2005). Do Medicaid and commercial CAHPS scores correlate within plans?: a New Jersey case study. *Med Care* (43(10): 1027-1033.

Elliott, M. N., A. M. Haviland, et al. (2013). Care experiences of managed care Medicare enrollees near the end of life. *Journal of the American Geriatrics Society* 61(3): 407-412.

Fenton, J. J., A. F. Jerant, et al. (2012). The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Archives of Internal Medicine.* 172(5): 405-411.

Fuertes, J. N., L. S. Boylan, et al. (2009). Behavioral Indices in Medical Care Outcome: The Working Alliance, Adherence, and Related Factors. *Journal of General Internal Medicine*. 24(1): 80-85.

Gary, T. L., E. M. Maiese, et al. (2005). Patient satisfaction, preventive services, and emergency room use among African-Americans with type 2 diabetes. *Disease Management*. DM 8(6): 361-371.

Harvard Pilgrim Health Care Quality Program Description (2017). Applicable to Harvard Pilgrim Health Care's Commercial, Marketplace Exchange & Medicare products. Accessible at https://www.harvardpilgrim.org/public/docs/2017-quality-program-description

Kahn, K. L., Tisnado, D. M., et al. (2007). Does ambulatory process of care predict health-related quality of life outcomes? Health Services Research. 42: 63-83.

Liu, Y., J. L. Malin, et al. (2013). Adherence to adjuvant hormone therapy in low-income women with breast cancer: the role of provider-patient communication. *Breast Cancer Research and Treatment.* 137(3): 829-836.

Lorig KR, Sobel DS, Stewart AL, et al. (1999). Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: A randomized trial. *Med Care* 37(1):5-14.

McWilliams JM, Landon BE, Chernew ME, Zaslavsky AM. (2014). Changes in patients' experiences in Medicare Accountable Care Organizations. *N Engl J Med.* Oct 30;371(18):1715-24.

Mittler, J., Landon, B., Fisher, E., Cleary, P., and Zaslavsky, A. (2010). Market variations in intensity of Medicare service use and beneficiary experiences with care. *Health Serv Res* 45(3): 647-669.

Ratanawongsa, N., A. J. Karter, et al. (2013). Communication and medication refill adherence: the Diabetes Study of Northern California. *JAMA Internal Medicine.* 173(3): 210-218.

Rodriguez HP, Scoggins JF, von Glahn T, Zaslavsky AM, Safran DG. (2009) Attributing sources of variation in patients' experiences of ambulatory care. *Med Care*. Aug;47(8):835-41.

Scanlon, D., Swaminathan, S., Lee, W., and Chernew, M. (2008). Does competition improve health care quality? *Health Serv Res* 43(6): 1931-1951.

Schneider, E C, Zaslavsky AM, et al. (2001). National quality monitoring of Medicare health plans: the relationship between enrollees' reports and the quality of clinical care. *Medical Care*. 39(12): 1313-1325.

Sorra J, Khanna K, Dyer N, Mardon R, Famolaro T. (2012) Exploring relationships between patient safety culture and patients' assessments of hospital care. *Journal of Patient Safety* 8(3):131–139.

Zolnierek, K. B. and M. R. Dimatteo (2009). Physician communication and patient adherence to treatment: a meta-analysis. *Medical care*. 47(8): 826-834.

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.

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

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

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.

The CAHPS Health Plan survey focuses on patient-centered care, which is a key element of health care quality (IOM, 2001). It measures aspects of health plan and health care quality that are important to consumers and for which consumers are the best or only source of information (Cleary, Edgman-Levitan, 1997; Cleary, 2016). Use of this measure will benefit both patients and health plans:

1) Patients can use information from the measures to help make more informed choices about which health plan to use.
2) Health Plans and their providers can use data from the surveys for quality improvement initiatives and incentives.
3) Researchers can use data files from the surveys to help answer important health services research questions.

Citations:

Cleary PD, Edgman-Levitan S. (1997). Health care quality. Incorporating consumer perspectives. JAMA. 1997 Nov 19;278(19):1608-12.

Cleary PD. Evolving concepts of patient-centered care and the assessment of patient care experiences; optimism and opposition. J Health Pol, Policy & Law, 2016, 41 (4): 675-696.

Institute of Medicine. "Crossing the Quality Chasm: A New Health System for the 21st Century". March 1, 2001. Accessible at http://iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx.

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 data for the Adult Health Plan CAHPS was submitted by users to the HP-CAHPS database for surveys administered September 2016 – June 2017. The data includes 152 Adult Medicaid health plans surveys and has 65,053 respondents. For the Getting needed care composite, the mean top box score/proportion = 0.55 (or 55%); standard deviation (std) = 0.05, minimum (min) = 0.37, median = 0.56, and max = 0.66 (or 66%). For the Getting care quickly composite, the mean top box score/proportion = 0.59 (or 59%); standard deviation (std) = 0.05, minimum (min) = 0.43, median = 0.60, and max = 0.71 (or 71%). For the Communication composite, the mean top box score = 0.74; std = 0.04, min = 0.61, median = 0.75, and max = 0.88 (or 88%). For the Health plan information and customer service composite, the mean top box score = 0.67; std = 0.05, min = 0.53, median = 0.68, and max = 0.80 (or 80%). For the global rating of all health care, the mean top box score = 0.53; std = 0.06, min = 0.33, median = 0.54, and max = 0.75 (or 75%). For the global rating of personal doctor, the mean top box score = 0.65; std = 0.05, min = 0.50, median = 0.66, and max = 0.88 (or 88%). For the global rating of specialist, the mean top box score = 0.66; std = 0.05, min = 0.48, median = 0.66, and max = 0.82 (or 82%). For the global rating of health plan, the mean top box score = 0.57; std = 0.07, min = 0.37, median = 0.58, and max

= 0.74 (or 74%). Scores by decile are reported in the attached Excel file named HP_CAHPS_Supplemental_Tables_Apr_2019.xlsx (worksheet tab titled "Table 1b.2a").

The data for the Child Health Plan CAHPS includes data submitted by users to the HP-CAHPS database for surveys administered September 2016 – June 2017. The data includes 169 health plans and has 103,283 respondents. For the Getting needed care composite, the mean top box score/proportion = 0.60 (or 60%); standard deviation (std) = 0.06, minimum (min) = 0.42, median = 0.60, and max = 0.72 (or 72%). For the Getting care quickly composite, the mean top box score/proportion = 0.73 (or 73%); standard deviation (std) = 0.06, minimum (min) = 0.55, median = 0.73, and max = 0.86 (or 86%). For the Communication composite, the mean top box score = 0.78; std = 0.04, min = 0.66, median = 0.78, and max = 0.90 (or 90%). For the Health plan information and customer service composite, the mean top box score = 0.66; std = 0.08, min = 0.0, median = 0.67, and max = 0.84 (or 84%). For the global rating of all health care, the mean top box score = 0.67; std = 0.05, min = 0.50, median = 0.68, and max = 0.79 (or 79%). For the global rating of personal doctor, the mean top box score = 0.75; std = 0.04, min = 0.60, median = 0.75, and max = 0.84 (or 84%). For the global rating of specialist, the mean top box score = 0.71; std = 0.08, min = 0.0, median = 0.72, and max = 0.86 (or 86%). For the global rating of health plan, the mean top box score = 0.68; std = 0.07, min = 0.44, median = 0.69, and max = 0.83 (or 83%). Scores by decile are reported in the attached Excel file named HP_CAHPS_Supplemental_Tables_Apr_2019.xlsx (worksheet tab titled "Table 1b.2a").

The recent trends for the Adult survey (from 2016 to 2017) are as follows: Getting needed care increased from 54 to 55%, Getting care quickly increased from 58 to 59%, Doctor communication no change (74%), plan service increased from 67% to 68%, rating of personal doctor no change (65%), rating of specialist increased from 65% to 66%, rating of health care increased from 53% to 54%, and rating of health plan increased from 57% to 58%.

The recent trends for the Child survey (from 2016 to 2017) are as follows: Getting needed care no change 61%, Getting care quickly decreased from 74% to 73%, Doctor communication no change (78%), plan service decreased from 68% to 67%, rating of personal doctor increased from 74% to 75%, rating of specialist increased from 70% to 72%, rating of health care increased from 67% to 68%, and rating of health plan increased from 68% to 69%.

Data also displayed in tab "Table 1b.2b" in HP_CAHPS_Supplemental_Tables_Apr_2019.

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.

Not applicable.

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.*

Current performance data by gender and race/ethnicity are shown in the attached Excel spreadsheet within the Table 1b.4 tab. The data used represents 152 Adult Medicaid Health Plans submitted to the CAHPS database. A total of 65,053 adult plan enrollees were surveyed. The top box scores by White race compared to non-White race are as follows (respectively): Getting needed care: 56% vs. 55%; Getting care quickly 61% vs 59%; Doctor communication 75% vs. 74%; Plan service 68% vs. 68%; Rate doctor 66% vs. 66%; Rate specialist 67% vs. 66%; Rate health care 55% vs. 53%; Rate plan 58% vs. 59%. These differences for White versus non-White respondents were generally small. Consistent with previous analyses of CAHPS data (e.g., Zweifler, Hughes, and Lopez, 2010; Weech-Maldonado et al, 2008), non-white respondents reported a more positive Global Rating of their health plan. For the remaining measures, there were either no differences, or White

respondents were more positive, though the differences were generally small (2 percentage points or less). The top box scores by male gender compared to female gender are as follows (respectively): Getting needed care: 56% vs. 55%; Getting care quickly 60% vs 60%; Doctor communication 74% vs. 74%; Plan service 68% vs. 67%; Rate doctor 64% vs. 66%; Rate specialist 64% vs. 66%; Rate health care 53% vs. 54%; Rate plan 56% vs. 59%. These differences by gender were small. Female respondents gave more positive ratings for half of the measures, but the differences were generally small. The largest differences were found for Global Rating of Specialist and Health Plan (3 percentage points higher for females).

Citations:

Weech-Maldonado, R, Elliot M, Oluwole, A, Schiller K, and Hays R. (2008). Survey response style and differential use of CAHPS rating scales by Hispanics. Med Care 46(9): 963-968.

Zweifler, J, Hughes, S, and Lopez, R. (2011). Controlling for race/ethnicity: A comparison of California commercial health plan CAHPS scores to NCBD benchmarks. Int J Equity Health 25(9): 4.

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

Not applicable.

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):*

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

Access to Care, Person-and Family-Centered Care

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

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://www.ahrq.gov/cahps/surveys-guidance/hp/index.html

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)*

Attachment **Attachment:** HP_CAHPS_Supplemental_Tables_Apr_2019.xlsx

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:** adult-med-eng-hp50-2152a-636588806384254718.pdf

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. <u>For maintenance of endorsement:</u> 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. <u>For maintenance of endorsement,</u> please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

NA

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.*

<u>IF an OUTCOME MEASURE</u>, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

We recommend that CAHPS Health Plan Survey items and composites be calculated using a top box scoring method. The top box score refers to the percentage of patients whose responses indicated that they "always" received the desired care or service for a given measure. The top box numerator for each of the four Overall Ratings items is the number of respondents who answered 9 or 10 for the item; with a 10 indicating the "Best possible."

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)*

<u>IF an OUTCOME MEASURE,</u> describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Respondents describe their experiences accessing and using care, and interacting with their health plans, over the past 6 months (Medicaid) or 12 months (Commercial Health Plans).

For each individual item, the top box numerator is the number of respondents who answered "Always" (the most positive response) for the item. The top box composite score is the average proportion of respondents who answered "Always" across the items in the composite.

There are two steps to calculating a composite score for a health plan:

1. Calculate the proportion of patient responses with the most positive response for each item in a composite.
2. Calculate the mean top box proportions across all items in a composite to determine the composite's top box score.

Example: Applying the Proportional Scoring Method to the composite "Getting Care Quickly":

Step 1 – Calculate the proportion of cases in the top box or most positive response for each item in a composite

Example: Items in "Getting Care Quickly" (2 items) have four response options: Never, Sometimes, Usually, Always. The top box percentage for each item in the composite is the proportion of respondents who answered "Always."

- Item #1 "Got care for illness/injury as soon as needed" = Proportion of respondents who answered "Always" = 80%

- Item #2 “Got non-urgent appointment as soon as needed” = Proportion of respondents who answered “Always” = 90%

Step 2 – Average the top box item scores to form the overall composite top box score

Calculate the average top box score across the items in the composite. In the above example, the calculation would be as follows:

Top box score for “Getting Care Quickly” = (Item1 * Item2) / 2 = (80% + 90%) / 2 = 85%

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

The eligible population for the survey includes all individuals who have been enrolled in a health plan for at least 6 (Medicaid) or 12 (Commercial) months with no more than one 30-day break in enrollment. Denominators will vary by item and composite.

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).

For each individual item, the top box denominator is the number of respondents who answered the item. For each composite score, the denominator is the number of respondents who answer at least one item within the composite. Composite scores are the average proportion of respondents who gave the highest rating across the items in the composite (as discussed in S.5).

Survey population (adult survey): All adult (age 18 and older) health plan enrollees who have been enrolled in a health plan for a specified period (6 months or longer for Medicaid version, 12 months or longer for Commercial version) with no more than one 30-day break in enrollment.

Survey population (child survey): Parents of children (age 0-17) enrolled in a health plan who have been enrolled in for a specified period (6 months or longer for Medicaid version, 12 months or longer for Commercial version) with no more than one 30-day break in enrollment.

Denominator for Measures 1-4 (composites): The number of respondents who answer at least one item within the composite.

Denominator for Measures 5-8 (ratings): The number of respondents who answered the item.

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

Individuals are excluded from the survey target population if:

1) They were not continuously enrolled in the health plan (excepting an allowable enrollment lapse of less than 30 days).

2) Their primary health coverage was not through the plan.

3) Another member of his or her household had already been sampled.

4) They had been institutionalized (put in the care of a specialized institution) or are deceased.

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.)*

The following should be excluded from the denominator:

1) Individuals not continuously enrolled in the health plan (excepting an allowable enrollment lapse of less than 30 days) or those for whom their primary health coverage is not through the plan.

2) Individuals from a household that has already been sampled.

Some users also exclude a survey from scoring and analysis if someone else answered the questions (as a proxy) for the respondent. (Question #38 on Adult survey.)

Survey code specifications for how to code an appropriately skipped item, multiple marks or blank items and for how to use the CAHPS Analysis Program can be found by downloading the Instructions for Analyzing Data from

CAHPS® Surveys available at https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html.

The CAHPS Analysis Program computes scores for users, sponsors and vendors. The goal of the CAHPS Analysis Program is to provide the user with a flexible way to make valid comparisons of performance across units (e.g., plans).

The CAHPS macro calculates scores at the unit level (e.g. health plan) for all survey measures including individual survey items, ratings, and multi-item composite measures. The output from the program then compares the performance of an entity to the overall performance of units. If a user wants to adjust their results for responder characteristics, the CAHPS macro can adjust unit scores for variations across units such as for respondent age, education, mental health status, and general health status (herein referred to as case-mix).

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.)*

HP-CAHPS users that have collected data for different groups (i.e., strata) of people can analyze the data separately or together. If groups are analyzed together, no changes to the CAHPS Analysis Program are necessary.

Users can estimate separate case-mix adjustments on two different populations using the macro parameter SPLITFLG = 1 in the CAHPS analysis program. (The default value = 0.) An example of splitting the case-mix adjustments separately on two populations is when comparing Medicaid Fee-for-Service populations with Medicaid Managed Care populations.

If survey users want to combine data for reporting from different sampling strata, they will need to create a text file that identifies the strata and indicates which ones are being combined and the identifier of the entity obtained by combining them.

Reference: Instructions for Analyzing Data from

CAHPS® Surveys: Using the CAHPS Analysis Program Version 4.1 available at

https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2015-instructions-for-analyzing-data.pdf.

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

Statistical risk model

If other:

S.12. Type of score:

Other (specify):

If other: 1. Top box score 2. Case-mix adjusted mean score

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.)*

Top Box Score Calculation:

1) Target Population = continuous enrollment in health plan for past 6 (12) months with no more than 30 day lapse in enrollment
2) Exclusions = lapse in enrollment or enrollment less than 6 (12) months, household already represented in sample, primary health care is not with this health plan
3) Screener items identify beneficiaries who meet the target process for each composite, such as whether the beneficiary sought any medical care, saw a personal doctor, saw a specialist, or interacted with the health plan's customer service. Composites are only calculated using enrollees who experienced a particular service/process.
4) Top box scores (percent with highest rating) are computed for each item
5) Top box scores are averaged across the items within each composite, weighting each item equally.

Users can adjust the survey data for characteristics such as self-reported respondent age, education, mental health status, and general health status. The CAHPS Analysis Program—often referred to as the CAHPS MACRO—is a free program written in SAS (version 6.0 or later) that enables survey users to case-mix adjust their data. The program also generates a distribution of survey results for each of the measures, calculates the mean score for both individual survey items and composite measures, and indicates whether an entity's scores are statistically different from the average. The most recent CAHPS Analysis Program can be found by downloading the Instructions for Analyzing Data from

CAHPS® Surveys available at https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html.

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.

Determining the Number of Sampling Units Needed: Plans Versus Products

The sample design should be based on the units for which users want to compare results, such as health insurance plans or products within health plans. For the purposes of this discussion, "Health insurance plan" is the entity that offers the health insurance (e.g., Plan A), and the "product" is the specific benefit plan design or coverage offered by the plan (e.g., Plan A's HMO product). Users should draw a sample for each health insurance plan or product about which they want to make inferences, separating plans into products, or other groups, such as if there are differences in geography, provider networks, or administrative structure. Also, each product might be treated separately if the benefits and coverage are different, since such differences might influence consumers' reports about their care. Users should analyze and report separate results for each of the unique groups for which they want to make inferences.

Defining the Sample Frame: Eligibility Guidelines

The sample that a vendor selects to survey should be drawn from a list of individuals (adults age 18 and older, or children 17 and younger) covered by the plan or product. This list, which typically would be provided by the sponsor, is the sample frame. Below are the CAHPS guidelines for determining who to include in the sample frame for the commercial survey (Medicaid survey):

- If surveying adults, include all individuals 18 years or older who have been enrolled in a health plan or product for 6 (12) months or longer, with no more than one 30-day break in enrollment during the 6 (12) months.

- If surveying children, include all individuals 17 years or younger who have been enrolled in a health plan or product for 12 (6) months or longer, with no more than one 30-day break in enrollment during the 12 (6) months.
- To identify those who have been enrolled in the plan or product for 12 (6) months or longer, use the anticipated start date of data collection to determine whether the person meets the 12 (6)-month eligibility requirement. For example, if the anticipated start date is March 1, 2016, include all those who have been continuously enrolled since March 1, 2015 (September 1, 2015).
- Allow the sample frame to include multiple individuals from the same household, but the sample drawn should not have more than one person (adult or child) per household.
- Include individuals with primary health coverage through the plan. Do not include individuals with only other types of coverage, like a dental-only plan.
- In the case of individuals who switch (or children who are switched) from one product to another within the same plan during the continuous enrollment period, count them as enrolled in the product in which they were enrolled the longest. For example, in the last 6 months, if the individual who was enrolled in a health plan's HMO product for 4 months switched to the same health plan's POS product, consider that person continuously enrolled in the health plan's HMO product.

The following section explains how to calculate the appropriate sample size for the CAHPS Health Plan Survey. The instructions are the same for both the Adult and Child versions as well as the Commercial and Medicaid versions.

Calculating the Sample Size for the Adult (Child) Questionnaire

It is recommended that the user select enough individuals to obtain approximately 300 completed adult (child) questionnaires per plan/product. For example, for an anticipated response rate of 50 percent, the user would need to start with a minimum sample size of 600.

If users anticipate that poor contact information (addresses and telephone numbers) will decrease the number of questionnaires that reach the sampled individuals, a larger sample may be needed.

If one or more of the plans do not have a membership large enough to draw the required sample size, the sample will be everyone in the health plan enrollee population who meets all of the eligibility criteria. Even under these circumstances, the sample may include only one adult (child) per household.

Sampling information is provided to users as part of the "Fielding the CAHPS Health Plan" survey document available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/hp/fielding-the-survey-hp50-2013.pdf.

Proxy Respondents

CAHPS Health Plan does allow for proxy respondents for mail and web-based mode.sd At the end of the survey, there is an item that asks "Did someone help you complete this survey?" If the answer is Yes, the follow-up question is "How did that person help you?" and they are to mark one or more of these response items:

1. Read the questions to me
2. Wrote down the answers I gave
3. Answered the questions for me
4. Translated the questions into my language
5. Helped in some other way

However, these the last two questions of the core questionnaire are not included in telephone scripts because telephone interviews should not be conducted with proxy respondents.

More details are provided in the document titled "Preparing a Questionnaire Using the CAHPS Health Plan Survey 5.0" updated 8/21/2017 and available for download at https://www.ahrq.gov/cahps/surveys-guidance/hp/index.html. (Click on Health Plan Survey 5.0 and Instructions).

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.

Users should choose a data collection protocol that maximizes the survey response rate at an acceptable cost. Some sponsors, as well as researchers conducting field tests, have found that the mail with telephone follow-up method is most effective: results from the CAHPS demonstration sites indicate that the telephone follow-up often adds 10 to 15 percentage points to the response rate obtained with mail alone.

This section provides you with a protocol for collecting responses by mail with telephone follow-up. Users can adapt this protocol to a mail-only or a telephone-only survey.

Conducting a Survey by Mail: Basic Steps

This section reviews the basic steps for collecting data through the mail and offers some advice for making this process as effective as possible.

- Set up a toll-free number and publish it in all correspondence with respondents. Assign a trained project staff member to respond to questions on that line. It is useful to maintain a log of these calls and review them periodically.
- Send an advance letter to the respondent. A well-written, persuasive advance letter authored by a recognizable organization (e.g., the sponsor or participating health care purchaser) will increase the likelihood that the recipient of the questionnaire will complete and return it within the deadline.

Tips for the letter:

  - Personalize the letter with the name and address of the intended recipient.
  - Have it signed by a representative of the sponsoring organization(s).
  - Check the cover letter for brevity and clarity and ensure that there are no grammatical or typographical errors.

Tips for the envelope:

  - Make it look "official" but not too bureaucratic; it should not look like "junk mail."
  - Place a recognizable sponsor's name—such as the name of a government agency, where applicable—above the return address.
  - Mark the envelopes "forwarding and address correction" in order to update records for respondents who have moved and to increase the likelihood that the survey packet will reach the intended respondent.

- Send the questionnaire with a cover letter and any special instructions to the respondent. Mail this package one week after the advance letter. Include a postage-paid return envelope to encourage participation.
- Send a postcard reminder to nonrespondents 10 days after sending the questionnaire. Some vendors recommend sending a reminder postcard to all respondents three to five days after mailing the survey instead of sending a postcard to nonrespondents only ten days after the survey is mailed. The reminder postcards serve as a thank you to those who have returned their questionnaires and as a reminder or plea to those who have not.
- Send a second questionnaire with a reminder letter to those still not responding 30 days after the first mailing.

- Begin follow-up by telephone with nonrespondents three weeks after the second questionnaire has been sent. Interviewers should attempt to locate respondents who have not responded to the mailed survey. The sponsor can either use a CATI script or a paper-and-pencil method to conduct the telephone interviews.
- The CAHPS consortium has found that a 40-50 percent response rate is achievable if users take steps to ensure the accuracy of the sample frame and carefully follow the recommended data collection protocol, including one or more attempts to follow up with non-respondents. Users are provided the following advice for improving response rates in the fielding guide:
- Improve initial contact rates by making sure that addresses and phone numbers are current and accurate (e.g., identify sources of up-to-date sample information, run a sample file through a national change-of-address database, send a sample to a phone number look-up vendor).
- Use all available tracing/tracking methods (e.g., directory assistance, Lexis-Nexis, Internet database services, GPS-based address verifications, and directories).
- Improve contact rates after data collection has begun (e.g., increase maximum number of calls, ensure that calls take place at different day and evening times over a period of days, mail second reminders, use experienced and well-trained interviewers).
- Consider using a mixed-mode protocol involving mail, telephone, and email data collection procedure (e.g., mail and telephone, email and mail, or email and telephone). In field tests, the combined approach was more likely to achieve a desired response rate than a single mode approach.
- Train interviewers on how to deal with gatekeepers.
- Train interviewers on refusal aversion/conversion techniques.

Data collection information is provided to users as part of the Fielding the CAHPS Health Plan Survey 5.0 CAHPS® Health Plan Survey and Instructions. Document No. 2013. Updated 8/4/2017 and available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/hp/fielding-the-survey-hp50-2013.pdf. More information on alternative modes of administration, and more detail on each specific step, can be found in the full fielding guide.

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.

CAHPS Health Plan Survey, Adult Version 5.0 (Medicaid and Commercial)

CAHPS Health Plan Survey, Child Version 5.0 (Medicaid and Commercial)

The English Adult Medicaid survey is available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/hp/adult-med-eng-hp50-2152a.pdf.

More surveys are available for download at https://www.ahrq.gov/sites/default/files/wysiwyg/health-plan5.0.zip

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)*

Health Plan

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

Outpatient Services

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

CAHPS_HP_NQF_submitted_Jan7_2019.docx,CAHPS_HP_NQF_submitted_Jan23_2019.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.

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 0006
Measure Title: CAHPS Health Plan Survey, Version 5.0
Date of Submission: 01/23/19

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: CAHPS Health Plan Survey Database*	☒ other: CAHPS Health Plan 2017 Database

*Metrics presented throughout are derived from analysis of the CAHPS Health Plan Survey, Adult Medicaid 5.0 and Child Medicaid 5.0.

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*).

Data administered September 2016 to June 2017 by Medicaid health plans using either the Child Medicaid 5.0 or Adult Medicaid 5.0 were used for testing. This data was accessed from the CAHPS Health Plan 2017 Database. The CAHPS Database aggregates the data to facilitate comparisons of CAHPS survey results by users, researchers, and other interested organizations. Information about the CAHPS database can be found at: http://cahpsdatabase.ahrq.gov/.

1.3. What are the dates of the data used in testing? September 1, 2016 – June 30, 2017

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

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

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*)

All health plans submitting Adult Medicaid Version 5.0 (152 plans) and Child Medicaid Version 5.0 (169 plans) results to the CAHPS database were included in the analysis. Health plan sponsors sample from their enrollees as described below in 1.6. Adult and child plans in this analysis each come from 33 states and the District of

Columbia, as shown in Table 1.5. A total of 65,053 respondents to the Adult survey and 103,283 respondents to the Child survey (completed by the child's parent, relative, or legal guardian) are included in the analysis. The Adult survey had an average of 428 respondents per plan, and the Child survey had an average of 611 respondents per plan.

Table 1.5 Geographic Distribution of CAHPS Health Plan Medicaid 5.0 Respondents, 2016-2017

Plan State	Adult Medicaid 5.0 Sample		Child Medicaid 5.0 Sample	
	Total Complete Records Within State	Total Plans Within State	Total Complete Records Within State	Total Plans Within State
ALABAMA	446	1	445	1
ARKANSAS	410	1	404	1
CALIFORNIA	3752	9	5116	8
COLORADO	613	2	364	2
DISTRICT OF COLUMBIA	790	2	2312	3
DELAWARE	584	1	421	1
FLORIDA	1723	6	4083	7
GEORGIA	1387	4	2847	5
HAWAII	652	2	2510	5
ILLINOIS	1591	4	4131	4
INDIANA	3320	7	3336	5
KANSAS	1590	3	4684	3
KENTUCKY	1619	4	1806	4
LOUISIANA	1692	4	2885	3
MASSACHUSETTS	1931	5	314	1
MARYLAND	4895	9	9952	9
MICHIGAN	5937	12	4881	12
MINNESOTA	5093	8	0	0
MISSOURI	0	0	1439	6
NEW HAMPSHIRE	374	1	926	1
NEW JERSEY	1353	3	1583	3
NEW MEXICO	1300	3	2606	3
NEVADA	481	1	1274	1
NEW YORK	2260	5	8974	20
OHIO	2372	5	4688	5
OREGON	4906	17	5149	17
PENNSYLVANIA	3771	9	4365	9
RHODE ISLAND	1050	2	1274	2
SOUTH CAROLINA	1160	3	2712	5
TENNESSEE	1045	1	2791	1
TEXAS	1387	3	3154	6
VIRGINIA	1971	5	3091	6
WASHINGTON	1793	5	6195	5
WISCONSIN	694	2	1298	2
WEST VIRGINIA	1111	3	1273	3
Total	**65,053**	**152**	**103,283**	**169**

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*)

The target population for the CAHPS Health Plan survey includes all individuals who had 6 (Medicaid plans) or 12 (Commercial plans) months of continuous plan enrollment with no more than one 30-day break in enrollment. CAHPS Health Plan fielding guidelines specify that a health plan's sample frame should be constructed using the anticipated start date of data collection to determine whether the person meets the continuous enrollment requirement. Only individuals with primary health coverage through the plan are included. Individuals who change products within a plan - for example, from an HMO to a PPO - are permitted on the sample frame and assigned to the product in which they were enrolled the longest.

Health plans are instructed to sample enough individuals to obtain approximately 300 completed surveys per plan/product. For example, assuming a 50 percent response rate, a health plan may randomly sample approximately 600 enrollees from the sample frame. Although multiple individuals in a single household may be on the sampling frame, the final sample must contain only one respondent per household. Where a household-level duplicate is sampled, that individual is discarded and replaced by another individual drawn at random from the frame. The fielding guidelines provide additional advice on drawing representative samples for multiple products and simultaneous sampling of adult and child enrollees: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/hp/fielding-the-survey-hp50-2013.pdf.

Tables 1.6a – 1.6e show descriptive characteristics of the individuals surveyed by the plans included in our analysis. Unfortunately, the data from the 2017 database did not include survey administration mode.

Table 1.6a. Gender Frequencies

Gender	Adult Survey %	Child Survey %
Female	61	47
Male	39	53

Table 1.6b. Race/Ethnicity Frequencies

Race Category	Adult Survey %	Child Survey %
White, Non-Hispanic	44	32
Black, Non-Hispanic	18	16
Hispanic or Latino	13	30
Other	21	16
Missing Race, Ethnicity	5	6

Table 1.6c. Age Category Frequencies

Adult Survey		Child Survey	
% 18 - 24	10	% 0 - 3	18
% 25 - 34	16	% 4 - 7	20
% 35 - 44	14	% 8 - 11	21
% 45 - 54	20	% 12-19	33
% 55 - 64	28	% Missing Age	7
% 65 - 74	5		
% 75+	2		
% Missing Age	5		

Table 1.6d. Self-reported Health Rating Frequencies

Response	Adult Survey %	Child Survey %
Excellent	10	36
Very good	21	33
Good	32	20
Fair	24	5
Poor	8	< 1
Missing	5	6

Table 1.6e. Self-reported General Mental Health / Emotional Health Frequencies

Response	Adult Survey %	Child Survey %
Excellent	19	42
Very good	22	24
Good	28	18
Fair	20	8
Poor	7	2
Missing	5	6

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

Not applicable; same data used for each aspect of testing below.

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.

HP-CAHPS 5.0 surveys collect patient-reported education level. In addition, information is available on the language in which the survey was completed. The tables below show the distribution of these two variables in the dataset.

Table 1.8a. Education Level Frequencies

Education Level	Adult Medicaid Survey %	Child Medicaid Survey %
8th grade or less	6	8
Some high school, but did not graduation	15	11
High school graduate or GED	36	30
Some college or 2-year degree	27	30
4-year college graduate	6	8
More than 4-year college degree	4	5
% Missing	6	8

Table 1.8b. % of Surveys Taken by Survey Language Type

Type	Adult Medicaid Survey %	Child Medicaid Survey %
English	93%	81%
Spanish	4%	14%
Other	3%	4%

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*)

There are multiple ways of estimating reliability. We used three approaches.

First, we estimated internal consistency reliability using the Cronbach's coefficient alpha for multi-item scales because this is commonly reported as a measure of the reliability of such scales. We estimated coefficient alpha for each composite. A reliability of at least 0.70 is considered acceptable for group-level comparisons (Nunnally and Bernstein, 1994). For composites with more than two items, we show the impact on Cronbach's alpha of deleting one of the items from the composite. However, CAHPS scores are designed to evaluate care across units of care such as plans, physician groups, and hospitals, not individual patients.

Therefore, next we assess reliability at the unit level which is the most relevant level of analysis for publicly reported CAHPS measures (Hays & Arnold, 1986, pp. 144-145). We calculated ICCs to assess the ratio between plan variance to within plan variation. The ICC provides the basis for determining the number of survey responses needed to obtain target-levels of reliability. Measures with relatively lower ICCs at the unit of interest indicate that a larger sample size of patient surveys are needed than for measures with higher ICCs. If a scale has a low ICC, that may imply that for a given sample size, and for units of comparable variability, the scale does not discriminate well among units or plans. ICCs above 0.05 indicate that the between group variance is greater than expected by chance and imply that nesting in groups does have an effect on the responses of individuals. The intraclass correlation is an appropriate estimate because the reliability of concern is for a measure at a single point in time rather than the average of the measure at two-time points.

Third, since CAHPS surveys are used to compare plans, site-level reliability (which is directly related to the standard error of measurement) is an important measure that is used to measure the number of responses to the survey needed to obtain reliable information (Hays, Shaul, et al., 1999). We measure site reliability on multi-item composite scores and global one-item scores, which partition within- and between-site variance. Higher levels of site reliability correspond to more accurate measurement of performance and better ability to distinguish performance among plans. Therefore, we do not feel conducting additional spline analysis, as described in Kaplan et al. (2009), is necessary.

Similar to internal consistency reliability (i.e., Cronbach's alpha), values of 0.70 and higher are considered acceptable for site-level reliability (Nunnally and Bernstein, 1994) and group comparisons. For example, CMS does not report (labeled as "Not available") any score for which reliability falls below 0.60 as that is considered very low reliability. CMS reports scores that meet the sample size threshold and for which reliability falls between 0.60 and 0.70, but flags these scores as having low reliability and alerts consumers to interpret such

scores with caution. Scores with reliability 0.70 or greater are reported without comment. Reliabilities of 0.85 or higher where possible are appropriate for applications such as pay-for-performance or actions that reward or classify individual plans.

The CAHPS grantees have reported the reliability of the CAHPS measures at the appropriate unit of comparison since the beginning of the project 23 years ago and for measure development throughout the project (e.g., Hays, Martino et al., 2014; Price, Stucky et al., 2018).

Citations:

Hays R, Arnold S. (1986). Patient and family satisfaction with care for the terminally ill. *Hospice Journal*, 7, 129-150.

Hays RD, Martino S, Brown J, Cui M, Cleary P, Gaillot S, Elliott M. (2014). Evaluation of a care coordination measure for the Consumer Assessment of Healthcare Providers and System (CAHPS®) Medicare Survey. *Medical Care Research and Review*, 71, 192-202.

Hays RD, Shaul JA, Williams VSL, Lubalin JS, Harris-Kojetin L, Sweeny SF, Cleary PD. , S. F. (1999) Psychometric properties of the CAHPS™ 1.0 Survey measures. *Medical Care,* 37, MS22-31.

Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. (2009). Improving the reliability of physician performance Assessment: Identifying the "physician effect" on quality and creating composite measures. *Medical Care*, 47(94), 378-387.

Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw Hill; 1994.

Raudenbush SW, Bryk AS. Hierarchical Linear Models. 2nd ed. Thousand Oaks, CA: Sage; 2002.

Price RA, Stucky B, Parast L, Elliott MN, Haas A, Bradley M, Teno JM. (2018). Development of valid and reliable measures of patient and family experiences of hospice care for public reporting. *Journal of Palliative Medicine*, 21, 924-932.

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*)

Tables 2a2.3a and 2a2.3b show the Cronbach's alpha for each composite in the Adult and Child surveys, respectively. For items within a composite consisting of 3 or more items, the Cronbach's alpha if the item were deleted is provided to determine if there was room for improving coefficient alpha by dropping an item. Table 2a2.3c shows the mean number of respondents per plan and plan-level reliability statistics for the surveys.

Table 2a2. 3a. Cronbach's Alpha Reliability Coefficient for CAHPS Health Plan Adult Medicaid Version 5.0 Sample, 2016-2017 (152 Plans, 65,053 Respondents)

Adult Survey Item	Standardized Cronbach's Alpha	Cronbach's Alpha if Item Deleted
GETTING NEEDED CARE COMPOSITE	0.61	
Easy to get care, tests, or treatment you needed		
Get appointment with specialist		
GETTING CARE QUICKLY COMPOSITE	0.67	
Got urgent care for illness/injury as soon as needed		
Got non-urgent appointment as soon as needed		
HOW WELL DOCTORS COMMUNICATE COMPOSITE	0.89	
Doctor explained things in a way that was easy to understand		0.88
Doctor listened carefully		0.84
Doctor showed respect for what patient had to say		0.85
Doctor spent enough time with patient		0.87
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	0.66	
Health plan customer service gave info or help patient needed		
Health plan customer service treated patient with respect		

Table 2a2. 3b. Cronbach's Alpha Reliability Coefficients for CAHPS Health Plan Child Medicaid Version 5.0 Sample, 2016-2017 (169 Plans, 103,283 Respondents)

Child Survey Item	Standardized Cronbach's Alpha	Cronbach's Alpha if Item Deleted
GETTING NEEDED CARE COMPOSITE	0.57	
Easy to get care, tests, or treatment you needed		
Got appointment with specialist as soon as needed		
GETTING CARE QUICKLY COMPOSITE	0.59	
Got urgent care for illness/injury as soon as needed		
Got non-urgent appointment as soon as needed		
HOW WELL DOCTORS COMMUNICATE COMPOSITE	0.85	
Doctor explained things in a way that was easy to understand		0.82
Doctor listened carefully		0.79
Doctor showed respect for what patient had to say		0.81
Doctor explained things in a way that was easy for child to understand		0.83
Doctor spent enough time with patient		0.82
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	0.67	
Health plan customer service gave info or help patient needed		
Health plan customer service treated patient with respect		

Table 2a2.3c. Plan-Level Reliability Statistics for CAHPS Health Plan Medicaid Version 5.0 Sample, 2016-2017

	Adult Medicaid 5.0 (152 health plans)			Child Medicaid 5.0 (169 health plans)		
Measures	Mean Respondents per plan	ICC	Plan-Level Reliability	Mean Respondents per plan	ICC	Plan-Level Reliability
Getting Needed Care	329	0.014	0.82	460	0.018	0.89
Getting Care Quickly	325	0.015	0.83	476	0.022	0.91
How Well Doctors Communicate	269	0.011	0.74	410	0.013	0.85
Health Plan Info and Customer Service	134	0.015	0.67	177	0.018	0.77
Global: Rating of personal doctor	327	0.012	0.80	509	0.006	0.76
Global: Rating of specialist	163	0.009	0.60	125	0.011	0.59
Global: Rating of all health care	313	0.011	0.78	444	0.008	0.79
Global: Rating of health plan	404	0.018	0.88	573	0.019	0.92

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

As shown in table 2a2.3a, removal of any questions in a composite would not result in a higher Cronbach's alpha. Therefore, we would not want to remove these questions. Cronbach's alphas in the adult survey ranged from 0.61 for the Getting Needed Care Composite to 0.89 for Doctor Communication. Cronbach's alphas in the child survey ranged from 0.57 for the Getting Needed Care Composite to 0.85 for Doctor Communication. Cronbach's coefficient alpha is not what is as important but rather target unit-level (e.g., plan) reliability. Nevertheless, this is an important aspect in health care that many plans are aiming to improve.

The ICC values range from 0.009 to 0.018 on the Adult survey and 0.006 to 0.022 on the Child survey. These values show that these measures would need a larger sample size to effectively discriminate among plans, however these values are also influenced by the sample size of the dataset used. Therefore, we also examined the site-level reliabilities.

All composites and global ratings completed by an average of 300 respondents per plan exhibit good plan-level reliability. Among Adults survey respondents, 6 of 8 measures have plan reliability greater than 0.70. Among Child survey respondents, 7 of 8 measures have plan-level reliability greater than 0.70; four of the seven with reliability equal to or greater than 0.85. One item – rating of specialist – achieved slightly lower reliability for both surveys. However, it should be noted that fewer than the recommended 300 individuals completed this item as a result of fewer people having recent experience with a specialist. Because the HP-CAHPS survey is used to compare health plans, the plan-level reliabilities are what is most important.

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)*

At the individual and plan level, we examined the relationships between each item's top box score, each composite's top box score and the top box score for the global measures of patient experience using Spearman rank-order correlations. We expect the composite measures to be moderately to strongly related to the overall ratings. As one example, the composite measuring how well respondents' personal doctor communicates is expected to be strongly related to respondents' overall rating of their personal doctor. Finding such a relationship supports interpretation of the composite as a valid measure of patient experience with a personal doctor.

We also examined Spearman rank-order correlations among the composites to assess the extent to which they measure different constructs. As measures of patient experience, we expected the composites to be correlated. However, very high intercorrelations indicate that the composites may not be unique enough to be considered separate measures.

One rule of thumb for correlations is:

- 0.10 is a small correlation
- 0.30 is a medium correlation and
- 0.50 is a large correlation.

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

Table 2b1. 3a. Individual-Level Correlation of Items, Composites and Global Ratings Mean Top Box Scores for CAHPS Health Plan Medicaid – ADULT Version 5.0 Sample, 2016-2017

Adult Survey Item (65,053 respondents)	Global Rating of Doctor	Global Rating of Specialist	Global Rating of Healthcare	Global Rating of Health Plan
GETTING NEEDED CARE COMPOSITE	**0.32**	**0.37**	**0.44**	**0.35**
Easy to get care, tests, or treatment you needed	0.33	0.30	0.44	0.35
Get appointment with specialist	0.22	0.35	0.30	0.26
GETTING CARE QUICKLY COMPOSITE	**0.23**	**0.22**	**0.30**	**0.23**
Got urgent care for illness/injury as soon as needed	0.22	0.23	0.33	0.27
Got non-urgent appointment as soon as needed	0.22	0.20	0.28	0.21
HOW WELL DOCTORS COMMUNICATE COMPOSITE	**0.57**	**0.30**	**0.41**	**0.29**
Doctor explained things in a way that was easy to understand	0.46	0.25	0.34	0.23
Doctor listened carefully	0.52	0.26	0.36	0.25
Doctor showed respect for what patient had to say	0.50	0.25	0.34	0.24
Doctor spent enough time with patient	0.49	0.26	0.36	0.26
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	**0.24**	**0.25**	**0.31**	**0.39**
Health plan customer service gave info or help patient needed	0.21	0.22	0.28	0.35
Health plan customer service treated patient with respect	0.21	0.22	0.24	0.32

Note: All correlations are statistically significant at p<0.001. Values are Spearman rank-order correlations among top box scores.

Table 2b1. 3b. Individual-Level Correlation of Items, Composites and Global Ratings Mean Top Box Scores for CAHPS Health Plan Medicaid – CHILD Version 5.0 Sample, 2016-2017

Child Survey Item (103,283 individuals)	Global Rating of Doctor	Global Rating of Specialist	Global Rating of Healthcare	Global Rating of Health Plan
GETTING NEEDED CARE COMPOSITE	**0.30**	**0.33**	**0.39**	**0.28**
Easy to get care, tests, or treatment you needed	0.31	0.27	0.40	0.29
Got appointment with specialist as soon as needed	0.19	0.31	0.26	0.22
GETTING CARE QUICKLY COMPOSITE	**0.21**	**0.18**	**0.27**	**0.18**
Got urgent care for illness/injury as soon as needed	0.18	0.16	0.25	0.17
Got non-urgent appointment as soon as needed	0.22	0.17	0.27	0.18
HOW WELL DOCTORS COMMUNICATE COMPOSITE	**0.49**	**0.25**	**0.38**	**0.23**
Doctor explained things in a way that was easy to understand	0.39	0.19	0.31	0.18
Doctor listened carefully	0.44	0.20	0.33	0.20
Doctor showed respect for what patient had to say	0.42	0.19	0.30	0.19
Doctor explained things in a way that was easy for child to understand	0.36	0.20	0.28	0.20
Doctor spent enough time with patient	0.42	0.21	0.32	0.19
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	**0.23**	**0.24**	**0.27**	**0.34**
Health plan customer service gave info or help patient needed	0.20	0.22	0.24	0.30
Health plan customer service treated patient with respect	0.20	0.20	0.23	0.31

Note: All correlations are statistically significant at p<0.001. Values are Spearman rank-order correlations among top box scores.

Table 2b1.3c. Plan-Level Correlation of Items, Composites and Global Ratings Mean Top Box for CAHPS Health Plan Medicaid - ADULT Version 5.0 Sample, 2016-2017

Adult Survey Item (152 Plans)	Global Rating of Doctor	Global Rating of Specialist	Global Rating of Healthcare	Global Rating of Health Plan
GETTING NEEDED CARE COMPOSITE	**0.53*** **	**0.44*** **	**0.52*** **	**0.54*** **
Easy to get care, tests, or treatment you needed	0.60***	0.41***	0.58***	0.55***
Get appointment with specialist	0.37***	0.40***	0.35***	0.41***
GETTING CARE QUICKLY COMPOSITE	**0.50*** **	**0.39*** **	**0.49*** **	**0.55*** **
Got urgent care for illness/injury as soon as needed	0.46***	0.38***	0.47***	0.48***
Got non-urgent appointment as soon as needed	0.42***	0.33***	0.38***	0.49***
HOW WELL DOCTORS COMMUNICATE COMPOSITE	**0.72*** **	**0.43*** **	**0.60*** **	**0.45*** **
Doctor explained things in a way that was easy to understand	0.59***	0.33***	0.48***	0.38***
Doctor listened carefully	0.66***	0.39***	0.50***	0.41***
Doctor showed respect for what patient had to say	0.64***	0.41***	0.59***	0.41***
Doctor spent enough time with patient	0.66***	0.40***	0.49***	0.37***
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	**0.34*** **	**0.26** **	**0.31*** **	**0.49*** **

Adult Survey Item (152 Plans)	Global Rating of Doctor	Global Rating of Specialist	Global Rating of Healthcare	Global Rating of Health Plan
Health plan customer service gave info or help patient needed	0.34***	0.25**	0.31**	0.45***
Health plan customer service treated patient with respect	0.24**	0.21*	0.21**	0.39***

***p<.001, **p<.01, *p<.05

Note: Values are Spearman rank-order correlations among top box scores.

Table 2b1.3d. Plan-Level Correlation of Items, Composites and Global Ratings Mean Top Box for CAHPS Health Plan Medicaid - CHILD Version 5.0 Sample, 2016-2017

Child Survey Item (169 Plans)	Global Rating of Doctor	Global Rating of Specialist	Global Rating of Healthcare	Global Rating of Health Plan
GETTING NEEDED CARE COMPOSITE	**0.43*** **	**0.32*** **	**0.52*** **	**0.34*** **
Easy to get care, tests, or treatment you needed	0.51***	0.18*	0.63***	0.38***
Got appointment with specialist as soon as needed	0.31***	0.38***	0.37***	0.25**
GETTING CARE QUICKLY COMPOSITE	**0.44*** **	**0.11**	**0.52*** **	**0.26*** **
Got urgent care for illness/injury as soon as needed	0.35***	0.05	0.41***	0.19*
Got non-urgent appointment as soon as needed	0.45***	0.16*	0.56***	0.29***
HOW WELL DOCTORS COMMUNICATE COMPOSITE	**0.57*** **	**-0.004**	**0.43*** **	**0.12**
Doctor explained things in a way that was easy to understand	0.48***	-0.02	0.35***	0.10
Doctor listened carefully	0.58***	0.04	0.52***	0.18*
Doctor showed respect for what patient had to say	0.62***	0.06	0.45***	0.24**
Doctor explained things in a way that was easy for child to understand	0.48***	-0.03	0.36***	0.06
Doctor spent enough time with patient	0.41***	-0.01	0.29***	0.06
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	**0.46*** **	**0.25** **	**0.52*** **	**0.57*** **
Health plan customer service gave info or help patient needed	0.44***	0.22**	0.47***	0.52***
Health plan customer service treated patient with respect	0.43***	0.23**	0.51***	0.55***

***p<.001, **p<.01, *p<.05

Note: Values are Spearman rank-order correlations among top box scores.

Table 2b1.3e. Plan-level Composite Top Box Intercorrelations for CAHPS Health Plan Medicaid - CHILD Version 5.0 Sample, 2016-2017

Composites	Getting Needed Care	Getting Care Quickly	Doctor Communication	Health Plan Service
Getting Needed Care	1	0.81***	0.64***	0.52***
Getting Care Quickly	0.80***	1	0.67***	0.47***
How Well Doctors Communicate	0.61***	0.54***	1	0.42***
Health Plan Info and Customer Service	0.52***	0.51***	0.43***	1

***p<.0001

Note: Values below the shaded diagonal are for the Adult survey sample, values above are for the Child survey sample. Values are Spearman rank-order correlations among top box scores.

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?*)

As hypothesized, most composites are strongly related to all of the global rating scales. For both Adult and Child surveys, the strongest predictor of personal doctor rating is how well that doctor communicated (e.g., Spearman's correlation of 0.72 for adults and 0.57 children at the plan level). The strongest predictors of health plan rating are respondents' ability to get the care they need and experiences with health plan customer service. All composites are substantially associated with overall healthcare rating in both Adult and Child surveys. There were some non-significant correlations for the Child survey between composites and ratings of specialists. This may occur because respondents are less likely to associate specialists with their health plan, personal physician, or overall access to healthcare. For the Child survey, there was no significant relationship between overall rating of health plan and doctor communication.

Although the composites should be correlated with each other, as they all measure aspects of patient experience, inter-correlations > 0.80 indicate that the composites may not be unique enough to be considered separate measures (O'Brien, 2007). Relationships are within acceptable range, with one exception. The correlation of top box means between "Getting Needed Care" and "Getting Care Quickly" slightly exceeded 0.80 (Adult, r = 0.80; Child, r=0.81). These are both measures of access, which explains their strong relationship among both adults and children.

Citation:

O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* 2007;41:673–690.)

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

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 2b4**.**

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 4 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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.

Health Plan CAHPS 5.0 results are not required to be risk adjusted by users. However, users of the survey, including public reporting entities, may voluntarily decide that they want to adjust the data to account for patient case-mix differences if comparing plans. For this purpose, users are able to find guidance and support in from the documents "Preparing Data for Analysis" (available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/preparing-data-for-analysis.pdf) and "Instructions for Analyzing Data from CAHPS Surveys" dated June 2017 (available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2015-instructions-for-analyzing-data.pdf). These documents contain instructions and specifications for coding the adjuster variables, imputing missing data for the adjusters, and for including them in analyses using the CAHPS Analysis Program in SAS. The variables chosen for adjustment and the steps for calculations of risk-adjusted scores are user-defined. Users must also decide whether or not to impute missing data for the adjusters at each adjuster's entity-level mean.

The CAHPS Analysis Program is a set of free programs written for SAS that enables survey users to conduct risk adjustment. The programs with test modules are available for download at https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html.

The CAHPS Analysis Program adjusts the data for case mix, generates a distribution of survey results for each of the measures, calculates the average score (the mean across all response categories) for both individual survey items and composite measures, and indicates whether an entity's scores are statistically different from the average. AHRQ's CAHPS Consortium developed the CAHPS Analysis Program to work with all CAHPS surveys. It is updated periodically to add functionality, produce additional types of output, and correct or debug issues with previous versions.

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.

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?

This section is used to describe the rationale for case-mix adjustment that is not required as part of the measures, but that CAHPS users may use. The standard methodology performed is case-mix adjustment by regression adjustment in a linear model. There is no ordering of risk factor inclusion. Without an adjustment,

differences in CAHPS scores between entities could be due to case-mix differences rather than true differences in quality.

The current CAHPS Analysis Program (described in section 2b3.1.1) suggests adjusting for general health status, mental health status, age, and education. Studies have found that patient and consumer survey responses about experiences and satisfaction with healthcare correlate with personal characteristics like general health, mental health/depression, education, and age (Simon et al., 2009; Rahmqvist and Bara, 2010; Zaslavsky et al., 2001; Martino et al., 2011; Elliott et al., 2009). Health status and age are two patient characteristics frequently found to be associated with patient reports about the quality of their medical care. People in worse health tend to report lower satisfaction and more problems with care than do people in better health perhaps because sicker patients have more complex health care needs and may tend to report more problems with coordination or communication. Older patients tend to report more satisfaction and fewer problems than do younger patients, although this association is usually not as strong as the one between health status and ratings (Hatfield and Zaslavsky, 2017; Eselius et al, 2008).

Education is a social factor self-reported by patients who take the CAHPS surveys. Studies have shown that more educated patients report more problems, perhaps because they have higher expectations rather than because they receive lower-quality care (Sofaer and Firminger, 2005). However, in a multivariate analysis using Medicare Advantage CAHPS data, Hatfield and Zaslavsky (2017) found that education had less influence on CAHPS dimension scores than self-reported general and mental health.

Different CAHPS surveys adjust for different variables and the variables included here are not the only adjustment factors, even for the health plan setting. CAHPS data can also be adjusted for other factors such as survey administration mode (Peipert et al., 2017). For example, a study by Drake and colleagues (2014) found that telephone respondents gave more positive responses than mail respondents did. Currently, the CAHPS database does not adjust for survey mode.

Citations:

Agency for Healthcare Research and Quality. (2017) THE CAHPS CLINICIAN & GROUP SURVEY DATABASE

How Results Are Calculated. Available at: https://www.cahpsdatabase.ahrq.gov/cahpsidb/Public/Files/Doc6_How_Results_are_Calculated_CG_2016.pdf

Drake KM, Hargraves JL, Lloyd S, Gallagher PM, Cleary PD. (2014) The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey. *Health Serv Res*. Jan 29. doi: 10.1111/1475-6773.12160. [Epub ahead of print]

Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, Giordano L. (2009) Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res.* Apr;44(2 Pt 1):501-18. doi: 10.1111/j.1475-6773.2008.00914.x.

Eselius LL, Cleary PD, Zaslavsky AM, Huskamp HA, Busch SH. (2008). Case-mix adjustment of consumer reports about managed behavioral health care and health plans. *Health Research and Educational Trust*. 43(6), 2014-2032.

Hatfield LA, Zaslavsky AM. (2017) Implications of Variation in the Relationships between Beneficiary Characteristics and Medicare Advantage CAHPS Measures. *Health Serv Res* Aug;52(4):1310-1329.

Martino, Elliott, Kanouse, Farley, Burkhart, Hays. (2011). Depression and the Health Care Experiences of Medicare Beneficiaries. *Health Services Research* 46 (6pt1): 1883–904.

Peipert JD, Brown JA, Cui M, Hays RD (2017). Differences in Mail and Telephone Responses to the CAHPS In-Center Hemodialysis Survey. *Ann Clin Nephrol* Vol.1 No.1: 1.

Rahmqvist M, Bara AC. (2010). Patient Characteristics and Quality Dimensions Related to Patient Satisfaction." *International Journal for Quality in Health Care* 22(2): 86–92.

Zaslavsky A, Zaborski L, Ding L, Shaul JA, Cioffi MJ, Cleary PD. (2001) Adjusting Performance Measures to Ensure Equitable Plan Comparisons. *Health Care Financing Review* 22 (3): 109–26.

Simon GC, Rutter M, Crosier J, Scott BH, Operskalski, Ludman E. (2009). Are Comparisons of Consumer Satisfaction with Providers Biased by Nonresponse or Case-Mix Differences? *Psychiatric Services* 60 (1): 67–73.

Sofaer S, Firminger K. (2005). "Patient Perceptions of the Quality of Health Services."

Annual Review of Public Health 26: 513–59.

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?

Eselius and colleagues (2008) published results of an analysis of case-mix adjustments for CAHPS focused on behavioral health services and health plans. After selecting appropriate adjusters based on explanatory power in separate linear regression models, the authors determined the impact of case-mix adjustment on their sample health plans. Specifically, they examined the size of the adjustments and the extent to which adjustments impacted the ranking of health plans. Case-mix adjustments had only modest effects on health plan ratings and rankings. The authors also found that mental health status was a strong predictor of patient experience. For this purpose, mental health status was added to HP-CAHPS version 5.0.

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.**

The analysis for the selection of education as a factor for adjustment is described in 2b3. 3a.

The table below is a subset of results reported in Table 1 (page 1314) of Hatfield and Zaslavsky's 2017 paper that focused on CAHPS case-mix adjustment and included education as a risk. Shown are the coefficients from a multivariate regression model using 818,896 Medicare individual CAHPS survey responses across 592 Medicare Advantage plans. The model predicted plan-level scores for each dimension. Predictors were: self-reported general and mental health status (poor, fair, good, very good, or excellent coded as 1–5) and self-reported education (no high school to graduate education coded as 1–6). The coefficients reported below were scaled to represent the effect of a unit change in the predictor relative to (divided by) the standard deviation of the health plan intercepts (adjusted health plan means) of each quality measure. In general, the results show that healthier and less educated people report more positive experiences with care. However, the impact of adjusting for education is minimal.

Table 2b3.4b. The Effects of Health Status and Education on CAHPS Experience of Care Scores from Hatfield and Zaslavsky's 2017 paper titled *Implications of Variation in the Relationships between Beneficiary Characteristics and Medicare Advantage CAHPS Measures*

CAHPS Quality Measure	Composite Score Range	*Global Adjustment Coefficients*		
		General Health	Mental Health	Education
Doctor Communication Composite Explains (1-4 score) Listens (1-4 score) Respects (1-4 score) Spends time (1-4 score)	4-16	0.79	1.20	-0.09
Rate care (1 item)	0-10	0.90	0.89	-0.18
Coordination of Care Composite Help managing care (1-3 score) MD recommends (1-4 score) Talk about meds (1-4 score) Quick test results (1-4 score) Test follow-up (1-4 score) Specialist coordination (1-4 score)	6-23	0.32	0.56	-0.15
Getting Care Quickly Illness care (1-4 score) Routine care (1-4 score)	2-8	0.29	0.36	0.05
Plan customer service Get info (1-4 score) Courteous (1-4 score) Easy forms (1-4 score)	3-12	0.34	0.45	0.20
Rate plan	0-10	0.47	0.47	0.35
Rate drug plan	0-10	0.45	0.40	0.43
Getting drug information Drug info (1-4 score) Courteous (1-4 score) Coverage info (1-4 score)	4-16	0.33	0.38	0.05
Getting needed drugs Easy to fill (1-4 score) Easy by mail (1-4 score)	2-8	0.35	0.62	0.13
Flu shot	0-1	0.39	0.01	0.14

Citations:

Hatfield LA, Zaslavsky AM. (2017) Implications of Variation in the Relationships between Beneficiary Characteristics and Medicare Advantage CAHPS Measures. *Health Serv Res.* Aug;52(4):1310-1329.

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*)

The Pearson product-moment correlation coefficient is widely used and understood. It assesses the linear association between the adjusted and unadjusted scores and ranges between -1 to 1. Because the ranking of scores is often important in public reports of CAHPS results, we also calculate Kendall's Tau (Kendall rank correlation coefficient). Tau is the correlation between rank orders of the adjusted and unadjusted scores.

The Kendall Tau statistic also has a range of–1 to +1, so that it has a range comparable to other correlation coefficients. Tau can be interpreted as the percent of pairs of units (e.g., plans) that switched ordering as a consequence of case-mix adjustment [100*(1-Tau)/2].

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

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

To quantify the effect of case-mix adjustment on the ranking of plans, we calculate the Pearson product moment, Kendall Tau correlation coefficients, and the maximum difference between the adjusted and unadjusted plan ratings. The adjustment factors include age, education, mental health status and general health status.

Table 2b3. 6a. Association between Adjusted and Unadjusted Mean Composite Scores: Adult Survey (152 Plans)

	Pearson Correlation	Kendall Correlation
GETTING NEEDED CARE COMPOSITE	0.97	0.83
GETTING CARE QUICKLY COMPOSITE	0.98	0.85
HOW WELL DOCTORS COMMUNICATE COMPOSITE	0.96	0.78
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	0.99	0.90

Table 2b3. 6b. Association between Adjusted and Unadjusted Mean Scores: Child Survey (169 Plans)

	Pearson Correlation	Kendall Correlation
GETTING NEEDED CARE COMPOSITE	0.98	0.89
GETTING CARE QUICKLY COMPOSITE	0.99	0.91
HOW WELL DOCTORS COMMUNICATE COMPOSITE	0.95	0.82
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	0.99	0.90

Table 2b3.6c Case-mix Adjusted Mean Scores for HP-CAHPS Adult Version 5.0 Sample, 2016-2017 (152 Plans)

Adult Survey Item	Maximum Difference Between Adjusted & Unadjusted Scores
GETTING NEEDED CARE COMPOSITE	0.08
Easy to get care, tests, or treatment you needed	0.09
Get appointment with specialist	0.09
GETTING CARE QUICKLY COMPOSITE	0.08
Got care for illness/injury as soon as needed	0.06
Got non-urgent appointment as soon as needed	0.13
HOW WELL DOCTORS COMMUNICATE COMPOSITE	0.07
Doctor explained things in a way that was easy to understand	0.06
Doctor listened carefully	0.07
Doctor showed respect for what patient had to say	0.06
Doctor spent enough time with patient	0.08
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	0.04
Health plan customer service gave info or help patient needed	0.05
Health plan customer service treated patient with respect	0.03

Table 2b3.6d. Case-mix Adjusted Mean Scores for CG-CAHPS Child Version 3.0 Sample, 2016-2017 (169 plans)

Child Survey Item	Maximum Difference Between Adjusted & Unadjusted Scores
GETTING NEEDED CARE COMPOSITE	0.12
Easy to get care, tests, or treatment you needed	0.14
Get appointment with specialist	0.10
GETTING CARE QUICKLY COMPOSITE	0.07
Got care for illness/injury as soon as needed	0.05
Got non-urgent appointment as soon as needed	0.09
HOW WELL DOCTORS COMMUNICATE COMPOSITE	0.09
Doctor explained things in a way that was easy to understand	0.09
Doctor listened carefully	0.08
Doctor showed respect for what patient had to say	0.07
Doctor explained things in a way that was easy for child to understand	0.11
Doctor spent enough time with patient	0.10
HEALTH PLAN INFO AND CUSTOMER SERVICE COMPOSITE	0.12
Health plan customer service gave info or help patient needed	0.15
Health plan customer service treated patient with respect	0.10

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

Not applicable.

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

Not applicable.

2b3.9. Results of Risk Stratification Analysis:

Not applicable.

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*)

There are different opinions about how to interpret the size of correlation coefficients. One rule of thumb is correlations of 0.10, 0.30 and 0.50 are often cited as small, medium and large, respectively.

The correlation statistics between pre- and post-adjusted plan rankings are above 0.80 for each of the composites. These high correlations show modest effects of case-mix adjustment on CAHPS Medicaid health plan mean scores. The adjustment effects were strongest for the communication composite followed by getting needed care ("access") composite as those correlations were the lowest. As shown in Tables 2b3.9c and 2b3.9d, maximum absolute value difference between adjusted and unadjusted scores for individual plans, ranges from 0.03 to 0.13 for Adult and 0.05 to 0.15 for Child.

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

_____

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 identify statistical significance of group-level differences, we use a statistical test (t-test) of case-mix adjusted mean scores for individual items, composites and global ratings for a plan versus the mean for all plans with p<.05 used as the criterion for determining significance. Because of the large sample size within the CAHPS database and within individual health plans, relatively small differences between health plans may be statistically significant.

The CAHPS analysis program allows users to test for both statistically significant and meaningful differences in performance. Users specify the size of the difference required for substantive significance in terms of an absolute size difference or a specified fraction of the distance between the entity and the nearer of upper and lower bounds on the measure.

Health Plans, purchaser groups, State agencies, and others have participated in the Health Plan CAHPS Database by submitting yearly results from the CAHPS Health Plan Survey. Database participants currently include State Medicaid agencies, Children's Health Insurance Programs (CHIP), and individual health plans serving Medicaid and CHIP enrollees. Comparative data for Medicare health plan enrollees are obtained from the Centers for Medicare & Medicaid Services (CMS). Products that assist with the determination of meaningful differences in performance have been developed from the data contributed to the CAHPS Health Plan Survey Database. These products include an annual Chartbook, the Online Reporting System, and Private Feedback Reports. For comparison purposes, the CAHPS online reporting system displays the percentage of health plans (HP) that scored at or below a particular top box score and a Report Builder feature allows users to create custom downloadable reports. Furthermore, organizations that contribute data to the CAHPS Database receive an Excel® report that displays their own results compared to the overall database average. More information about the CAHPS Health Plan Survey Database can be found at: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/cahps-database/about/cahps-database-2017-hp-chartbook.pdf. Instructions for Analyzing Data from CAHPS Surveys is available at: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2015-instructions-for-analyzing-data.pdf.

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*)

From the sample used for testing, the tables below report how many plans' mean item or composite score was significantly above or below the overall database mean item or composite score.

Table 2b4.2a CAHPS Health Plan Adult Medicaid 5.0: Plans Statically Significantly Above or Below the 2017 CAHPS Health Plan Database Average (N=152)

Survey Question or Composite	Count of Plans Significantly Above	Count of Plans Significantly Below	% Statistically Different from CAHPS Database Average
Getting Care Quickly Composite	28	20	32%
Got care for illness/injury as soon as needed	18	12	20%
Got non-urgent appointment as soon as needed	33	23	37%

Survey Question or Composite	Count of Plans Significantly Above	Count of Plans Significantly Below	% Statistically Different from CAHPS Database Average
Getting Needed Care Composite	31	23	36%
Easy to get care, tests, or treatment	32	23	36%
Get appointment with specialist	21	15	24%
How Well Doctors Communicate Composite	23	11	22%
Doctor explained things in a way that was easy to understand	19	12	20%
Doctor listened carefully	13	14	18%
Doctor showed respect for what you had to say	14	11	16%
Doctor spent enough time with you	21	12	22%
Health Plan Information and Customer Service Composite	14	13	18%
Health plan gave info or help needed	15	14	19%
Health plan treated patient with respect	13	8	14%
Global Rating of All Health Care	24	18	28%
Global Rating of Personal Doctor	29	18	31%
Global Rating of Health Plan	40	32	47%
Global Rating of Specialist	13	4	11%

Table 2b4.2b CAHPS Health Plan Child Medicaid 5.0: Plans Statistically Significantly Above or Below the 2017 CAHPS Health Plan Database Average (N=169)

Survey Question or Composite	Count of Plans Significantly Above	Count of Plans Significantly Below	% Statistically Different from CAHPS Database Average
Getting Care Quickly Composite	51	35	51%
Got care for illness/injury as soon as needed	41	19	36%
Got non-urgent appointment as soon as needed	43	37	47%
Getting Needed Care Composite	43	32	44%
Easy to get care, tests, or treatment	46	33	47%

Survey Question or Composite	Count of Plans Significantly Above	Count of Plans Significantly Below	% Statistically Different from CAHPS Database Average
Get appointment with specialist	32	22	32%
How Well Doctors Communicate Composite	28	24	31%
Doctor explained things in a way that was easy to understand	31	25	33%
Doctor listened carefully	18	18	21%
Doctor showed respect for what you had to say	24	12	21%
Doctor explained things in a way that was easy for child to understand	15	10	15%
Doctor spent enough time with you	37	30	37%
Health Plan Information and Customer Service Composite	31	21	31%
Health plan gave info or help needed	28	15	25%
Health plan treated patient with respect	29	11	24%
Global Rating of All Health Care	36	20	33%
Global Rating of Personal Doctor	22	19	24%
Global Rating of Health Plan	56	36	54%
Global Rating of Specialist	16	5	12%

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?*)

Results show that the measures as specified sufficiently discriminate between health plans in terms of patient reported quality. For both the Adult and Child surveys, more than 20 percent of health plans are significantly different from the benchmark for most items. Patient experiences with health plans are linked to their persistence in the plans. For example, one study found that the mean voluntary disenrollment rate among Medicare managed care enrollees is four times higher for plans in the lowest 10 percent of overall CAHPS Health Plan survey ratings than for those in the highest 10 percent (Lied et al, 2003).

Citation:

Lied TR, Sheingold SH, Landon BE, Shaul JA, Cleary PD. (2003). Beneficiary reported experience and reported voluntary disenrollment in Medicare managed care. *Health Care Finance Rev* 25(1):55–66.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Section not applicable – one set of specifications.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

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*)

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*)

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*)

Survey Non-Response

We are unable to assess the extent of missing data due to survey nonresponse from the CAHPS database, because we cannot calculate or empirically verify survey response rates provided by users for individual health plans. More specifically, we do not get any sample frame information from submitters in order to calculate survey non-response or even response rates. Klein and colleagues (2011) measured a 49% survey response rate for the 2007 Medicare CAHPS survey. They found that Asians, African Americans, and Hispanics responded at adjusted response rates 7–17 percentage points lower than non-Hispanic Whites ($p < .001$ for each). Among seniors, response rates dropped beyond age 75. Breakoff from telephone surveys was most common among African Americans and older respondents.

The CAHPS consortium has found that a 50 percent response rate is achievable if users take steps to ensure the accuracy of the sample frame and carefully follow the recommended data collection protocol, including one or more attempts to follow up with non-respondents (AHRQ, 2008). This section presents the guidance provided to users for calculating response rates.

The response rate is the total number of completed questionnaires divided by the total number of respondents selected. For CAHPS analyses and reports, this rate is adjusted as shown in the following formula:

$$\frac{\text{Number of completed returned questionnaires}}{\text{Total number of individuals selected} - (\text{deceased} + \text{ineligible})}$$

In calculating the response rate, users do not exclude respondents who refused, whom they were unable to reach because of bad addresses or phone numbers, or who were unable to complete the questionnaire because of language barriers or because they were institutionalized or incompetent.

Listed below is an explanation of the categories included and excluded in the response rate calculation:

Numerator Inclusions:

- **Completed questionnaires.** A questionnaire is considered complete if responses are available for 50 percent of key CAHPS items and at least one composite item or rating item.

Denominator Inclusions:

- **Refusals.** The respondent (or parent or guardian of the sampled child) refused in writing or by phone to participate.
- **Nonresponse.** The respondent (or parent or guardian of the sampled child) is presumed to be eligible but did not complete the survey for some reason (never responded, was unavailable at the time of the survey, was ill or incapable, had a language barrier, etc.).
- **Bad addresses/phone numbers.** In either case, the respondent (or parent or guardian) is presumed to be eligible but was never located.

Denominator Exclusions:

- **Deceased.** In some cases, a household or family member may inform you of the death of the respondent or the sampled child.
- **Ineligible - not enrolled in the plan.** The respondent or the sampled child disenrolled from the plan, was never in the plan, or was enrolled in the plan for less than 6 (or 12) months.

Users are provided the following advice for improving response rates (AHRQ, 2008):

- Improve initial contact rates by making sure that addresses and phone numbers are current and accurate (e.g., identify sources of up-to-date sample information, run a sample file through a national change-of-address database, send a sample to a phone number look-up vendor).
- Use all available tracking methods (e.g., directory assistance, Lexis-Nexis, CD-ROM directories, Internet database services and directories).
- Improve contact rates after data collection has begun (e.g., increase maximum number of calls, ensure that calls take place at different day and evening times over a period of days, mail second reminders, use experienced and well-trained interviewers).
- Consider using a mixed-mode protocol involving both a mail and telephone data collection procedure. In field tests, the combined approach was more likely to achieve a desired response rate than did either mode alone.
- Train interviewers on how to deal with gatekeepers.
- Train interviewers on refusal aversion/conversion techniques.

Item Non-Response

The method used to construct CAHPS scores, discussed in sections S4-S11 of the main NQF submission form, maximizes the use of available data by averaging available individual-level responses in construction of an overall score for the health plan. For each individual item, the top box score is percentage of respondents who answered “Always” (the most positive response) for the item. The top box composite score is the average proportion of respondents who answered “Always” across the items in the composite. Because of this methodology, the health plan-level statistics presented here have no missing data.

We provide the percentage of cases with missing values at the item level below. It is important to note that most CAHPS patient experience items are applicable only for enrollees who have utilized certain services during the past 6 or 12 months. For example, if a respondent has not seen a personal doctor during the reference period, they are skipped through items about their experiences with doctors. As a result, many CAHPS Health Plan Survey items have high percentages of missing data overall, but when skip patterns are considered, the percentages of inappropriate missing data are much lower. The tables below show the percentage of cases with truly missing data on each item (i.e., the respondent should have answered the item but did not) as well as the percentage of cases that were appropriately skipped through the item.

Klein and colleagues (2011) found that among Medicare respondents, older age was the strongest predictor of item missingness (e.g., those 85 years and older failed to answer items at twice the rate of those aged 65–74

years, p < .001). Non-Hispanic Whites had lower rates of item missingness than other racial/ethnic groups (p < .001 for each; one-third lower than African Americans).

Citation:

AHRQ. 2017. "Fielding the CAHPS® Health Plan Survey 5.0. Sampling Guidelines and Protocols." Available at: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/hp/fielding-the-survey-hp50-2013.pdf.

Klein DJ, Elliott MN, et al. (2011). Understanding nonresponse to the 2007 Medicare CAHPS survey. *Gerontologist* Dec. 51(6): 843-55.

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*)

Table 2b6.2a. CAHPS Health Plan Adult Medicaid 5.0: Item-level Percent Missing (65,053 Respondents)

Survey Question or Composite	% Truly Missing	% Missing due to Appropriate Skip
Getting Care Quickly Composite		
Got care for illness/injury as soon as needed	3.5	56.9
Got non-urgent appointment as soon as needed	4.9	26.4
Getting Needed Care Composite		
Easy to get care, tests, or treatment	5.6	21.6
Get appointment with specialist	5.2	54.4
How Well Doctors Communicate Composite		
Doctor explained things in a way that was easy to understand	6.6	30.8
Doctor listened carefully	6.8	30.8
Doctor showed respect for what you had to say	6.8	30.8
Doctor spent enough time with you	6.9	30.8
Health Plan Information and Customer Service Composite		
Health plan gave info or help needed	5.9	63.0
Health plan treated patient with respect	6.0	63.0
Global Rating of All Health Care	5.3	21.6
Global Rating of Personal Doctor	6.4	17.1
Global Rating of Specialist	5.8	56.1
Global Rating of Health Plan	5.5	N/A

Table 2b6.2b. CAHPS Health Plan Child Medicaid 5.0: Item-level Percent Missing (103,283 Respondents)

Survey Question or Composite	% Truly Missing	% Missing due to Appropriate Skip
Getting Care Quickly Composite		
Got care for illness/injury as soon as needed	2.3	63.4
Got non-urgent appointment as soon as needed	3.7	24.8
Getting Needed Care Composite		
Easy to get care, tests, or treatment	5.9	21.7
Get appointment with specialist	5.6	72.4
How Well Doctors Communicate Composite		
Doctor explained things in a way that was easy to understand	6.4	26.7
Doctor listened carefully	6.6	26.7
Doctor showed respect for what you had to say	6.6	26.7
Doctor explained things in a way that was easy for child to understand	7.7	47.5
Doctor spent enough time with you	7.2	26.7
Health Plan Information and Customer Service Composite		
Health plan gave info or help needed	6.6	64.5
Health plan treated patient with respect	6.7	64.5
Global Rating of All Health Care	5.7	21.7
Global Rating of Personal Doctor	6.1	10.7
Global Rating of Specialist	5.8	73.8
Global Rating of Health Plan	6.3	N/A

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

For survey non-response, CAHPS studies have shown that non-response weighting does not reduce bias beyond case-mix adjustment (Elliott, Edwards, et al., 2005; Elliott, Zaslavsky et al., 2009). More specifically, Elliott, Zaslavsky, and colleagues (2009) found that although individuals with lower response propensity had less positive evaluations of care, there was no evidence that non-response weighting based on available data, such as mode and patient characteristics, improved the accuracy of hospital scores beyond what could be achieved with patient mix adjustment.

For item non-response, less than 8 percent of cases are truly missing on all items, which suggests that our item-level results are likely not biased by systematic missing data due to item nonresponse.

The CAHPS Health Plan Survey adequately improves the quality of responses by using several screening questions to direct respondents through survey skip patterns. The screening questions result in a high percentage of missing data due to appropriate skips, as some health plan enrollees have not used the services queried in the survey. Survey item screeners have been found to reduce measurement error by ensuring that respondents who are not 'qualified' to answer a question are screened out instead of providing invalid responses (Rodriguez et al., 2009)

Citations:

Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD. (2005). Patterns of unit and item nonresponse in the CAHPS Hospital Survey. *Health Serv Res* 2005 Dec;40(6 Pt 2):2096-119.

Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, Giordano L. (2009). Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res* Apr;44(2 Pt 1):501-18..

Rodriguez HP, Glahn Tv, Li A, Rogers WH, Safran DG. (2009). The effect of item screeners on the quality of patient survey data: a randomized experiment of ambulatory care experience measures. *Patient*. Jun 1;2(2):135-41.

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: Survey sampling uses administrative enrollment data that is maintained by all health plans and easily accessible to produce a sampling frame. The data are collected by a survey of health plan enrollees.

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**.

Patient/family reported information (may be electronic or paper)

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).

CAHPS surveys can be administered via the Internet, but the surveys elicit reports about access and care in a way that is not captured in any electronic system.

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: Feasibility_Scorecard_v1.0-636588868092535968.xlsx

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.

The procedures for administering HP-CAHPS surveys have been standardized by the Centers for Medicare & Medicaid Services (CMS) and NCQA for many years. Because NCQA-accredited health plans are required to submit HP-CAHPS survey results to the NCQA, those plans most often contract with an NCQA-certified survey vendor to accurately collect and report CAHPS survey results. NCQA requires strict adherence to its standardized procedures and protocols for survey administration and collection. NCQA staff monitors each survey vendor's work and provides ongoing technical support to survey vendors. CMS contracts with vendors who are required to adhere to strict standards for survey administration and analysis. Both the NCQA and CMS place a high value on aligning requirements to assist in streamlining CAHPS measurement for Health Plans and for those plan enrollees who are being surveyed.

Achieving a desired response rate may be difficult for users. Phone is not optimal as the only mode of survey administration, but it is commonly used as a follow-up for CAHPS mail surveys. Phone follow-up can improve CAHPS response rates compared to mail-only (Burkhart et al., 2014; Fowler et al., 2002; Gallagher et al., 2005; Klein et al., 2011). A study of Medicare beneficiaries found that response rates continue to improve when up to 4 follow-up calls are made (Burkhart et al., 2014). In addition, phone follow-up calls help to achieve better representation of patients in terms of income, literacy/education, health status, age, gender, and race/ethnicity, above and beyond mail surveys alone (Tesler and Sorra, 2017).

The CAHPS Consortium continues to conduct research to develop and test survey administration methods that can improve the efficiency of data collection, enhance response rates, and gather more information about the experiences of those segments of the patient population that are hard to reach through more traditional means. This research includes: 1) studies comparing the effect of administration modes on response rates, survey scores, and data collection costs (e.g., mode comparisons have included in-office distribution vs. mail; email vs. mail); 2) studies assessing the effect of survey length on response rates and survey scores; 3) studies examining the impact of incentives on response rates; and 4) studies comparing the effect of different survey formats and design on survey responses. As part of this ongoing work, the Consortium sponsored a one-day invitational research meeting in September 2018 that convened a small group of survey users, researchers, CAHPS stakeholders, and policymakers to share results from recent research on survey methodologies that affect response rates and the representativeness of CAHPS survey data.

A summary of AHRQ's CAHPS Fall 2018 Research meeting is available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/survey-methods-research/summary-research-meeting.pdf.

References:

Burkhart Q, Haviland A, Kallaur P, et al. (2014) How much do additional mailings and telephone calls contribute to response rates in a survey of Medicare beneficiaries. Field Methods. 27(4):409-25.

Fowler Jr FJ, Gallagher PM, Stringfellow VL, et al. (2002) Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. Med Care 40(3):190-200.

Gallagher PM, Fowler FJ, Stringfellow VL. (2005) The nature of nonresponse in a Medicaid survey: causes and consequences. J Off Stat 21(1):73-87.

Klein DJ, Elliott MN, Haviland AM, et al. Understanding nonresponse to the 2007 Medicare CAHPS survey. (2011) Gerontologist 2011;51(6):843-55.

Tesler, R. and Sorra, J. CAHPS Survey Administration: What We Know and Potential Research Questions. (Prepared by Westat, Rockville, MD, under Contract No. HHSA 290201300003C). Rockville, MD: Agency for

Healthcare Research and Quality: October 2017. AHRQ Publication No. 18-0002-EF. Accessible at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/about-cahps/research/survey-administration-literature-review.pdf.

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)*.

The CAHPS Health Plan survey is available free of charge. In addition to the survey instrument, users can access comprehensive fielding, analysis, and reporting guides as well as SAS programming code (the CAHPS MACRO) that performs analysis of survey results. All of these tools are available at: https://www.ahrq.gov/cahps/surveys-guidance/hp/index.html . Requirements for submitting data to the CAHPS database, and for using the CAHPS name on an instrument, include:

- All core items must be present on the user's questionnaire
- No changes to core item wording are permitted
- Instruments must not omit any of the survey items related to respondent characteristics

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 NCQA Health Plan Rankings http://healthinsuranceratings.ncqa.org/2018/Default.aspx CMS Medicare Advantage https://www.medicare.gov/find-a-plan/results/planresults/planratings/compare-plan-ratings.aspx?PlanType=MAPD#plan_rating_health Qualified Health Plan (QHP) Quality Ratings https://www.healthcare.gov/quality-ratings/ NCQA Health Plan Rankings http://healthinsuranceratings.ncqa.org/2018/Default.aspx CMS Medicare Advantage https://www.medicare.gov/find-a-plan/results/planresults/planratings/compare-plan-ratings.aspx?PlanType=MAPD#plan_rating_health Qualified Health Plan (QHP) Quality Ratings https://www.healthcare.gov/quality-ratings/ Payment Program OPM's Federal Employee Health Benefits Plan Performance Assessment https://www.opm.gov/healthcare-insurance/healthcare/carriers/2017/2017-12a1.pdf Regulatory and Accreditation Programs NCQA Health Plan Accreditation http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx Quality Improvement (external benchmarking to organizations) CAHPS Database https://cahps.ahrq.gov/cahps-database/comparative-data/index.html Quality Improvement (Internal to the specific organization) Health Share of Oregon https://www.healthshareoregon.org/

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 Payment Program

a) Office of Personal Management FEHB Health Plan Performance Assessment project.

Implemented in 2016, OPM assesses of the annual performance of health plans contracted under the Federal Employees Health Benefits Program (FEHB) program. The performance assessment framework includes a discrete set of qualitative and quantifiable performance measures used to assess key aspects of performance. That overall assessment is linked to health plan profit factors. There are four primary categories of health plan performance to be assessed: clinical quality, customer service (CAHPS 5.0), resource use, and contract oversight. Each year Federal Employees Health Benefits (FEHB) plans send the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) survey to a sample of plan members to evaluate their health plan experiences. OPM publicly displays plan-level customer experience scores on their website after the consumer enters a zip code at https://www.opm.gov/healthcare-insurance/healthcare/plan-information/compare-plans/quality/.

c. The Federal Employees Health Benefits (FEHB) Program provides private health insurance to about 8.3 million federal employees, retirees, and their dependents. There are approximately 86 plans or carriers.

D. Public Reporting (1)

a. Name of Program and Sponsor: NCQA Health Insurance Plan Ranking

b. The purpose of publishing rankings is to make quality information on health plans available to consumers. NCQA's Health Insurance Plan Ratings lists private (commercial), Medicare and Medicaid health insurance plans based on their combined HEDIS®, CAHPS® and NCQA Accreditation standards scores. Commercial and Medicaid CAHPS data are submitted to NCQA; Medicare CAHPS data are submitted to a CMS contractor. Medicaid plans have the option to be scored on either Adult CAHPS or Child CAHPS data. NCQA's Health Insurance Plan Rankings use NCQA's established rankings methodology, which has been used and widely recognized since 2005. In 2018, NCQA rated over 1000 plans. The methodology used to rank health plans, including information on how CAHPS scores are used, can be found at https://www.ncqa.org/wp-content/uploads/2018/09/201808013_Health_Plan_Ratings_Methodology.pdf.

D. Public Reporting (2)

a. Name of Program and Sponsor: Medicare Advantage (MA) and Prescription Drug Plan (PDP) Program

b. CMS publicly reports plan-level CAHPS scores for consumers of Medicare Advantage Plans and Part D Prescription Drug Plans. The results from the Medicare CAHPS surveys are published in the Medicare & You handbook each Fall and on the Medicare Options Compare Web site (www.medicare.gov). The measures derived from the surveys are used by beneficiaries to help choose between plans, help contracts identify areas for quality improvement, and allow the public and research community to assess Medicare program performance. Beginning in 2012, several measures from MA & PDP CAHPS have been included in the Star Ratings for Medicare Advantage Quality Bonus Payments. General background information about the scoring of CAHPS®-based measures in the MA-PDP program is presented in the MA-PDP CAHPS® Survey: Quality Assurance Protocols and Technical Specifications (http://www.ma-pdpcahps.org/). State-level scores can be found at https://ma-pdpcahps.org/globalassets/ma-pdp/historic-data/state_national_mean_maffs_2018_final.pdf.

D. Public Reporting (3)

a) Patient Protection and Affordable Care Act - Exchange and Insurance Market Standards

b) Federal Health Insurance Exchanges and State-based exchanges on the federal platform (also known as the Health Insurance Marketplace) conduct the Qualified Health Plan (QHP) Survey, which is a version of the CAHPS Health Plan Survey. For 2019, CMS requires that QHP issuers use a HEDIS® Compliance Auditor and follow the HEDIS® Compliance Audit standards to validate the QHP Enrollee Survey sample frame and the clinical measure data. For the 2019 plan year, HealthCare.gov is continuing a pilot program to present health insurance plan quality ratings (or "star ratings") for some plans in Michigan, Montana, New Hampshire, Virginia, and Wisconsin (https://www.healthcare.gov/quality-ratings/). CMS intends to release guidance specifying the form and manner in which CMS will display 2019 QHP quality rating information at HealthCare.gov. For example, on HealthCare.gov, CMS anticipates referring to the Quality Rating System global rating as the "Overall Quality Rating," the Clinical Quality Management summary indicator as "Medical Care," the "Enrollee Experience" summary indicator as "Member Experience," and the "Plan Efficiency, Affordability, & Management" summary indicator as "Plan Administration." More information is available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/2019-QRS-and-QHP-Enrollee-Survey-Technical-Guidance_FINAL_20181016_508.pdf.

E. Quality Improvement with Benchmarking

a. Name of Program and Sponsor: CAHPS Database, Agency for Healthcare Research and Quality

b. The primary purpose of the CAHPS Database is to facilitate comparisons of CAHPS survey results by and among survey sponsors. This compilation of survey results from a large pool of survey users into a national

database enables participants to compare their own results to relevant benchmarks (i.e., reference points such as national and regional averages). The CAHPS Database also offers an important source of primary data for research related to patient and consumer assessments of quality as measured by CAHPS surveys.

c. The 2017 CAHPS Database includes Health Plan CAHPS 5.0 data from 152 Adult Medicaid Health Plans, 169 Child Medicaid Health Plans, and 23 Children´s Health Insurance Programs (CHIP).

F. Quality Improvement Internal

a. Health Share of Oregon

b. Health Share of Oregon, an organization serving Medicaid patients, used HP-CAHPS scores and improvement strategies to improve their patients' experience with customer service. As recommended in the CAHPS Improvement Guide, they developed an understanding of high-quality service from the perspective of members and patients and established service standards so that staff are clear on what is expected of them and how they should be interacting with members and patients. This case study was presented during an April 2016 Webcast hosted by the Agency for Healthcare Research and Quality. This Webcast highlighted the CAHPS Ambulatory Improvement Guide and the strategies that health plans can use to improve their enrollees' experiences. More details about this plan's internal improvement can be found at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/quality-improvement/reports-and-case-studies/health-share-of-oregon-case-study.pdf. Strategies for Improving Patient Experience with Ambulatory Care can be found in Section 6 of the Improvement Guide, available on AHRQ's website at https://www.ahrq.gov/cahps/quality-improvement/improvement-guide/improvement-guide.html.

c. Health Share of Oregon serves Oregon Health Plan members (Medicaid) in Clackamas, Multnomah, and Washington counties.

G. Regulatory and Accreditation Programs

a. Name of Program and Sponsor: NCQA Health Plan Accreditation

b. NCQA's accreditation program certifies that health plans meet basic requirements for consumer protection and quality improvement and is considered the industry's gold standard. CAHPS Health Plan Survey 5.0H, Adult Version (for Commercial and Medicaid plans) and CAHPS Health Plan Survey 5.0H, Child Version (for Medicaid plans) are part of the accreditation measure set. HEDIS Measures for Accreditation Scoring, can be found at https://www.ncqa.org/wp-content/uploads/2018/08/20190000_HEDIS_Measures_SummaryofChanges.pdf.

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?*)

Not applicable.

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.*)

Not applicable.

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.

Development

The CAHPS Health Plan survey has a long history of use dating from 1997. The CAHPS Health Plan survey has gone through four revisions since that time, using field and psychometric testing conducted by multiple partners, including NCQA, CMS, and other stakeholders to increase the scientific rigor and relevance of the survey and the usability of the data. All survey development has been conducted by the CAHPS Consortium, a public-private research collaborative.

Steps which have contributed to the content and design of the CAHPS Health Plan Survey over time have included:

- Literature review and review of existing measures
- Development and consultation with technical expert panels
- Focus groups with consumers
- Cognitive testing
- Field testing
- Public comment
- On-going collaboration and harmonization with key partners and stakeholders
- Input from the NCQA Task Force and review and approval by the NCQA Committee on Performance Measurement to ensure harmonization with NCQA Health Plan accreditation requirements

Implementation

The AHRQ CAHPS Database is a data repository designed to facilitate and support useful comparisons of selected CAHPS survey results. Each year, the CAHPS Database compiles data submitted voluntarily by users of the CAHPS Health Plan Survey (HP-CAHPS) and the CAHPS Clinician & Group Survey (CG-CAHPS).

Comparative results are published in an annual Chartbook as well as through an Online Reporting System (ORS) that displays the database average, regional, and other benchmarks. The HP-CAHPS Database ORS can be found at: https://cahpsdatabase.ahrq.gov/CAHPSIDB/Public/about.aspx

Organizations that contribute data receive private reports in Excel that display their own results compared to the overall database average. Contributors also have access to technical assistance through the CAHPS User Network. In addition, each year the CAHPS Database team conducts a series of outreach calls to survey vendors and Database users to gather feedback on the products and services provided, along with suggestions for improvement.

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.

The publicly available HP-CAHPS Database Online Reporting System is updated annually with new data submitted by CAHPS Health Plan survey users. Updates are sent out via email by GovDelivery for those who have signed up to receive updates from the CAHPS Databases. Organizations that contribute data to the CAHPS database receive a private feedback report in Excel that displays their own results compared to the overall database average.

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.

As part of CAHPS development and maintenance, the CAHPS Consortium has sought input from multiple users, including accreditors, health plans, and the public. Throughout the development process, the CAHPS Consortium has incorporated the data or input from these various sources in an incremental process of revision and refinement to develop measurement that is more precise and to produce survey data that would better meet the information needs of consumers and other stakeholders.

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

The CAHPS Consortium hears user feedback during research studies and development. Users can contact the CAHPS Database team with questions or comments by phone at 888-808-7108 or email at CAHPSDatabase@westat.com. We not aware of any substantial problems experienced by health plans.

4a2.2.3. Summarize the feedback obtained from other users

CAHPS consortium uses feedback from focus groups with patients in developing survey content and design. For HP-CAHPS version 5.0, the item about how often it was easy to get care was moved from the Your Health Plan section to the Your Health Care section because respondent feedback was that they had difficulty attributing this item to the health plan.

We not aware of any substantial problems experienced by respondents.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

The 5.0 version of the CAHPS Health Plan Survey incorporated some minor changes into the wording of core items based on input gathered in consultation with stakeholders. For example, questions about access to urgent and non-urgent appointments were modified to ask respondents if they were able to get an appointment "as soon as they needed," rather than as soon as "they thought" they needed for consistency across all CAHPS surveys.

The HP-CAHPS version 5.0 also changed the placement of one core item that also resulted in the deletion of a screener item and added a new item on self-reported mental health (for case-mix adjustment purposes).

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.

Progress on Improvement

Referring to data shown in the attached Excel spreadsheet Table 1b.2b tab, among adults, four measures demonstrated improvement between 2016 and 2017 and the remaining two had no change. For the Child survey, improvements were found for four of the eight measures. The largest improvement of 2 percentage points was seen in Global Rating of Specialist. Two measures had no change and two measures decreased by 1 percentage point.

The top box scores for both surveys indicate that there is room for continued improvement. The highest mean top box score in 2017 was 78% for Doctor communication (Child survey), and most scores are in the 50%-70% range. This data is also presented in the 2017 Health Plan Survey Database Chartbook (available at https://cahpsdatabase.ahrq.gov/files/2017CAHPSHealthPlanChartbook.pdf).

From 2007 to 2017, the top box score for all four composites and the provider ratings increased for the Adult Medicaid population. From 2007-2017, for the Child Medicaid population, the top box composite scores improved for only two of the composites: Getting needed care and Health plan information and customer service, yet there was improvement in the provider ratings. (These changes are displayed in figures 1 -3, pages

8-13 of the 2017 Chartbook, accessible at https://www.ahrq.gov/sites/default/files/publications2/files/cahps-database-2017-hp-chartbook_0.pdf). More trends can be seen in Figures 5-8.

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.

No unexpected findings.

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

Transparency of CAHPS scores to support consumer choice have influenced improvement in quality of care (Elliott, Lehrman, et al., 2010; Elliott, Cohea, et al., 2015; Cleary, 2016).

Sources:

- Elliott MN, Lehrman WG, Goldstein EH, Giordano LA, Beckett MK, Cohea CW, Cleary PD. (2010) Hospital survey shows improvements in patient experience. Health Aff 29(11): 2061-2067.
- Elliott MN, Cohea CW, Lehrman WG, Goldstein E, Cleary PD, Giordano LA, Beckett MK, Zaslavsky AM. (2015) Accelerating improvement and narrowing gaps: Trends in patients' experiences with hospital care reflected in HCAHPS public reporting. Health Serv Res, 2015, 50 (6): 1850-67.
- Cleary PD, Evolving concepts of patient-centered care and the assessment of patient care experiences; optimism and opposition. (2016) J Health Pol, Policy & Law 41 (4): 675-696.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

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

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

5a. Harmonization of Related Measures

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

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

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

not applicable

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.

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Agency for Healthcare Research and Quality

Co.2 Point of Contact: Caren, Ginsberg, caren.ginsberg@ahrq.hhs.gov, 301-427-1894-

Co.3 Measure Developer if different from Measure Steward: Agency for Healthcare Research and Quality

Co.4 Point of Contact: Caren, Ginsberg, caren.ginsberg@ahrq.hhs.gov, 301-427-1894-

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.

AHRQ

Caren Ginsberg, PhD, CAHPS Project Officer; Caren.Ginsberg@ahrq.hhs.gov

Elma Chowdhury; Elma.Chowdhury@AHRQ.hhs.gov

CAHPS Grantee RAND Team

Ron Hays, PhD, Principal Investigator, RAND; drhays@ucla.edu

Marc Elliott, PhD, Co-Principal Investigator, RAND; marc_elliott@rand.org

Julie Brown, RAND; julie_brown@rand.org

CAHPS Grantee Yale Team

Paul D. Cleary, PhD, Principal Investigator, Yale Team; Paul.Cleary@yale.edu

Susan Edgman-Levitan, PA, Co-Principal Investigator, Yale Team; SEDGMANLEVITAN@PARTNERS.ORG

Lee Hargraves, PhD, University of Massachusetts Medical School; lee.hargraves@umb.edu

CAHPS User Network, Westat

Joann Sorra, PhD, Project Director; JoannSorra@westat.com

Stephanie Fry; StephanieFry@westat.com

John Rauch; JohnRauch@westat.com

Lise Rybowski; lise@severyngroup.com

Dale Shaller, MPA; d.shaller@comcast.net

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 1997

Ad.3 Month and Year of most recent revision: 05, 2012

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

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

Ad.6 Copyright statement: CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality. CAHPS surveys are in the public domain.

Ad.7 Disclaimers: None.

Ad.8 Additional Information/Comments: