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 #: 0005

Corresponding Measures:

De.2. Measure Title: CAHPS Clinician & Group Surveys (CG-CAHPS) Version 3.0 -Adult, Child

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

De.3. Brief Description of Measure: The Consumer Assessment of Healthcare Providers and Systems Clinician & Group Survey 3.0 (CG-CAHPS) is a standardized survey instrument that asks patients to report on their experiences with primary or specialty care received from providers and their staff in ambulatory care settings over the preceding 6 months.

The CG-CAHPS 3.0 survey can be used in both primary care and specialty care settings. The adult survey is administered to patients aged 18 and over. The child survey is administered to the parents or guardians of pediatric patients under the age of 18. Patients who had at least one visit to a selected provider during the past 6 months are eligible to be surveyed.

CG-CAHPS Survey Version 1.0 was endorsed by NQF in July 2007 (NQF #0005) and version 2.0 received maintenance endorsement in early 2015. Version 3.0 was released in July 2015. The development of the survey is through the CAHPS Consortium and sponsored by the Agency for Healthcare Research and Quality. The survey is part of the CAHPS family of patient experience surveys and is available at https://cahps.ahrq.gov/surveys-guidance/cg/about/index.html

The Adult CG-CAHPS Survey 3.0 has 31 questions including one overall rating of the provider and 13 questions used to create these four multi-item composite measures of care or services provided:

1. Getting Timely Appointments, Care, and Information (3 items)
2. How Well Providers Communicate With Patients (4 items)
3. Helpful, Courteous, and Respectful Office Staff (2 items)
4. Providers' Use of Information to Coordinate Patient Care (3 items)

The Child CG-CAHPS Survey 3.0 has 39 questions including one overall rating of the provider and 12 questions used to create these four multi-item composite measures of care or services provided:

1. Getting Timely Appointments, Care, and Information (3 items)
2. How Well Providers Communicate With Patients (4 items)
3. Helpful, Courteous, and Respectful Office Staff (2 items)

4. Providers' Use of Information to Coordinate Patient Care (2 items)

1b.1. Developer Rationale: CG-CAHPS assesses aspects of health care delivery that are important to patients and for which patients are the best or only source of information (Solomon et al, 2005). Furthermore, the CG-CAHPS survey focuses on patient-centered care (Cleary, 2016), one of the six central aims identified by the Institute of Medicine for improving the health care system (IOM, 2001). A focus on the patient experience has the potential to enhance clinical outcomes, improve patient safety, and reduce unnecessary medical services. Moreover, assessing patient experience through surveys that include data on the demographic characteristics of respondents, such as race and ethnicity, can help identify the extent to which positive experiences are distributed equitably across patients (Haviland et al., 2003).

Citations:

Cleary, PD. Evolving Concepts of Patient-Centered Care and the Assessment of Patient Care Experiences: Optimism and Opposition. Journal of Health Politics, Policy and Law. Vol. 41, No. 4. August 2016.

Haviland, M. et al, (2003). Do Health Care Ratings Differ by Race or Ethnicity? Joint Commission Journal on Quality and Safety. 29(3), pp. 134-145.

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.

Solomon, L., Hays, R. D., Zaslavsky, A., & Cleary, P. D. (2005). Psychometric properties of the Group-Level Consumer Assessment of Health Plans Study (CAHPS?) instrument. Medical Care, 43, 53-60.

S.4. Numerator Statement: The CG-CAHPS Survey item and composites are often reported 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 the Overall Rating of Provider is the number of respondents who answered 9 or 10 for the item, with 10 indicating "Best provider possible".

For more information on the calculation of reporting measures, see

"Preparing Data from CAHPS® Surveys for Analysis" (AHRQ, 2017) accessible at

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

and the CAHPS Analysis Instructions accessible at
https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2015-instructions-for-analyzing-data.pdf (updated June 2017).

S.6. Denominator Statement: The measure's denominator is the number of survey respondents. The target populations for the surveys are patients who have had at least one visit to the selected provider in the target 6-month time frame. This time frame is also known as the look back period. The sampling frame is a person-level list and not a visit-level list.

S.8. Denominator Exclusions: Among eligible respondents, for a given item, respondents with a missing response is excluded. Among eligible respondents, for a composite measures, respondents who did not answer at least one item in the composite are excluded from the composite measure's denominator.

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

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Clinician : Group/Practice

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+ after a visit to a clinican.
- 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.
    - Developer cites QI activities such as shadowing, coaching and training, and offers other studies demonstrating the connection between workflow modifications and improved patient communication results.
    - Other interventions found to impact patient experience:
        - Clinic hour expansion
        - Joining a larger medical group
        - Improving infrastructure
        - Access to medical record

- Improving virtual access
- Provision of same-day or next-day appointments and access to a consistent clinician, group or care team
- Improvement in communication

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:** Developer indicated new evidence was presented since the last maintenance renewal (2015), but that evidence was not highlighted with red text per the instructions.

Questions for the Committee:

- Is the evidence presented which shows a link between clincian processes/ structures and consumer experience satisfaction persuasive?
- Is the correlation between patient satisfaction and more distal patient outcomes persuasive?

Guidance from the Evidence Algorithm

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

Preliminary rating for evidence: ☒ Pass ☐ No Pass

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

Maintenance measures – increased emphasis on gap and variation

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

Mean "top box"scores across dozens to hundreds (636 with at least 10 observations each) of providers corresponding to 112,000 consumers are reported. "Top box" achievement to the numerator means that an item is rated at 9 or 10 on a presumed (it is not clear from the numerator detail description) 10 point scale where 10 means the best possible quality of care is perceived by the cosumer. By domain here are the means (and standard deviations) observed.

Adult survey:

- Access = .68 (.10)
- Communication = .87 (.05)
- Care coordination = .73 (.07)
- Office site = .81 (.08)
- One item overall = .82 (.08)

Child survey (77 practices, 12K consumers):

- Access = .75 (.10)
- Communication = .90 (.05)
- Care coordination = .81 (.07)
- Office site = .76 (.10)
- One item overall = .83 (.09)

Disparities

- Gender specific differences were not evident in either the adult or child samples. In the adult samples there were some suggestions form the data that older consumers generally scored their providers higher than younger clients. For example, mean scores for 18-24 year olds on the communications composite was 84% and for 55-64 year olds it was 89%. Race data was not available for analysis.

Question 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.*

- Studies cited where patient experience is correlated with better preventive services, taking medications as prescribed and lower utilization of unneccessart health services
- Evidence indicates that the data from the CAHPS results MAY be used to improve Patient Experience.

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?*

- There was variability between medical groups and between clinics within a group. There has been some stratification by health status,age and education. Race data wasn't available for analysis but would be available for groups to evaluate differences if they collect this information
- There is definitely room for improvement. I wondered how "Care Coordination" was measured (I couldn't tell from the tool provided). It seems to be the one area that showed the least statistical reliability. Also, the Spanish speaking and Black populations may be under sampled.

Criteria 2: Scientific Acceptability of Measure Properties

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

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

Reliability

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

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

Validity

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

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

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

Evaluators: NQF Scientific Methods Panel (SMP)

Methods Panel Review (Combined)

Scientific Methods Panel Votes: Measure Passes

- Reliability: H-1, M-4, L-0, I-1
- Validity: H-1, M-5, L-0, I-0

Reliability

- Three reliability checks were run
- Data use: 635 (reduced from 656 to elimate sites with <10 observations) adult providers corresponding to 137K consumer; 77 (reduces from 81 sites) child providers corresponding to 12K consumers. Data was from 2016 to 2017.
- Outlier: Table 1.5a that Florida is a bit of outlier because it has only one practice site serving 24,850. The developers note that they only test at the practice level, not at the individual clinician level.
- Reliability test 1: Element level Cronbach's alpha (Ca) to check within domain consistency
    - Results: Access, Communication, Courteous/Helpful composites Ca >.75 (for adult and child)
    - Results: Care Coordination Ca < .39 (below reasonable)
- Reliability test 2: Score level ICC was used to consider the variability between entitites verus within entities. This application is not clear, other than that the developers say they are presenting are assessing "the ration between practice to within practice variation." The ICC they say is acceptable is >0.05 since it reveals that differences that are greater than expected by chance (presumably based on the heuristic of p<0.05?). The refernce of Hasy and Arnold, 1986, pp 144-145) is given as support for this approach. More basic description of the form of these comparisons would be useful.
    - ICCs always were below .046 (lower than desireable, suggesting high samples are needed)
- Reliabilty test 3: Here they seem to do a variation of the Cronbach's alpha by "We measure site (practice) reliability on multi-item composite scores and global one-item scores, which partition within- and between-site variance." It is otherwise not clear how these comparison are actually done otherwise, though they cite Hays, Shaul et al., 1999. More basic description of the form of these comparisons would be useful.
    - Site-level reliabily always greater than 0.71 (suggesting reasonable between practice versus within practice variability).

Validity

- Spearman rank order correlations between subscores and eachother, and between the overall rating were assessed. These were done both at the individual patient (n=112K) level and at the practice level (n=635).
- Results: At the individual patient level, Access, Comunication Courteous/Helpful, and Care Coordination composite scores all had correlations with the global score that exceeded 0.23 (p<.001). At the practice level these correlations exceeded 0.29 (p<0.001).

- Different factors correlated with one another though they did not suggest the different factors were completely redundant, as depicted in the table below.

Composites	Access	Provider Communication	Courteous/ Helpful Staff	Care Coordination
Access	1	0.39***	0.67***	0.41***
Provider Communication	0.49***	1	0.28*	0.59***
Office Staff Service	0.52***	0.44***	1	0.36**
Care Coordination	0.47***	0.68***	0.42***	1

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

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

- Missing data said to effect less than 5% of the individual ratings, though it was noted that accounting for response bias did not reduce any bias beyond the limited amount addressed by case-mix adjustment.
- No explicit exclusions were applied.
- Risk adjustment was not conducted, though a method and empirical coefficients to support case-mix illness adjustment is presented. Results presented further demonstrated a very high correlations (>0.85) between adjusted and unadjusted scores, obviating considerably the need for such risk adjustment.
- Performance gap, though somewhat subtle per section 1b (see above), is apparent in the presentations in Tables 2b4.2a-b which show that 30 to 40% of the sites perform at rates that are statistically distinct from the average rates observed.

Questions for the Committee regarding reliability:

- Do the results, which suggest marginal differences for Coordination of Care but stronger differences for the other factors, indicate the measure can be repeated with reasonable fidelity?
- 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:

- Are the validity tests sufficient to indicate the assessment is a true measure of quality of care related to patient satisfaction?
- Any concerns about missing data, or the absence of exclusions?

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

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

Measure Number: 0005

Measure Title: CAHPS Clinician & Group Survey **(CG-CAHPS)** Version 3.0

Measure Title: 0005 **– Adult, Child**

Type of measure:

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

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

Data Source:

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

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

☐ **Enrollment Data** ☐☒ **Other:**CAHPS Clinician and Group Database

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: None

   MP#2: Not so much a concern, but clarification that clinician-level identification in the CAHPS database is not reliable, so only practice-level are submitted

   MP#3: No testing was conducted at individual clinician level even though the measures are specified for both individual clinician and group practice level.

   The developer should clearly specify if these are risk adjusted measures and provide reliability and validity testing results accordingly.

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**

   MP#2:

   a. Clinician-level data not reliable and disregarded.

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   MP#4: Both critical data elements used in the measure and performance measure score was assessed. Cronbach's coefficient alpha used for multi-item scales and for each composite. Reliability of 0.70 acceptable per literature. For unit level reliability calculated ICCs to assess the ratio between practice variance to within practice variation. Site level reliability was conducted.

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

   Cronbach's alphas were calculated to assess the internal consistency of multi-item scales. However, it is worth noting that the measures as specified were not based on the scale scores. Instead of using the original response categories for each item, responses were dichotomized using top-box scaling. Second, a top-box response rate was calculated for each item each site, 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 patient level while the measures as specified were composited at the site level.

   MP#4: Cronbach's alpha used for each composite in surveys. Adequate sample sizes reported. Assessed site level reliability using ICC with acceptable results. Assessed ICC values for practices, which show that a larger sample size would be needed to effectively discriminate among practices.

   MP#1: Cronbach's alpha was generally good for each composite, but for the Care Coordination composite, in both the adult and child versions, internal consistency was fair (alpha= 0.55) to poor (alpha = 0.39) in the adult and child versions, respectively. Site level signal-to-noise estimates were good.

   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.

   MP#2:

   Cronbach's alpha, ICC and site-level reliability (signal-to-noise). Cronbach's alpha all good except care coordination for adults (0.55) and children (0.39). These are not acceptable. All others good as scaled. ICCs were all below 0.05. submitter states: "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." I do not understand this statement and generally considered all ICCs to be problematic, as they indicate that clinicians and sites may not be able to be differentiated. Site-level reliabilities all good and suggested by the submitters as most important in review.

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

   Submission document: Testing attachment, section 2a2.3

   MP#6: Submitters discussed inadequate results of reliability in some of the composite scores using the Cronbach's alpha method. They do note that the practice site level of reliability is adequate and that is what it is designed to measure.

   MP#3: The results of measure score reliability were very good, please note that the results were based on around 150 respondents per site. For sites with fewer survey responses, measure score reliability may be lower.

At the respondent level, internal consistency estimates were in general very good, although the results for Care Coordination were low, particularly for Child survey, only 0.39, unusually low. More importantly, given how the measures are specified, Cronbach's alphas as calculated do not provide the information required.

MP#1: Cronbach's alpha was generally good for each composite, but for the Care Coordination composite, in both the adult and child versions, internal consistency was fair (alpha= 0.55) to poor (alpha = 0.39) in the adult and child versions, respectively. Site level signal-to-noise estimates were good.

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.

MP#2: Test sample adequate for generalization. Care Coordination not acceptable based on poor scalability of multi-item scale (data element); others seem good on reliability, at least site-level. Moderate confidence

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

   Site-level reliability testing was appropriate

   ☒ **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** (considering 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: As noted, composite score reliability did not demonstrate strong reliability

MP#4: No testing done at the individual clinician level.

MP#3: NQF requires reliability testing for both data element and measure score for instrument-based measures. Reported internal consistency testing is inadequate.

MP#1: Site level STN estimates were acceptable.

MP#5: See response to item 7 above.

MP#2:

a. **Minor concern with care coordination going forward. Confused over ICC results and their position on same. Site-level reliability good**

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

MP#6: No exclusions

MP#2:

a. Not so much exclusions, but very low representation of Spanish speakers and other URM data, especially on adult side

Submission document: Testing attachment, section 2b2.

MP#4: NA

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#6: Since they are only using these at the practice level, I have no concerns

MP#2:

a. No concern here is care coordination excluded or moved to single/separate items

Submission document: Testing attachment, section 2b4.

MP#4: None

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 concerns

Submission document: Testing attachment, section 2b5.

MP#4: Different modes of data collection (mail, telephone, web based, or mixture) are indicated with recommendation for which ones produce the most comparable results.

MP#3: No concern.

MP#5: N/A

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

MP#6: Submitters adequately address concerns of missing data and provide supporting rationale for unbiased results from missing data

a. No concerns

Submission document: Testing attachment, section 2b6.

MP#4: None

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:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

MP#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: 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 in comparing practices.

MP#1: Analysis shows a pretty modest but acceptable effect of the risk adjustment model.

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 included relatively few Black patients, so an attempt to analyze for effects of race may have failed on the basis of inadequate sample for analysis. Education is included in the adjustment model.

MP#2: Clinician and Group CAHPS 3.0 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 practices. The recommended and provided software applies linear regression models.

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

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. In the submission, it states: 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. I don't this aligns with our emerging standard which would push these numbers up. generally, data support practice-level comparisons and not clinician-level.

    Submission document: Testing attachment, section 2b2.2

    MP#4: Appropriate

    MP#1: Composite and individual patient experience measures were correlated with global rating provider. Inter-correlations were also computed for each composite. 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#3: Both item-level top-box scores and composite scores were correlated with the global rating of provider at the individual and practice site level. Validity of the global rating item was assumed.

    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 practice level

    MP#2:

    a. Results support use at the practice comparison level. Generally, scores do seem to be reasonable measures of quality experience

    Submission document: Testing attachment, section 2b2.3

    MP#4: Results show that all four composites are strongly related to the global provider rating scale.

    MP#3: For adult survey, reasonably high correlations were reported between composite scores and the global rating at the practice level, with only 1 correlation below 0.5. For child survey, 3 correlations were

below 0.5 at the practice level. The correlations between survey items and the global rating were in general lower.

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 practices.

    MP#3: The correlations with the global rating were mostly around 0.5 for all four composite scores at the practice level.

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

    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.

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**

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. 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.

    MP#2:

    a. **Suggest care coordination be sent back for more reliable assessment, perhaps even selecting the best single question?**

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?*

- This survey is already widely used and I don't have concerns that it can't be consistently implemented.
- See above re: Care Coordination. I think this is an important measure that is consistently implemented. It would be interesting to determine if the wide variation in scores between sites was a function of actual difference in policy. For example, some clinics have worked hard on front line staff training, referrals and care coordination. Others simply have not. This would account for the wide variability between practice groups.

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

- The developers have tested at the practice level and not the individual clinician level. For public reporting, this is the analysis that is used. Individual groups may use at the clinician level for improvement initiatives. The care coordination domain doesn't have the reliability of the other domains - this is the revised composite for the 3rd version and the developers do make a comment about this emerging area
- No.

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

- No, The composite scores correlated with the global scores.
- No.

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?

- Missing data (5% at an individual level) did not effect individual ratings. There were some differences by race in the return rate with lower return rates in Asians, African American and Hispanics and in seniors over 75 (mentioned in a referenced study - not included in the analysis by the developer). Results showed a 30-47% difference from CAHPS database average
- No

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?*

- There are suggestions to adjust for health status, mental health status, age and education. It is not required as part of the measure but is available to users.Results showed high correlations between adjusted and unadjusted scores
- Note - one reviewer argues that this is a process rather than an outcome measure. I disagree. Satisfaction with care IS an outcome. And, satisfaction impacts other outcomes.

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.

- Electronic and paper versions are available, mail, phone, e-mail, and web-based modes available and deployed.
- These different modalities have been compared, but those results are not presented in this application.
- Several peer-reviewed publications about response bias and response rates are cited.
- CAHPS and related tools are available free of charge here: https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html.
- Developers note that facilities 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:

- Does the Committee feel that the developer has adequately assessed the feasibility of administering the survey?

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

RATIONALE

- A low rating is assigned because measure developer has not evaluated the burden on providers 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?*

- The surveys are in use and the methods have been updated to reflect changing technology and patient preferences: mail, phone, email, web. The costs for those using it may vary depending on the method with the vendors.
- The feasibility was questioned because information on cost of utilizing approved surveying services was not noted. However, cost will be the same across practices. It may be that it is not feasible for smaller practices, but workable for large groups.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- Developer sites multiple accountability programs where the measure is used
    - MIPS under the Quality Payment Program (CMS)—Payment
    - Comprehensive Primary Care Plus Program (CMMI)—Payment
    - PCMH Recognition Program (NCQA)—Recognition
    - Massachusetts Health Quality Partners—Public Reporting
    - Greater Detroit Area Health Council—Public Reporting
    - Minnesota Department of Health—Public Reporting
    - AHRQ CAHPS Database—Quality Improvement and Benchmarking
- Developer also cites known use for internal quality improvement purposes

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

- The instrument was developed in the late 1990s and has since been refined with substantial input from entities being measured.
- In response to feedback from pediatric experts, the Child CG-CAHPS was adjusted working with the American Academy of Pediatrics to add 11 items that address development and prevention.
- Field testing with Massachusetts health Quality Partners lead to scale adjustments
- Users offered feedback on 3.0 version leaded to multiple changes:
    - The survey reference time period.
    - Access composite measure.
    - Communication composite measure.
    - Office staff composite measure.
    - Care coordination composite measure.
    - Measures in the PCMH item set.
- Based on the stakeholder and user feedback obtained through public comment, technical expert review, and further testing, the following key changes were made in the release of the CG-CAHPS 3.0 version:
    - One instrument, in contrast to the three instruments available for the 2.0 version.
    - Use of a 6-month reference time period rather than a 12-month reference period.
    - New and modified composite measures:
    - New composite measure for "Care Coordination."
    - Modified composite measure for "Access."
    - Modified composite measure for "Communication."
    - A modified Patient-Centered Medical Home Item Set.
    - Shift of development and prevention items from the core Child Survey to the Patient-Centered Medical Home Item Set.
    - Overall reduced length.

***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

- Mixed results about improvement were presented.

- The last three years of CAHPS data coalesced in a chartbook showed mostly slight declines in performance.
- Data from the Washington Health Alliance was only summarized by noted "room to improve" was evident.
- Data from Minnesota and Massachusetts indicated "slightly" increasing performance with time.
- The developers present no explanation for these results.

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

Unexpected findings (positive or negative) during implementation

Potential harms

- None presented

Additional Feedback:

- None

***Questions for the Committee**:*

- Are the trending data sufficient to demonstrate the quality metric is yielding improvement?
- Should the developer have interpreted and explained the mixed results in more detail?

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

RATIONALE

- A low rating is assigned because improvement in the measure has not been consistently demonstrated, and because the discourse does not clarify why such inconsistency (especially the slight decreases) are evident**.**
- 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 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?*

- The surveys is in use for public reporting by CMS and private organizations. NCQA uses it for Patient Centered Medical Home recognition. Medical groups use it for their own internal quality improvement. Resources are available from AHRQ for groups for improvement and case studies were provided
- This information COULD be used to target improvement activities. It would be interesting if the scores were related to targeted improvement efforts.

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.

- Developer does not note unexpected findings. An unintended consequence may be how it is used at an individual clinician level - improvement balanced with accountability depending on the reliability at the clinician level/sample size
- I say moderate, although reviewers said low. It is used by some and could be used to pinpoint areas for improvement.

Criterion 5: Related and Competing Measures

Related or competing measures

- Developer cites the following: "CAHPS for Accountable Care Organizations (ACO) and CAHPS for MIPS -- both use the core CG-CAHPS items. These versions have been recommended by the NQF Measure Applications Partnership (MAP) Clinician Workgroup for use by CMS in the Medicare Shared Savings Program and the Merit-Based Incentive Payment System, respectively. (See: http://www.qualityforum.org/Publications/2018/03/2018_MAP_Clinicians_-_Final_Report.aspx"
- 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?*

- There are CAHPS for ACO's and MIPS - the developer has been trying to align the core items
- Several related but not competing measures. Harmonization not required.

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

CG_evidence_April_2019.docx,CG_evidence_Submitted_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*)**:** 0005

Measure Title: CAHPS Clinician and Group Survey Version 3.0

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:

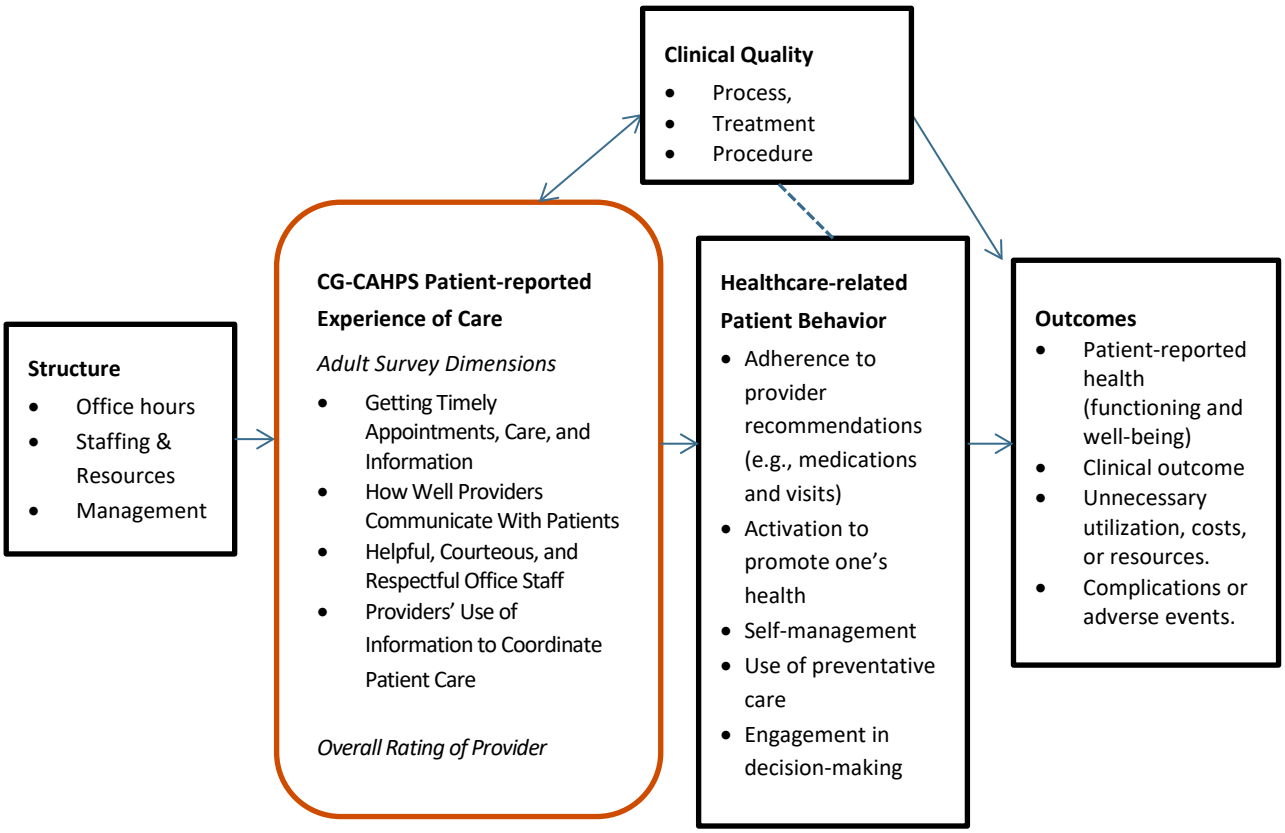
☐ Structure:

☐ 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 CG-CAHPS and patient behaviors and clinical outcomes. Patient care experiences are

influenced by structural features and processes 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 measures, developed to complement more technical quality measures, are measures for which the patients are the best or only source of information and/or perspective, such as the degree to which care felt patient-centered (Anhang Price et al., 2014). Several studies provide evidence that patients value the CAHPS measures and find them meaningful. For example, Safran et al (2001) found that patients who reported the poorest-quality relationships with their physicians were three times more likely to voluntarily leave the physicians' practice than patients with the highest-quality relationships.

Collins et al (2017) found that a patient's "most important CAHPS domain" varied across subgroups; racial and ethnic patient subgroups differentially valued various aspects of the care experience. To efficiently reduce disparities and improve quality, Collins et al recommend tailoring quality improvement programs to the factors most important to the racial, ethnic, and language mix of the patient population of interest. Quigley et al (2014) found that the importance of provider communication varied significantly by practice specialty type, yet respectful treatment was consistently import across all specialties.

Patients also 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).

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.

Quigley, D.D, Elliott, M.N., Burkhart, Q, Farley, D.O., Skootsky, S.A., & Hays, R.D. (2014). Specialities differ in which aspects of doctor communication predict overall physician ratings. *Journal of General Internal Medicine,* 29, 447-454.

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 measures key components of patient experience, 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. We reviewed the literature on the determinants of patient care experiences measured by CAHPS and their associations with other indicators of health care quality. CAHPS is also an actionable measure that helps clinicians and health plans target interventions that will improve the quality and patient-centeredness of care.

REVIEW OF THE EVIDENCE

Prior research has identified several features of healthcare delivery structure, including clinic accessibility, patient flow, and management, 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 of INFLUENTIAL RELATIONSHIPS

Health Care Process and Quality Improvement Influence on Patient Experience

Providers routinely use patient-reported measures such as CAHPS to guide quality improvement (QI) activities to improve their patients' experience with care (Friedberg et al., 2011; Davies, Shaller et al., 2013; Quigley et al, 2015).

For example, to improve scores for "how often the office staff were as helpful as you thought they should be", Dean Clinic, a large integrated health care delivery system in Wisconsin, further surveyed their patients to ask them what "helpfulness" means to a them and to ask how office staff could be more helpful. From this feedback, the Clinic learned about ways that the office staff can be more welcoming, friendly, and appreciative of patients. With input from both staff and management, Dean Clinic developed action plans to improve patient experience. The service department shadowed staff and provided feedback. To improve consistency in service across all sites, the Clinic developed an orientation for all new employees on customer service expectations. They also offered ongoing training in the form of service workshops, videos, and Webinars, as well as targeted interventions for the lowest scoring offices (AHRQ, 2013). Several quality improvement initiatives to improve provider communication and engagement include shadowing, coaching, and training

(AHRQ, 2013; Hardee & Kasper, 2008; Quigley, Palimaru et al., 2017). Lastly, Friedberg et al (2011) found that physician groups that aim to improve access, communication with patients, and customer service do so by addressing office workflow, providing additional training for nonclinical staff, and adopting or enhancing an electronic health record.

Using data from the CAHPS survey and a newly installed electronic health record system, in 2015, leaders of Northeast Valley Health Corporation (Los Angeles County) pinpointed interventions including reallocating staff resources and not scheduling well-child visits first thing in the morning. System leaders found that when the first appointments of the day ran long, there was a cascading effect on the rest of the day's schedule. By 2015, NEVHC instituted changes systemwide. By 2017, total average cycle time was reduced from 82 minutes to 65 minutes; the proportion of patients with a cycle time under 60 minutes rose from 34 percent to 48 percent; and the proportion of patients seen within 15 minutes of appointment time rose from 38 percent to 47 percent. Case study found at https://www.ahrq.gov/news/newsroom/case-studies/201718.html

More examples of interventions to improve patient experience with primary and specialty care as measured by the CG-CAHPS Survey can be found in the CAHPS Improvement Guide available at https://cahps.ahrq.gov/quality-improvement/improvement-guide/improvement-guide.html. The Guide also includes information on analyzing survey results and identifying root causes of performance problems.

The Centers for Medicare and Medicaid Services Merit-based Incentive Payment System (MIPS) Quality Payment Program measures Medicare Part B providers in four performance categories to derive a score that could affect a provider's Medicare reimbursement positively or negatively starting at 4% in 2019 (based on 2017 performance). MIPS has a performance category called “Improvement Activities” that includes an inventory of activities that assess how a physician group can improve care processes, enhance patient engagement in care, and increase access to care. Several of these activities are geared to improve patient experience of care as measured by CG-CAHPS. A list of MIPS activities can be found at https://qpp.cms.gov/mips/improvement-activities.

Structure

Physician clinic hours of operation and availability for appointments have been found to predict patient experience in several studies. In a study with survey data from 61,839 patients of 1729 primary care physicians in California, system-level factors, such as belonging to a larger medical group and the physician’s zip code-based Primary Care Services Areas, explained between 28% to 48% of variation in patient care experience, with the highest proportion explained for the access to care composite (Rodriguez et al. 2009). Improving the infrastructure supporting certain aspects of care may have broad effects because system changes can influence multiple outcomes (Cleary, 2016).

“Expanded practice access” is a highly weighted CMS MIPS improvement activity. To improve the patient experience in access to care, practices may consider providing 24/7 access to clinicians, groups, or care teams for advice about urgent and emergent care (e.g., eligible clinician and care team access to medical record, cross-coverage with access to medical record, or protocol-driven nurse line with access to medical record). Access expansion can include one of the following: 1) expanded hours in evenings and weekends with access to the patient medical record (e.g., coordinate with small practices to provide alternate hour office visits and urgent care); 2) use of alternatives to increase access to care team by clinicians and groups, such as e-visits, phone visits, group visits, home visits and alternate locations (e.g., senior centers and assisted living centers); and/or 3) provision of same-day or next-day access to a consistent clinician, group or care team when needed for urgent care or transition management. Another example is implementation of “open access” scheduling, in which some physician time is always reserved for same-day appointments, to improve patient access to care (Murray & Tantau, 1998).

Health-related Patient Behavior and Disease Management

One dimension of the CG-CAHPS measure captures the 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). Sequist and colleagues (2008) found that measures of patient experience, including doctor-patient communication, clinical team interactions, and health promotion support, were positively associated with some prevention and disease management clinical process measures in clinical practices and among individual clinicians.

Outcomes

Out of 40 evidence papers with outcome measures, Doyle's (2013) meta- analysis found 29 studies with 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 complexity 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; Xu et al. 2014).

Utilization

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). 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

Agency for Healthcare Research and Quality (AHRQ). How Two Provider Groups Are Using the CAHPS® Clinician & Group Survey for Quality Improvement. Brief available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/quality-improvement/reports-and-case-studies/cgcahps-webcast-brief-2014.pdf.

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.

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.

Davies E, Shaller D, Edgman-Levitan S, Safran DG, Oftedahl G, Sakowski J, Cleary PD. (2008) Evaluating the use of a modified CAHPS survey to support improvements in patient-centered care: lessons from a quality improvement collaborative. *Health Expect.* Jun;11(2):160-76.

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., 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.

Friedberg MV, SteelFisher GK, Karp M, and Schneider EC. (2011) Physician groups' use of data from patient experience surveys. J Gen Intern Med. 26(5): 498-504.

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.

Hardee JT, Kasper IK. (2008) A Clinical Communication Strategy to Enhance Effectiveness and CAHPS Scores: The ALERT Model. *Perm J.* Summer;12(3):70-4.

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.

Murray M and Tantau C. (1998) Must patients wait? *Jt Comm J Qual Improv* 24(8): 423-5.

Quigley DD, Mendel PJ, Predmore ZS, Chen AY, Hays RD. (2015) Use of CAHPS™ patient experience survey data as part of a patient-centered medical home quality improvement initiative. *J Healthc Leadersh.* Jul 7;7:41-54.

Quigley, D.D., Palimuru A.I., Chen A.Y., & Hays R.D. (2017). Implementation of practice transformation: Patient experience according to practice leaders. *Quality Management in Health Care.* 26 (3), 140-151.

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.

Sequist, T. D., E. C. Schneider, et al. (2008). Quality monitoring of physicians: linking patients' experiences of care to clinical quality and outcomes. *Journal of general internal medicine*. 23(11): 1784-1790.

Xu, X., Buta, E., Anhang Price, R., Elliott. N. N., Hays, R. D., & Cleary, P. D. (2014). Methodological considerations when studying the association between patient-reported care experiences and mortality. *Health Services Research.* 50(4), 1146-61.

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

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

☐ 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 evidence.

1b. Performance Gap

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

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

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

CG-CAHPS assesses aspects of health care delivery that are important to patients and for which patients are the best or only source of information (Solomon et al, 2005). Furthermore, the CG-CAHPS survey focuses on patient-centered care (Cleary, 2016), one of the six central aims identified by the Institute of Medicine for improving the health care system (IOM, 2001). A focus on the patient experience has the potential to enhance clinical outcomes, improve patient safety, and reduce unnecessary medical services. Moreover, assessing patient experience through surveys that include data on the demographic characteristics of respondents, such as race and ethnicity, can help identify the extent to which positive experiences are distributed equitably across patients (Haviland et al., 2003).

Citations:

Cleary, PD. Evolving Concepts of Patient-Centered Care and the Assessment of Patient Care Experiences: Optimism and Opposition. Journal of Health Politics, Policy and Law. Vol. 41, No. 4. August 2016.

Haviland, M. et al, (2003). Do Health Care Ratings Differ by Race or Ethnicity? Joint Commission Journal on Quality and Safety. 29(3), pp. 134-145.

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.

Solomon, L., Hays, R. D., Zaslavsky, A., & Cleary, P. D. (2005). Psychometric properties of the Group-Level Consumer Assessment of Health Plans Study (CAHPS?) instrument. Medical Care, 43, 53-60.

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 CG-CAHPS 3.0 was submitted by users to the CG-CAHPS database for surveys administered January 2016 – March 2017. The data includes 636 practice sites that had at least 10 completed surveys and has 112,398 respondents. For the Access composite, the mean top box score/proportion at the practice site level = 0.68 (or 68%); standard deviation (std) = 0.10, minimum (min) = 0.38, median = 0.67, and max = 1.0 (or 100%). For the Communication composite, the mean top box score = 0.87; std = 0.05, min = 0.59, median = 0.88, and max = 1.0 (or 100%). For the Care Coordination composite, the mean top box score = 0.73; std = 0.07, min = 0.35, median = 0.74, and max = 0.91 (or 91%). For the Office Staff composite, the mean top box score = 0.81; std = 0.08, min = 0.51, median = 0.82, and max = 1.0 (or 100%). For the one-item overall rating of provider, the mean top box score = 0.82; std = 0.08, min = 0.29, median = 0.83, and max = 1.0 (or 100%). Scores by decile are reported in the attached Excel file named CG_ CAHPS_Main_Supplementary_Tables_Apr_2019.xlsx (worksheet tab titled “Table 1b.2 Adult”).

The data for the Child CG-CAHPS 3.0 includes data submitted by users to the CG-CAHPS database for surveys administered January 2016 – March 2017. The data includes 77 practice sites that had a least 10 completed surveys and has 12,474 respondents. For the Access composite, the mean top box score/proportion at the

practice site level = 0.75 (or 75%); standard deviation (std) = 0.10, minimum (min) = 0.45, median = 0.76, and max = 0.93 (or 93%). For the Communication composite, the mean top box score = 0.90; std = 0.05, min = 0.68, median = 0.90, and max = 0.99 (or 99%). For the Care Coordination composite, the mean top box score = 0.81; std = 0.07, min = 0.57, median = 0.81, and max = 0.95 (or 95%). For the Office Staff composite, the mean top box score = 0.76; std = 0.10, min = 0.48, median = 0.76, and max = 0.97 (or 97%). For the one-item overall rating of provider, the mean top box score = 0.83; std = 0.09, min = 0.36, median = 0.84, and max = 0.98 (or 98%). Scores by decile are reported in the attached Excel file named CG_ CAHPS_Main_Supplementary_Tables_Apr_2019.xlsx (worksheet tab titled "Table 1b.2 Child").

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.*

For the adult survey (N=112,398), differences in the top box scores by gender are small. Male respondents gave more positive ratings for each of the composites (ranging from 1-2 percentage points higher) but were no different on the overall rating of provider.

For the child survey (N=12,474 respondents), differences in the top box scores by the gender of the adult respondents are very small. There was no difference for the Access composite, Staff service composite, or the global rating of provider. Female respondents scored one percentage point higher for communication and care coordination composites.

For the adult survey, access composite, the top box mean scores by age category were 18-24 = 63%, 25-34 = 65%, 35-44 = 64%, 45-54 = 65%, 55-64 = 67%, 65-74 = 70%, and 75+ = 69%. For the communication composite, the mean scores by age category were 18-24 = 84%, 25-34 = 85%, 35-44 = 87%, 45-54 = 88%, 55-64 = 89%, 65-74 = 70%, and 75+ = 87%. For the care coordination composite, the mean scores by age category were 18-24 = 70%, 25-34 = 72%, 35-44 = 74%, 45-54 = 75%, 55-64 = 76%, 65-74 = 76%, and 75+ = 71%. For the staff service composite, the mean scores by age category were 18-24 = 74%, 25-34 = 76%, 35-44 = 77%, 45-54 = 79%, 55-64 = 82%, 65-74 = 84%, and 75+ = 85%. For the overall rating, the top box mean scores by age category were 18-24 = 72%, 25-34 = 73%, 35-44 = 76%, 45-54 = 80%, 55-64 = 82%, 65-74 = 86%, and 75+ = 85%. Older patients generally reported more positive patient experiences. The largest difference was found for Global Rating of Provider and Helpful, Courteous, and Respectful Office Staff. The associations of age with more positive experiences with care has been observed consistently in different CAHPS surveys. For example, O'Malley et al (2005) found that younger age patients (18-24) scored significantly lower (less positive) than patients 25-34 and patients in the eight age categories above 34 all scored higher than those 25-34 on the CAHPS Hospital Survey. We did not compare age differences for the adult respondents (most often parents) of the child survey.

Race/ethnicity was not compared due to lack of variability in our dataset; 88% of the adult survey population was white. However, researchers have found that Asians fare worse than Whites in terms of rating their patient experience about care (Weech-Maldonado et al, 2003; Chung et al, 2018).

Response scores by gender and age category are also provided in worksheets "Table 1b.4a" and "Table 1b.4b." in the attached Excel file named CG_CAHPS_Main_Supplementary_Tables_Apr_2019.xlsx.

Citations:

Chung S, Mujal G, Liang L, Palaniappan LP, Frosch DL. Racial/ethnic differences in reporting versus rating of healthcare experiences. Medicine (Baltimore). 2018 Dec;97(50):e13604.

O´Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L and Cleary PD (2005) Case-Mix Adjustment of the CAHPS® Hospital Survey. Health Services Research. Dec. Volume 40, Issue 6p2, pages 2162–2181.

Weech-Maldonado R, Morales LS, Elliott M, Spritzer K, Marshall G, and Hays RD. Race/Ethnicity, Language, and Patients´ Assessments of Care in Medicaid Managed Care. Health Services Research. 2003. 38(3): p. 789-808.

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/cg/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:** CG_CAHPS_Main_Supplementary_Tables_Apr_2019-636915274790156222.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-eng-cg30-2351a.docx

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Patient

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

Yes

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.

CG-CAHPS Survey Version 1.0 was endorsed by NQF in July 2007. CG-CAHPS Survey Version 2.0 received maintenance endorsement in Jan 2015. Version 3.0 was implemented in mid-July 2015. Changes to the measure specifications in version 3.0 of the Adult survey include:

1. Use of a 6-month reference “look-back” time-period rather than a 12-month reference period.

Reason: This change made the survey consistent with the versions implemented by CMS, including the CAHPS Survey for ACOs and the CAHPS Survey for PQRS. An analysis led by Yale randomized patients to a 12-month or 6-month survey version yielded similar CAHPS scores at the practice site level.

2. New composite measure "Providers’ Use of Information to Coordinate Patient Care" consisting of three items:

a) Follow up on test results (stand-alone time already in CG-CAHPS core survey)

b) Knows important information about medical history (previously in Communication composite)

c) Provider talked about all prescription medicines being taken (from the PCMH Item Set)

Reason: Care coordination is an important aspect of patient experience that is commonly assessed by CAHPS surveys. The Consortium’s goal was to develop a care coordination composite measure that could be standardized across CAHPS surveys. The CAHPS Medicare Survey includes a 10-item measure but a shorter measure is desirable. Since two of the items were already part of the core survey, this new composite required the addition of only one item to the core survey.

3. The "Access" composite measure was reduced to 3 items from 5 items.

Reason: The item “Got answer to medical question after office hours” and “Wait time for appointment to start” were moved from this core survey to a supplemental item set. Because very few patients seek answers after office hours, this item does not add sufficiently to a core measure of access. While the item about wait time for appointments remains an important assessment of access, the item-scale correlation for this item was lower than the correlations for the other items included in the recommended composite measure. The items retained in the composite measure are important to patients and stakeholders, have good reliability, and include multiple aspects of access. In addition, items that used the wording “phoned this provider’s office” were revised to “contacted this office” to conform to the preference of stakeholders for items that encompass the multiple ways in which patients may reach out to the provider’s office. Cognitive testing of this revision was conducted to ensure that respondents correctly understand the intent of this revised wording.

4. The "Communication" composite measure was reduced to 4 items from 6 items.

Reason: The 4-item composite is consistent with the communication measure in the CAHPS Health Plan Survey. The ”Gives easy to understand instructions” was dropped because it was similar to ”Explains things in a way that is easy to understand.” The item, ”Knows important information about medical history,” was moved to the new Care Coordination composite.

These changes reduce the length of the core Adult Survey from 34 items to 31 items.

Additional changes to the CHILD SURVEY: Several questions about prevention and development were moved from the core survey into the Patient-Centered Medical Home supplemental item set.

Reason: These items are recommended for use by primary care providers and may not be applicable to the specialty care providers whose young patients are eligible for the CG-CAHPS Child survey. Therefore, they were removed from the core Child Survey and added to the PCMH item set. The Child Survey Version 3.0 is 39 items reduced from 55.

Comparability of Results for the 2.0 and 3.0 Versions of the CG-CAHPS Adult are available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/Comparison_of_CG-CAHPS_v3_and_v2.pdf.

The CAHPS® Clinician & Group Survey: Overview of the Questionnaires (overview-of-questionnaries-cg30-2350.pdf) is available through Clinician & Group Survey 3.0 and Instructions ZIP file at https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

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

The CG-CAHPS Survey item and composites are often reported 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 the Overall Rating of Provider is the number of respondents who answered 9 or 10 for the item, with 10 indicating "Best provider possible".

For more information on the calculation of reporting measures, see

"Preparing Data from CAHPS® Surveys for Analysis" (AHRQ, 2017) accessible at

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

and the CAHPS Analysis Instructions accessible at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2015-instructions-for-analyzing-data.pdf (updated June 2017).

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

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

For each individual item, the top box numerator is the number of respondents who answered "Always" (the most positive response) for the item.

There are two basic steps to calculating a composite score for a practice site:

1. Calculate the proportion of patient responses in the top box or most positive response category 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.

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

Example: Items in "Helpful, Courteous, and Respectful Office Staff" (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 "Clerks and receptionists at this provider's office were as helpful as you thought they should be." = Proportion of respondents who answered "Always" = 80%
- Item #2 "Clerks and receptionists at the provider's office treat you with courtesy and respect." = 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 "Helpful, Courteous, and Respectful Office Staff" = (Item1 + Item2) / 2 = (80% + 90%) / 2 = 85%

More detail can be found in https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/preparing-data-for-analysis.pdf

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

The measure's denominator is the number of survey respondents. The target populations for the surveys are patients who have had at least one visit to the selected provider in the target 6-month time frame. This time frame is also known as the look back period. The sampling frame is a person-level list and not a visit-level list.

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 item in a composite and the provider rating item, the top box denominator is the number of respondents who answered the item per aggregate-level entity (e.g., a physician or practice site). For a composite score, the denominator is the number of respondents who answered 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).

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

Among eligible respondents, for a given item, respondents with a missing response is excluded. Among eligible respondents, for a composite measures, respondents who did not answer at least one item in the composite are excluded from the composite measure's denominator.

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.)*

Surveys will not be obtained from the following:

1) Deceased patients. For example, the individual has died between the visit(s) and receipt of the questionnaire.
2) If the potential respondent has a language barrier, the instrument is not available in the respondent's preferred language, and no one was available to translate the questions for the respondent.

The following should be excluded from the sample where date was obtained but is not useable due to ineligibility:

1) The respondent reports he or she has not visited the sampled entity (e.g., a physician or practice site). This might be indicated by a "no" response to Question 1 (e.g., "Our records show that you got care from the provider named below in the last 6 months. Is that right?").
2) Individuals from a household that has already been sampled.
3) When a proxy was used (someone answered the questions on behalf of the target respondent) and the users do not intent to add a proxy indicator with case-mix adjustment

Survey respondents who did not answer at least one item of a measure are excluded from a measure's denominator.

Survey code specifications --- including how to code an appropriately skipped item, multiple marks or blank items --- can be found in this document: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/preparing-data-for-analysis.pdf.

More instruction is available as downloadable files from the CAHPS analysis web page https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html. Files are:

- SAS programs with test modules (ZIP, 170 KB)—updated June 2017
- Instructions for Analyzing Data from CAHPS Surveys (PDF, 927 KB)

The CAHPS Analysis Program computes scores for practices, sponsors and vendors. The goal of the CAHPS Analysis Program is to provide the user with a flexible way to analyze CAHPS survey data to make standardized comparisons of performance.

The CAHPS macro calculates scores at the unit level (e.g. practice site) 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, global rating of mental health, and global rating of general health (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.)*

CG-CAHPS users that have collected data for different clinical practices may decide to analyze the data separately or together. If practices are to be analyzed together, no changes to the CAHPS Analysis Program are necessary. If a team decides to analyze the practices separately and the data file contains more than one group, it is important to set up selection criteria in the CAHPS Analysis Program or split the data set.

Users can separate case-mix adjustments on two different subgroups 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 urban and rural locations.

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.) Tob-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: Patients that had at least one visit during the past 6-months to a selected provider
2) Exclusions = Patients who did not answer at least one item of the composite measures or rating item.

3) Screener items. Example: Patients who answered "No" to the first item indicating that they did not receive care from the provider entity in the last 6 months

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.

Case-mix Adjusted Top box or Mean Scores:

The steps for user-defined calculations of risk-adjusted scores can be found in 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 (updated June 2017)

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.

Defining the Sample Frame: Eligibility Guidelines

The sample is drawn from a list of individuals (adults age 18 and older, or children 17 and younger) who have received care from a given provider, practice site, or medical group during the specified time interval (see below). The list is called a sample frame.

The source of sample information will vary by survey sponsor. The data to identify individual patients may be found only in the records of medical practices or health systems.

Guidelines for determining sample frame:

- Include only patients who have had at least one visit to the selected practice in the last 6 months. This time frame is also known as the look back period.
- The sampling frame is a person-level list and not a visit-level list. Therefore, patients should appear only once in the sampling frame regardless of how many visits they have had in the look back period.
- The adult questionnaires include all adults 18 years or older.
- The child questionnaire includes all children 17 years or younger.
- Draw the sample irrespective of reason for visit and duration of patient-provider relationship, so that the full range of patients is represented.
- Include all patients who meet the sampling criteria even if they are no longer receiving care from the practice, site/clinic or provider.
- To identify the sampling frame, use the anticipated start date of data collection to determine the reference period. For example, if your anticipated start date is September 1, 2017, include all those who have had at least one visit since March 1, 2017.
- Allow the sample frame to include multiple individuals from the same household, but the sample you draw should not have more than one person (adult or child) per household. In other words, the sample that is selected for data collection should be de-duplicated to ensure that only one person per household receives a survey.
- All CAHPS survey items have been designed for the general population. Appropriate screening items are included for items targeted to assess a specific experience. In order to ensure that results are comparable to those produced by other sponsors and vendors, targeted sampling, such as selecting only patients with particular conditions or experiences, is not recommended. Targeted sampling should only be used to supplement the general population sample, if desired.
- In order to administer the survey, the name of the provider must be available, even if surveying at the site/clinic or practice level. If the sampling frame does not accurately identify the provider that the patient

saw, select a larger sample to account for errors in connecting health care received to a specific provider. For example, errors can occur if administrative billing data are used for the sampling frame and visits with physician assistants or nurse practitioners are billed under the supervisory physician.

Sample size requirements: Recommended Number of Completes

The CAHPS Clinician & Group Surveys can be used to assess care at the individual provider, practice site/clinic, or medical group level. A practice site/clinic is based on a single geographic location. A medical group may contain multiple practice sites/clinics and is defined by a specific list of providers.

The recommendations regarding the number of completed questionnaires per provider, practice site, and group apply to the survey with core items only.

Individual providers: For applications of the survey intended to report or assess performance for individual providers, the Consortium recommends at least 50 completed surveys per provider to achieve .70 provider-level reliability.

Practice site or clinic: The recommended number of completed surveys is based on the number of providers at the site. The CAHPS Consortium's recommendations for sample sizes at the practice site level are: 50 completed surveys for 1 provider at the site, 100 surveys for 2 providers at the site, 150 for 3 providers, 175 for 4 to 9 providers, 200 for 10-13 providers, and 250 for 14-19 providers at the site. If there are 20 providers or more, it is recommended to have at least 300 completed surveys. These recommendations are set to achieve between .70 and .80 site-level reliability for all composite measures. Sample size requirements increase with the number of providers practicing at the site.

Medical group: 300 completed surveys. For applications of the survey intended to report or assess performance for a larger entity, such as a multisite medical group, with no interest in assessing individual physicians, we refer to analyses conducted for the CAHPS Group Practice Surveys, which preceded the Clinician & Group Surveys, and were confirmed with data from the CAHPS Database. Consequently, the Consortium recommends a minimum of 300 completed surveys per medical group to achieve .70 group-level reliability.

Recommended sample sizes are developed based on extensive assessment of the number needed to produce target levels of reliability and the ability to distinguish between providers. Sample size recommendations are based on studies of the number of completed questionnaires necessary to achieve adequate provider-level, practice-level, or group-level reliability for a measure. That is, how many completed questionnaires are needed to reliably distinguish among different units of measurement (such as individual providers or practices)? To answer this question, the CAHPS team examined data from multiple field trials as well as data from the CAHPS Database. The unit-level reliability coefficient indicates the extent to which the patients within that unit (e.g., the patients of an individual provider, practice, or group) agree with one another in terms of their reported experiences. This coefficient can take any value from 0.0 to 1.0, where 1.0 signifies a measure for which every patient (e.g., all patients surveyed for a given practice) reports an identical experience. High levels of reliability are ideal, but achieving higher levels requires more completed questionnaires. To balance the goal of reliability with the need for a feasible sample size, the CAHPS team adopted the widely accepted coefficient of 0.70 as the threshold. A reliability level of at least 0.70 is strongly recommended for "high stakes" purposes such as public reporting or payment incentives, given the larger errors around estimated scores below this threshold.

Proxy Respondents

CAHPS does allow for proxy respondents for mail and web-based mode. 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 information can be found in the "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" document (June 2017) available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.pdf

and "Preparing a Questionnaire Using the CAHPS Clinician & Group Survey"

Document No. 2032. Updated 9/1/2015. Accessible at

https://www.ahrq.gov/cahps/surveys-guidance/item-sets/cg/index.html

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.

Preparing Sample Files for Data Collection

Once the users have selected their sample, the vendor assigns a unique identification (ID) number to each sampled person. This unique ID number should not be based on an existing identifier such as a Social Security number or a patient ID number. This number will be used only to track the respondents during data collection.

Some sample frames may not include complete and accurate contact information, requiring the combination of information from two (or more) sources – such as administrative records and contact records from a health system, medical group or clinician office. When information from two sources differs, sponsors and their survey vendors should consult with each other to decide which sources of information are most accurate and should be used. This may be a complex, multistep process that requires time and rigorous quality control. In addition, because the sponsor may be responsible for some elements of this process and the vendor for others, it is important to carefully coordinate this process.

The pieces of information that are most critical to the success of data collection are accurate and complete patient [parent/guardian] and provider names and contact information appropriate for the mode of administration (i.e., addresses for mail surveys, telephone number for telephone administration, and e-mail addresses for web-based administration). In cases of incomplete address information or if there is reason to believe that this information may be inaccurate, sponsors and/or vendors may be able to use tracing/tracking methods such as Lexis-Nexis, CD-ROM directories, Internet database services, GPS-based address verifications, and directories.

Each survey sponsor will need to choose the data collection mode that maximizes the response rate at an acceptable cost. Based on field test results, the Consortium recommends the following modes:

- Mail only.
- Telephone only.
- Mixed mode (mail and telephone, e-mail with link to a web-based survey and mail, or e-mail and telephone).

Based on user experience, we anticipate that survey sponsors who employ one of these modes will achieve response rates that approximate 30 to 40 percent or higher.

Results from the field tests, as well as the experiences of organizations that have fielded similar surveys, indicate that the mail with telephone follow-up method is most effective: results from survey research literature indicate that follow-up by telephone often adds 10 to 15 percentage points to the response rate.

The Consortium is aware that many organizations already conduct patient surveys to gather information for internal quality improvement purposes. These organizations use several different modes of survey

administration, most of which include mail or telephone-based administration. The Consortium's support of the use of multiple modes of survey administration is intended to minimize disruption to current survey processes. Thus, organizations that conduct mail surveys can continue using mail, those that conduct telephone surveys can continue using telephone, and likewise for other modes -- with the caveat that, if data are not collected using the recommended modes, the results will not be comparable to data gathered with those modes.

Tracking Returned Questionnaires

Most vendors have established methods for tracking the sample. The Consortium suggests setting up a system to track the returned surveys by the unique ID number that is assigned to each respondent in the sample. This ID number should be placed on every questionnaire that is mailed and/or on the call record of each telephone case.

To maintain respondent confidentiality, the tracking system should not contain any of the survey responses. The survey responses should be entered in a separate data file linked to the sample file by the unique ID number. (This system will generate the weekly progress reports that sponsors, and vendors should review closely.)

Each respondent in the tracking system should be assigned a survey result code that indicates whether the respondent completed and returned the questionnaire, completed the telephone interview, was ineligible to participate in the study, could not be located, is deceased, or refused to respond. The tracking system should also include the date the survey was returned or the telephone interview completed. The interim result code reflects the status of the case during the different rounds of data collection, and the final result code reflects the status at the end of data collection. These result codes are used to calculate response rates.

Data collection information is provided to users as part of the "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" document (June 2017) available at

https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.pdf

A review of the strategies for CAHPS survey administration can be found in "CAHPS Survey Administration: What We Know and Potential Research Questions" (Oct 2017), accessible at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/about-cahps/research/survey-administration-literature-review.pdf.

Citation:

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.

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 Clinician & Group, Adult Survey 3.0

CAHPS Clinician & Group, Child Survey 3.0

Available in English and Spanish at https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html

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)*

Clinician : Group/Practice

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_CG_NQF_submitted_Jan7_2019.docx,CAHPS_CG_NQF_submitted_Apr19_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*)**:** 0005
Measure Title: CAHPS Clinician & Group Survey Version 3.0
Date of Submission: 4/19/19

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 Clinician and Group Database	☒ other: CAHPS Clinician and Group Database

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*).

This dataset used for testing is CAHPS Clinician & Group Survey Version 3.0 data voluntarily submitted by users that administered the survey between January 1, 2016 and March 31, 2017.

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? January 1, 2016 – March 31, 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*)

CG-CAHPS is specified for both individual clinician and group/practice. However, we only have testing for the group or "practice site" because the clinician-level identification in the CAHPS database is not reliable. The site-level measured entity is referred to as a "practice site". The practice site is an outpatient facility in a specific location. It can be considered as a medical office. Each practice site located in a building containing multiple medical offices is considered a separate site, as providers in a single practice site are likely to share administrative and clinical support staff. Practice site level survey results are calculated across the respondents within a specific site.

The data used for descriptive analysis includes Adult CG-CAHPS 3.0 surveys administered January 2016 through March 2017. The data includes 656 practice sites and 137,416 respondents. For testing and performance scores, practice sites with less than 10 surveys were excluded as were sites submitted from health care systems that did not wish to identify specific practice sites. After the restriction (N=635 sites), the average number of respondents per site is 177 and ranges from 11 individuals per site to 1,563 individuals per site. The total number of unique provider IDs in the adult testing sample was 5,304 resulting in an average of 10 providers per site. However, these numbers are not exact and should be interpreted with caution as the coding of physician IDs may not be consistent across survey vendors.

The data used for descriptive analysis includes Child CG-CAHPS Version 3.0 surveys submitted by users to the CG-CAHPS database that were administered January 2016 through February 2017. The data includes 81 practice sites and 12,488 respondents. For testing and performance scores, practice sites with less than 10 surveys were excluded. After the restriction (N=77), the average number of respondents per site is 162 and ranged from 10 individuals per site to 3,028 individuals per site. The child survey is completed by the child's parent, relative, or legal guardian. The total number of unique provider IDs in the child sample was 547 resulting in an average of 8 providers per site. However, these numbers should again be interpreted with caution as the coding of physician IDs may not be consistent across survey vendors.

Table 1.5a Distribution of States in CG-CAHPS Adult Version 3.0 Sample

Count of Complete Records & Practice Sites within State for CG Adult		
Practice State	**Total Complete Records Within State**	**Total Practices Within State**
CALIFORNIA	29,355	145
FLORIDA	24,850	1
LOUISIANA	30	1
MAINE	5,374	39
MASSACHUSETTS	19,969	100
MICHIGAN	21,819	198
MINNESOTA	24,699	99
NORTH DAKOTA	2,407	7
OHIO	7,450	62
WISCONSIN	1,463	4
Total	***137,416***	***656***

Table 1.5b Distribution of States in CG-CAHPS Child Version 3.0 Sample

Count of Complete Records & Practice Sites within State for CG Child		
Practice State	**Total Complete Records Within State**	**Total Practices Within State**
CALIFORNIA	1,464	11
FLORIDA	4,606	5
MAINE	472	4
MASSACHUSETTS	1,124	10
MICHIGAN	3,667	39
MINNESOTA	1,155	12
Total	***12,488***	***81***

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*)

In each practice site, surveys are completed by patients who had at least one visit during the past 6 months from the user's "survey start date." For example, if the start date was January 1, 2016, the survey sample included responses from surveyed patients who have had at least one visit since July 1, 2015.

Although multiple individuals in a single clinician group may be in the sampling frame, the final sample must contain only one respondent per household. Where a duplicate household is sampled, it is discarded and replaced by another random draw from the frame. The fielding guidelines provide additional advice on drawing representative samples for multiple products and simultaneous sampling of adult and child enrollees. Fielding guidelines are available at: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.pdf.

Tables 1.6a – 1.6f show descriptive characteristics of the individuals surveyed by the Adult Version 3.0 and the Child Version 3.0. Practice sites had adult patients that were predominantly white and non-Hispanic (75%) and older than 54 (72%). Practice sites had child patients who were predominantly white (60%) or Hispanic (15%).

Table 1.6a. Gender Frequencies

Gender	Adult Survey %	Child Survey %
Female	53	46
Missing Gender	10	2

Table 1.6b. Race/Ethnicity Frequencies

Race Category	Adult Survey %	Child Survey %
White, Non-Hispanic	75	60
Black, Non-Hispanic	3	4
Hispanic or Latino	5	17
Other	11	12
Missing Race, Ethnicity	7	6

Table 1.6c. Age Category Frequencies

Adult Survey		Child Survey	
% 18 - 24	2	% 0 - 3	35
% 25 - 34	7	% 4 - 7	19
% 35 - 44	7	% 8 - 11	19
% 45 - 54	12	% 12-19	26
% 55 - 64	22	% Missing Age	0.1
% 65 - 74	27		
% 75+	23		
% Missing Age	0		

Table 1.6d. Survey Mode Frequencies

Survey Mode	Adult Survey %	Child Survey %
Mail	47	62
Phone	13	2
Interactive Voice Response	33	20
Web / Internet	7	13
Missing	0	3

Table 1.6e. Self-reported General Health Rating Frequencies

Response	Adult Survey %	Child Survey % (Child Health)
Excellent	12	47
Very good	31	31
Good	32	12
Fair	15	3
Poor	4	< 1
Missing	6	6

Table 1.6f. Self-reported Mental Health / Emotional Health Rating Frequencies

Response	Adult Survey %	Child Survey % (Child Health)
Excellent	29	52
Very good	33	25
Good	23	11
Fair	8	4
Poor	1	1
Missing	6	7

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.

Social risk factors collected include education and language. The tables below show the distribution of these two variables in the dataset.

Table 1.8a. Education Level Frequencies

Education Level	Adult Survey %	Child Survey % (Respondent Education)
8th grade or less	2%	2%
Some high school, but did not graduation	3%	2%
High school graduate or GED	22%	12%
Some college or 2-year degree	28%	27%
4-year college graduate	16%	24%
More than 4-year college degree	21%	25%
% Missing	9%	9%

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

Type	Adult Survey %	Child Survey %
English	99	93
Spanish	1	4
Other	0	3

2a2. RELIABILITY TESTING

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 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 practice variance to within practice 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. 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 groups/units, 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 practices. Therefore, we do not feel conducting additional spline analyses, 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 practices.

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; Hays, Berman 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 R, Berman L, Kanter M, et al. (2014). Evaluating the Psychometric Properties of the CAHPS Patient-Centered Medical Home Survey. *Clin Ther.* May ; 36(5): 689–696.e1.

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 R, Shaul J, William V, et al. (1999) Psychometric Properties of the CAHPS™ 1.0 Survey Measures. *Medical Care.* Volume 37(3) SUPPLEMENT, March 1999, pp MS22-MS31.

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.

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 practice site and practice site-level reliability statistics for the surveys.

Table 2a2. 3a. Cronbach's Alpha Reliability Coefficient for CG-CAHPS Adult Version 3.0 Sample, Jan 2016 - March 2017 (656 practice sites, 137,416 Respondents)

Adult Survey Item	Standardized Cronbach's Alpha	Cronbach's Alpha if Item Deleted
ACCESS COMPOSITE	**0.77**	
Got appointment for urgent care as soon as needed (Q6)		0.60
Got appointment for check-up or routine care as soon as needed (Q8)		0.66
Got answer to phone question during regular office hours on same day (Q10)		0.81
COMMUNICATION COMPOSITE	**0.89**	
Provider explained things clearly (Q11)		0.87
Provider listened carefully (Q12)		0.84
Provider showed respect (Q14)		0.85
Provider spent enough time (Q15)		0.87
COURTEOUS/HELPFUL STAFF	**0.81**	
Office staff was helpful (Q21)		n/a
Office staff showed courtesy and respect (Q22)		n/a
CARE COORDINATION COMPOSITE	**0.55**	
Provider knew important info about medical history (Q13)		0.42
Provider's office followed up with test results (Q17)		0.46
Talked about prescription medicines at each visit (Q20)		0.48

Table 2a2. 3b. Cronbach's Alpha Reliability Coefficient for CG-CAHPS Child Version 3.0 Sample, Jan 2016-Feb 2017 (81 practice sites, 12,488 Respondents)

Child Survey Item	Standardized Cronbach's Alpha	Cronbach's Alpha if Item Deleted
ACCESS COMPOSITE	**0.75**	
Got appointment for urgent care as soon as needed (Q13)		0.58
Got appointment for check-up or routine care as soon as needed (Q15)		0.67
Got answer to phone question during regular office hours on same day (Q17)		0.75
COMMUNICATION COMPOSITE	**0.88**	
Provider explained things clearly (Q18)		0.86
Provider listened carefully (Q19)		0.82
Provider showed respect (Q21)		0.83
Provider spent enough time (Q22)		0.86
COURTEOUS/HELPFUL STAFF	**0.84**	
Office staff was helpful (Q26)		n/a
Office staff showed courtesy and respect (Q27)		n/a
CARE COORDINATION COMPOSITE	**0.39**	
Provider knew important information about your child's medical history (Q20)		**n/a**
Provider's office followed up with test results (Q24)		**n/a**

Table 2a2. 3c. Practice Site-Level Reliability Statistics for Clinician and Group CAHPS, Version 3.0, January 2016 – March 2017

	Adult (635 practice sites)			Child (77 practice sites)		
Measures	**Average # of Respondents per site**	**ICC**	**Site-Level Reliability**	**Average # of Respondents per site**	**ICC**	**Site-Level Reliability**
ACCESS: Getting Timely Appointments, Care, and Information (3 items)	149	0.043	0.87	141	0.029	0.81
COMMUNICATION: How Well Providers Communicate with Patients (4 items)	175	0.027	0.83	160	0.015	0.71
CARE COORDINATION: Providers' Use of Information to Coordinate Patient Care (3 items for Adult and 2 items for Child surveys)	174	0.033	0.85	158	0.021	0.78
COURTEOUS/HELPFUL STAFF: Helpful, Courteous, and Respectful Office Staff (2 items)	169	0.038	0.87	155	0.046	0.88
GLOBAL: Patients' Overall Rating of Provider (1 item)	167	0.032	0.84	155	0.029	0.82

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 question, except adult survey question #10 in the Access composite, would not result in a higher Cronbach's alpha. Therefore, we would not want to remove these questions. At the individual level, Care Coordination, had internal consistency reliability less than 0.70 on both the Adult (α = 0.55) and Child (α = 0.39) surveys. We know that this particular Care Coordination measure is not as reliable as we would like, but, of course, Cronbach's coefficient alpha is not what is as important but rather target unit-level (e.g., practice) reliability. Nevertheless, this is an important aspect in health care that many providers are aiming to improve.

The ICC values range from 0.027 to 0.043 on the Adult survey and 0.015 to 0.046 on the Child survey. These values show that these measures would need a larger sample size to effectively discriminate among practices, however these values are also influenced by the sample size of the dataset used. Therefore, we also examined the site-level reliabilities. All of the composites and global ratings, including Care Coordination, have site-level reliabilities that exceed 0.70, ranging from 0.71 to 0.88. For the adult survey, 3 out of 5 are greater than or equal to 0.85 which is considered very good. The child survey has one composite that is greater than or equal to 0.85.

Again, because the CG-CAHPS survey is used to compare practice sites, the site-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)*

The HP-CAHPS composites were developed over ten years ago using thorough cognitive interviewing and stakeholder involvement. Details about the development and testing were part of the measure's first NQF endorsement in July 2007. 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.

CG-CAHPS version 3.0 (implemented in July 2015) included a new 3-item composite for assessing care coordination. Care coordination is an important aspect of the patient experience. The Consortium's goal was to develop a care coordination composite measure that could be standardized across CAHPS surveys. Since two of the items were already part of the core survey, the new composite required the addition of only one item to the core survey. During development, testing of the composite found a strong unique association with the CAHPS global rating of health care, controlling for the other CAHPS core composite scores. This measure can be used to evaluate relative unit performance and characteristics associated with better care coordination (Hays, Martino, et al., 2014). Items in "Providers' Use of Information to Coordinate Patient Care" composite are:

1. Follow up on test results (stand-alone item already in CG-CAHPS core survey)
2. Knows important information about medical history (previously in Communication composite)
3. Provider talked about all prescription medicines being taken (from the PCMH Item Set)

At the individual and practice site 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 measure of "How would you rate your provider?" using Spearman rank-order correlations. For example, the composite measuring how well patients' provider communicates is expected to be strongly related to the patients' overall rating of their provider. Finding such a relationship supports interpretation of the composite as a valid measure of patient experience with a provider.

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.

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.

Doyle CL, 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.

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.

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

Table 2b1. 3a. Individual-Level and Practice-Level Correlation of Items, Composites and Global Rating Mean Top Box Scores for Adult CAHPS Clinician and Group Version 3.0 Sample, 2016-2017

Adult Survey Item	Global Rating of Provider (N = 112,458 Respondents)	Global Rating of Provider (N – 635 practices
ACCESS COMPOSITE	**0.27*** **	**0.50*** **
Got appointment for urgent care as soon as needed (Q6)	0.28***	0.42***
Got appointment for check-up or routine care as soon as needed (Q8)	0.25***	0.45***
Got answer to phone question during regular office hours on same day (Q10)	0.27***	0.42***
PROVIDER COMMUNICATION COMPOSITE	**0.51*** **	**0.72*** **
Provider explained things clearly (Q11)	0.44***	0.65***
Provider listened carefully (Q12)	0.49***	0.68***
Provider showed respect (Q14)	0.45***	0.67***
Provider spent enough time (Q15)	0.43***	0.65***
COURTEOUS/HELPFUL STAFF	**0.26*** **	**0.42*** **
Office staff was helpful (Q21)	0.24***	0.41***
Office staff showed courtesy and respect (Q22)	0.23***	0.40***
CARE COORDINATION COMPOSITE	**0.39*** **	**0.60*** **
Provider knew important info about medical history (Q13)	0.44***	0.70***
Provider's office followed up with test results (Q17)	0.30***	0.41***
Talked about prescription medicines at each visit (Q20)	0.23***	0.29***

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

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

Table 2b1. 3b. Individual-Level and Practice-Level Correlation of Items, Composites and Global Rating Mean Top Box Scores for Child CAHPS Clinician and Group Version 3.0 Sample, 2016-2017

Child Survey Item	Global Rating of Provider (12,471 respondents)	Global Rating of Provider (77 practices)
ACCESS COMPOSITE	**0.23*** **	**0.42*** **
Got appointment for urgent care as soon as needed (Q13)	0.24***	0.33**
Got appointment for check-up or routine care as soon as needed (Q15)	0.21***	0.24*
Got answer to phone question during regular office hours on same day (Q17)	0.26***	0.26*
COMMUNICATION COMPOSITE	**0.50*** **	**0.66*** **
Provider explained things clearly (Q18)	0.40***	0.59***
Provider listened carefully (Q19)	0.47***	0.67***
Provider showed respect (Q21)	0.44***	0.48***
Provider spent enough time (Q22)	0.44***	0.58***
COURTEOUS/HELPFUL STAFF	**0.23*** **	**0.35** **
Office staff was helpful (Q26)	0.22***	0.34**
Office staff showed courtesy and respect (Q27)	0.21***	0.31**
CARE COORDINATION COMPOSITE	**0.39*** **	**0.47*** **
Provider knew important information about your child's medical history (Q20)	0.42***	0.67***
Provider's office followed up with test results (Q24)	0.25***	0.29*

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

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

Table 2b1. 3c. Practice site-level Top Box Composite Intercorrelations for CG-CAHPS Surveys, Version 3.0 Sample

Composites	Access	Provider Communication	Courteous/ Helpful Staff	Care Coordination
Access	1	0.39***	0.67***	0.41***
Provider Communication	0.49***	1	0.28*	0.59***
Office Staff Service	0.52***	0.44***	1	0.36**
Care Coordination	0.47***	0.68***	0.42***	1

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

Note: Values below the shaded diagonal are for the Adult survey, values above are for the Child survey. 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?*)

The adult and child survey individual-level results show that all four composites are strongly related to the global provider rating scale, with the Communication composite having the strongest relationship followed by Care Coordination.

As hypothesized and found at with the individual-level results, the Adult and Child survey site-level results show that each of the four composites are strongly related to the global rating scale, again with the Communication composite having the strongest relationship.

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). As shown in Table 2b1.3c, relationships among the composites met our expectations at both the individual and site level, with none exceeding 0.70.

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.

Clinician and Group CAHPS 3.0 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 practices. For this purpose, users are able to find guidance and support in from the documents "Preparing Data from CAHPS Surveys 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; Eselius et al, 2008; 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 patient experience 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 better patient experience 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 of 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 practice setting. For example, a study by Drake and colleagues (2014) found that telephone respondents gave more positive responses than mail respondents did. However, within a practice or hospital there is often little variability in mode (Elliot et al, 2009). As reported in Table 1.6.d, 47 percent of the CG-CAHPS Adult results and 62 percent of the CG-CAHPS Child results submitted to the 2017 database were completed by mail.

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

AHRQ (2017) "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

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, L.L., Cleary, P.D., Zaslavsky, A.M., Huskamp, H.A., & Busch, S.H. (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.

Hepner KA, Brown JA, Hays RD. (2005) Comparison of mail and telephone in assessing patient experiences in receiving care from medical group practices. *Evaluation and the Health Professions*, 28, 377-389.

Martino, Elliott, Kanouse, Farley, Burkhart, and 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., and A. C. Bara. (2010) Patient Characteristics and Quality Dimensions Related to Patient Satisfaction." *International Journal for Quality in Health Care* 22(2): 86–92.

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

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

Annual Review of Public Health 26: 513–59.

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

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 Health Plan CAHPS; a survey with similar composites to CG-CAHPS about access to care, provider communication and service. 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-reported experience. For this purpose, this question was added to CG-CAHPS: "In general, how would you rate your overall mental or emotional health?"

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 three patient care dimensions that are common to both Health Plan CAHPS and Clinician and Group CAHPS: Doctor Communication composite, Overall rating of care, and Care coordination. 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 report 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

Note: Adjustment coefficients represent the effect of a unit change in the predictor relative to (divided by) the standard deviation of the heath plan intercepts (adjusted health plan means).

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., physician practices) 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 practices, we calculate the Pearson product moment, Kendall Tau correlation coefficients, and the maximum difference between the adjusted and unadjusted practice ratings. The adjustment factors include age, education, mental health status and general health status. We also calculate the maximum absolute value difference between adjusted and unadjusted scores among practices.

Table 2b3. 6a. Association between Adjusted and Unadjusted Mean Composite Scores: Adult Survey (635 Practices)

Composite	Pearson Correlation	Kendall Correlation
Access	0.99	0.90
Provider Communication	0.98	0.88
Courteous/Helpful Staff	0.98	0.85
Care Coordination	0.99	0.92

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

Composite	Pearson Correlation	Kendall Correlation
Access	0.99	0.91
Provider Communication	0.99	0.88
Courteous/Helpful Staff	1.00*	0.96
Care Coordination	0.99	0.91

* Actual correlation was 0.9969

Table 2b3.6c Case-mix Adjusted Mean Scores for CG-CAHPS Adult Version 3.0 Sample, Jan 2016 - March 2017 (635 practice sites, 112,458 Respondents)

Adult Survey Item	Maximum Difference Between Adjusted & Unadjusted Scores
ACCESS COMPOSITE	0.13
Got appointment for urgent care as soon as needed (Q6)	0.15
Got appointment for check-up or routine care as soon as needed (Q8)	0.13
Got answer to phone question during regular office hours on same day (Q10)	0.17
COMMUNICATION COMPOSITE	0.08
Provider explained things clearly (Q11)	0.07
Provider listened carefully (Q12)	0.08
Provider showed respect (Q14)	0.07
Provider spent enough time (Q15)	0.11
COURTEOUS/HELPFUL STAFF	0.11

Adult Survey Item	Maximum Difference Between Adjusted & Unadjusted Scores
Office staff was helpful (Q21)	0.14
Office staff showed courtesy and respect (Q22)	0.09
CARE COORDINATION COMPOSITE	0.11
Provider knew important info about medical history (Q13)	0.13
Provider's office followed up with test results (Q17)	0.16
Talked about prescription medicines at each visit (Q20)	0.09

Table 2b3.6d Case-mix Adjusted Mean Scores for CG-CAHPS Child Version 3.0 Sample, Jan 2016-Feb 2017 (77 practice sites, 12,471 Respondents)

Child Survey Item	Maximum Difference Between Adjusted & Unadjusted Scores
ACCESS COMPOSITE	0.09
Got appointment for urgent care as soon as needed (Q13)	0.10
Got appointment for check-up or routine care as soon as needed (Q15)	0.08
Got answer to phone question during regular office hours on same day (Q17)	0.10
COMMUNICATION COMPOSITE	0.03
Provider explained things clearly (Q18)	0.03
Provider listened carefully (Q19)	0.03
Provider showed respect (Q21)	0.03
Provider spent enough time (Q22)	0.07
COURTEOUS/HELPFUL STAFF	0.04
Office staff was helpful (Q26)	0.04
Office staff showed courtesy and respect (Q27)	0.03
CARE COORDINATION COMPOSITE	0.06
Provider knew important information about your child's medical history (Q20)	0.05
Provider's office followed up with test results (Q24)	0.11

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 practice rankings are above 0.80 for each of the composites. These high correlations show modest effects of case-mix adjustment on CAHPS practice scores. As shown in Tables 2b3.6c and 2b3.6d, maximum absolute value difference between adjusted and unadjusted scores for individual practices, ranges from 0.07 to 0.17 for Adult and 0.03 to 0.11 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 rating for a practice versus the mean for all practices with p<.05 used as the criterion for determining significance. Because of the large sample size within the CAHPS database and within individual practice sites, relatively small differences between practice sites may be statistically significant.

The CAHPS analysis program allows users to perform testing 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

Products that assist with the determination of meaningful differences in performance have been developed from the data contributed to the CAHPS Clinician & Group Survey Database. Database 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 practices 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 Clinician & Group Survey Database can be found at: https://cahpsdatabase.ahrq.gov/files/2016CAHPSClinicianGroup.htm.

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 practice sites' mean item or composite score was significantly above or below the overall database mean item or composite score. The comparison against an the database as an anchor is an approach for estimating magnitude of differences in patient experience scores.

Quigley, D. D., Elliott, M. N., Setodji, C. M., & Hays, R. D. (2018). Quantifying magnitude of group-level differences in patient experiences with health care. Health Services Research., 53 (4), 3027-3051.

Table 2b4.2a Adult CG-CAHPS Version 3.0: Number of Sites Significantly Above or Below the 2017 CAHPS Database Average (N=635 Practice Sites)

Adult Survey Item (Total of 635 Sites)	Count of Practice Sites Significantly Above	Count of Practice Sites Significantly Below	% statistically different from CAHPS Database Average
ACCESS COMPOSITE (3 items)	**149**	**102**	**40%**
Got appointment for urgent care as soon as needed (Q6)	131	71	32%
Got appointment for check-up or routine care as soon as needed (Q8)	149	101	40%
Got answer to phone question during regular office hours on same day (Q10)	121	77	32%
COMMUNICATION COMPOSITE	**162**	**70**	**37%**
Provider explained things clearly (Q11)	128	58	29%
Provider listened carefully (Q12)	157	50	33%
Provider showed respect (Q14)	155	50	32%
Provider spent enough time (Q15)	143	79	35%
COURTEOUS/HELPFUL STAFF	**171**	**102**	**43%**
Office staff was helpful (Q21)	161	105	42%
Office staff showed courtesy and respect (Q22)	170	85	40%
CARE COORDINATION COMPOSITE	**178**	**94**	**43%**
Provider knew important info about medical history (Q13)	136	71	33%
Provider's office followed up with test results (Q17)	199	96	47%
Talked about prescription medicines at each visit (Q20)	149	87	37%
RATING OF PROVIDER	**141**	**64**	**32%**

Table 2b4.2b Child CG-CAHPS Version 3.0: Number of Sites Significantly Above or Below CAHPS Database Average (N=77 Practice Sites)

CHILD SURVEY ITEM	Count of Practice Sites Significantly Above	Count of Practice Sites Significantly Below	% statistically different from CAHPS Database Average
ACCESS COMPOSITE	**19**	**11**	**39%**
Got appointment for urgent care as soon as needed (Q13)	17	8	32%
Got appointment for check-up or routine care as soon as needed (Q15)	17	11	37%
Got answer to phone question during regular office hours on same day (Q17)	21	5	34%
COMMUNICATION COMPOSITE	**17**	**5**	**29%**
Provider explained things clearly (Q18)	14	2	21%
Provider listened carefully (Q19)	15	3	23%
Provider showed respect (Q21)	18	2	26%
Provider spent enough time (Q22)	18	8	34%
COURTEOUS/HELPFUL STAFF	**18**	**12**	**39%**
Office staff was helpful (Q26)	17	12	38%
Office staff showed courtesy and respect (Q27)	19	11	39%
CARE COORDINATION COMPOSITE	**15**	**6**	**27%**
Provider knew important information about your child's medical history (Q20)	13	3	21%
Provider's office followed up with test results (Q24)	14	4	25%
RATING OF PROVIDER	**21**	**6**	**35%**

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 generate results that sufficiently discriminate between practice sites in terms of patient reported experience of care quality. For many items, more than a third of the practice sites are significantly below (or above) the average score.

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.

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 data included in the CAHPS database, because we cannot calculate or empirically verify survey response rates provided by users for practice sites. 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 individuals selected subtracting out those who are deceased or ineligible. For CAHPS analyses and reports, this rate is adjusted as shown in the following formula:

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

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:

The total number in the denominator should include the following:

- **Refusals.** The individual (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 sampled individual (or parent or guardian) is presumed to be eligible but was never located.
- **Respondents.** The individual returned a questionnaire, whether complete, incomplete, or partially complete.

Denominator Exclusions:

- **Deceased**. In some cases, a household or family member may inform you of the death of the sampled individual or child.
- **Ineligible.** The sampled individual or child did not receive care from the participating medical group or health system in the last 6 months.

Users are provided the following advice for improving response rates:

- Improve initial contact rates by making sure that addresses, phone numbers, and email addresses 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, CD-ROM directories, free or subscription-based 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 mail, email, and telephone data collection procedures. In field tests, the combined approach was more likely to achieve a desired response rate than a single mode alone.
- Train interviewers on how to deal with gatekeepers (someone such as a relative who stands between the interviewer and the respondent, making it difficult or impossible to complete the interview).
- Train interviewers on refusal aversion/conversion techniques.

Item Non-Response

The method used to construct CAHPS scores maximizes the use of available data by averaging available individual-level responses in construction of an overall score for the practice site. For each individual item, the top box score is the 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 practice site-level statistics generally have a small percentage of 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 patients who have utilized certain services during the past 6 months. As a result, some CG-CAHPS items have high percentages of missing data overall, but when skip patterns are considered, the percentages of 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).

Citations:

AHRQ "Fielding the CAHPS® Clinician & Group Surveys" Document No. 2033. Updated 6/12/2017. Accessible at: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.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. Adult CG-CAHPS Version 3.0: Item-level Percent Missing (635 Practices, 112,458 respondents)

Adult Survey Item	% Missing at Practice Site Level	% Missing Individual Level	% Missing due to Appropriate Skip
ACCESS COMPOSITE			
Got appointment for urgent care as soon as needed (Q6)	1%	5%	56%
Got appointment for check-up or routine care as soon as needed (Q8)	1%	6%	30%
Got answer to phone question during regular office hours on same day (Q10)	1%	5%	57%
COMMUNICATION COMPOSITE			
Provider explained things clearly (Q11)	0%	2%	—
Provider listened carefully (Q12)	0%	2%	—
Provider showed respect (Q14)	0%	3%	—
Provider spent enough time (Q15)	0%	3%	—
COURTEOUS/HELPFUL STAFF			
Office staff was helpful (Q21)	0%	6%	—
Office staff showed courtesy and respect (Q22)	0%	6%	—
CARE COORDINATION COMPOSITE			
Provider knew important info about medical history (Q13)	0%	3%	—
Provider's office followed up with test results (Q17)	1%	10%	27%
Talked about prescription medicines at each visit (Q20)	0%	12%	13%
RATING of PROVIDER	**0%**	**6%**	—

"—" denotes not applicable.

Table 2b6.2b. Child CG-CAHPS Version 3.0: Item-level Percent Missing (77 practice sites, 12,471 Respondents)

Child Survey Item	% Missing at Practice Site Level	% Missing Individual Level	% Missing due to Appropriate Skip
ACCESS COMPOSITE			
Got appointment for urgent care as soon as needed (Q13)	0%	4%	50%
Got appointment for check-up or routine care as soon as needed (Q15)	1%	4%	28%
Got answer to phone question during regular office hours on same day (Q17)	1%	3%	57%
COMMUNICATION COMPOSITE			
Provider explained things clearly (Q18)	0%	2%	—
Provider listened carefully (Q19)	0%	2%	—
Provider showed respect (Q21)	0%	2%	—
Provider spent enough time (Q22)	0%	3%	—
COURTEOUS/HELPFUL STAFF			
Office staff was helpful (Q26)	0%	5%	—
Office staff showed courtesy and respect (Q27)	0%	5%	—
CARE COORDINATION COMPOSITE			
Provider knew important information about your child's medical history (Q20)	0%	3%	—
Provider's office followed up with test results (Q24)	5%	5%	61%
RATING of PROVIDER	**0%**	**5%**	—

"—" denotes not applicable.

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 13 percent of cases are truly missing on most items, which suggests that our item-level results are likely not biased by systematic missing data due to item nonresponse.

The CAHPS Clinician & Group survey adequately improves the quality of responses by using several item-screeners within a skip-pattern of questions. The screening questions result in a high percentage of missing due to appropriate skips for these three items: 1) Got appointment for urgent care as soon as needed, 2) Got answer to phone question during regular office hours on same day, and 3) Got appointment for check-up or routine care as soon as needed. 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: Collected by survey of providers´ patients.

3b. Electronic Sources

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

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

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).

Though mixed-mode administration (i.e., mail and phone) is a viable strategy for the collection of CAHPS surveys, mail continues to be the most frequent mode for most CAHPS surveys. Users then create electronic databases of results after receipt of the completed hard copy survey through scanning or data entry. However, vendors may set up their database before data collection by populating the frame to assist in identifying non response.

Traditionally, the rationale for not using electronic sources more broadly is that mail and telephone are the best ways to obtain representative samples of patients based on the contact information that is available for sampling and data collection. E-mail has been added as a mixed mode strategy for physician groups with reliable email addresses for their patient population.

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.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.

To address data collection efficiency and to improve response rates, the CAHPS Consortium endorsed e-mail notification for web-based surveys as an additional mode of data collection (Drake et al., 2014; McInnes et al., 2012). The CAHPS Consortium recommends a mixed mode that would have two e-mail reminders and a follow-up by mail or telephone to all who are in the survey sample. The follow-up to the entire sample is necessary to get a representative set of responses from a practice's population, as not all patients may have e-mail.

Studies have shown that 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 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 have been 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.

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. Aug;49(4):1387-99.

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.

McInnes DK, Brown JA, Hays RD, Gallagher P, Ralston JD, Hugh M, Kanter M, Serrato CA, Cosenza C, Halamka J, Ding L, Cleary PD. (2012) Development and evaluation of CAHPS questions to assess the impact of health information technology on patient experiences with ambulatory care. Med Care. Nov;50 Suppl:S11-9.

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 Clinician and Group survey is available to users free of charge. In addition to the survey instrument, users can access comprehensive fielding, analysis, and reporting guides as well as SAS programming code that performs analysis and significance testing. These tools are available at: https://www.ahrq.gov/cahps/surveys-guidance/cg/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 Massachusetts Health Quality Partners http://www.mhqp.org/measure_and_report/?content_item_id=157 Greater Detroit Area Health Council http://www.gdahc.org/ Minnesota Community Measurement, in conjunction with the Minnesota Department of Health State Quality Reporting and Measurement System http://www.mnhealthscores.org/ Massachusetts Health Quality Partners http://www.mhqp.org/measure_and_report/?content_item_id=157 Greater Detroit Area Health Council http://www.gdahc.org/ Minnesota Community Measurement, in conjunction with the Minnesota Department of Health State Quality Reporting and Measurement System http://www.mnhealthscores.org/ Payment Program CMS Merit-based Incentive Payment System (MIPS) under the Quality Payment Program https://qpp.cms.gov/mips/overview Performance-Based Incentive Payment as part of Comprehensive Primary Care Plus (CPC+) https://innovation.cms.gov/initiatives/comprehensive-primary-care-plus Regulatory and Accreditation Programs CMS Reporting Requirements of Medicare Accountable Care Organizations http://acocahps.cms.gov/ Minnesota Community Measurement, in conjunction with the Minnesota Department of Health State Quality Reporting and Measurement System http://www.mnhealthscores.org/?p=our_reports&sf=clinic&category=16&sub_category=&search_phrase=&zipcode=&within=5&category_section= (2) Professional Certification or Recognition Program Distinction in Patient Experience Reporting by the National Committee for Quality Assurance (NCQA) Patient-Centered Medical Home (PCMH) Program https://www.ncqa.org/programs/health-care-providers-practices/patient-centered-medical-home-pcmh/distinction-in-patient-experience-reporting/ 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) Northeast Valley Health Cooperation (NEVHC) https://nevhc.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. (1) Payment Program

1. Merit-based Incentive Payment System (MIPS) under the Quality Payment Program - Centers for Medicare and Medicaid Services (CMS)

2. Purpose: For physicians who provide services to Medicare patients, MIPS is a new approach to paying for Medicare Part B items and services in which clinicians may receive an increase or decrease in their payments based on their performance. Beginning in 2018, the performance category counts towards the MIPS final score. Performance is measured through the data clinicians report in four areas - Quality, Improvement Activities, Advancing Care Information, and Cost. In the area of Quality, practices pick 6 measures of performance to submit. The CG-CAHPS for MIPS survey incorporates supplemental items to produce patient experience measures beyond those generated by the core survey. CG-CAHPS is a high priority quality measure for general practice and family medicine physician groups. The CAHPS for MIPS survey scores are available for public reporting on the Physician Compare website annually

3. This is a national program that includes most physicians in the U.S. Clinicians participating in MIPS may participate as individuals or as a member of a group. The CAHPS sample frame includes all patients from a practice-supplied roster.

A. (2) Payment Program

1. Comprehensive Primary Care Plus (CPC+) Program - Centers for Medicare and Medicaid Innovation (CMMI)

2. Purpose: (CPC+) is a national advanced primary care medical home model that aims to strengthen primary care through regionally-based multi-payer payment reform and care delivery transformation. CPC+ includes two primary care practice tracks with incrementally advanced care delivery requirements and payment options to meet the diverse needs of primary care practices in the United States. CPC+ includes a Performance-Based Incentive Payment that prospectively pays and retrospectively reconciles a performance-based incentive based on how well a practice performs on patient experience measures, clinical quality measures, and utilization measures that drive total cost of care. Patient experience measures are based on the CAHPS Clinician & Group Survey 3.0 with PCMH supplemental items administered at the practice level.

3. This national program includes 2,932 practices in 18 regions.

B. Recognition

1. Patient-Centered Medical Home Recognition by the National Committee for Quality Assurance

2. Purpose: NCQA recognizes clinicians and practices in key areas of performance. The recognition program includes an optional Distinction in Patient Experience reporting for those recognized practices that use the PCMH version of the CG-CAHPS survey. The NCQA Distinct in Patient Experience reporting for PCMH Recognition will expire in 2019. Medical practices can retain Distinction through its expiration date.

3. NCQA's Patient-Centered Medical Home Recognition Program is the most widely adopted Patient-Centered Medical Home evaluation program in the country. More than 12,000 practices (with more than 60,000 clinicians) are recognized by NCQA. And more than 100 payers support NCQA recognition through financial incentives or coaching. Results can be seen at https://reportcards.ncqa.org/#/clinicians/list?recognition=Patient-Centered%20Medical%20Home.

D. (1) Public Reporting

1. Massachusetts Health Quality Partners (MHQP)

2. Purpose: In 2018, MHQP announced the winners of its first annual "MHQP Patient Experience Awards."

MHQP receives over 65,000 responses to the annual CAHPS PCMH survey of commercially-insured patients across the state. The results were used to identify the top performing primary care practices.

3. For the 14th year, MHQP has released the results of its annual survey assessing patient experiences in primary care. The results of the survey from over 60,000 patients are posted on http://www.healthcarecompassma.org/ and the winners are listed on http://www.mhqp.org/2018/11/26/mhqp-announces-winners-of-new-patient-experience-awards/.

D. (2) Public Reporting

1. Greater Detroit Area Health Council

2. Purpose: Sponsored by the Greater Detroit Area Health Council, physician Organizations (POs) and physician offices across Michigan voluntarily participate in a patient survey project called the Michigan Patient Experience of Care Initiative (MiPEC). These practices are leaders in improvement and promoting transparency and accountability. The website shows how each PO compares to the average for all POs.

3. Scores from the Adult and Pediatric CG-CAHPS survey collected in 2017 are reported for 11 physician organizations within Michigan. http://www.mycarecompare.org/patient-survey/

D. (3) Public Reporting

1. Minnesota Department of Health

2. Purpose: Beginning in 2012, the Minnesota Department of Health mandated use of the adult and child CG-CAHPS survey for all physician clinics with a sufficient number of patients. MDH contracts with MN Community Measurement, an independent, non-profit community organization, to oversee the collection and reporting of survey results. MN Community Measurement publishes the survey results on its consumer-facing website, MNHealthScores. The goal is to accelerate the improvement of health by sharing information on health care quality, cost and patient experience with all Minnesotans.

3. In 2019, MN Community Measurement reports quality results and ranks CAHPS results for 674 clinics online at http://www.mnhealthscores.org/

E. Quality Improvement with Benchmarking

1. CAHPS Database, Agency for Healthcare Research and Quality

2. The CAHPS Database Online Reporting System is a Web-based platform for viewing results from the Clinician & Group Survey (CG-CAHPS). 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). Participation in the CAHPS Database is free and open to all survey users on a voluntary basis. Organizations that contribute data receive a Private Feedback Report in Excel that displays their own results compared to the overall average. Contributors also have access to technical assistance through the CAHPS User Network. The reporting system presents results for composite measures, rating measures, and individual survey items, organized according to survey version and field period. Displays include:

- "Top Box" Scores
- Bar charts
- Chartbook
- Percentiles
- Report Builder

3. The most recent data set compiled from CG-CAHPS survey data (including PCMH) includes results from 3,443 practices representing 366,994respondents. Data submitted for the latest Chartbook were collected from January 2017 through March 2018.

F. Internal Quality Improvement

1. Northeast Valley Health Cooperation

2. Northeast Valley Health Cooperation (NVHC), a federally qualified health center in California, analyzed their trend in CG-CAHPS scores to determine areas that needed improvement. More specifically, NVHC analyzed variation between sites to determine priority focus areas. Strategies were implemented to improve communication and wait times in the physician office. A presentation by NVHC can be viewed at

https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/news-and-events/events/20160315/debra-rosen-slides.pdf.

3. NVHC services Los Angeles County. They have 14 sites and 65,910 users/patients.

G. Regulatory

1. Reporting Requirements of Medicare Accountable Care Organizations

2. Purpose: The CAHPS Survey for Accountable Care Organizations (ACOs) is a version of the CAHPS Clinician & Group Survey for accountable care organizations participating in Medicare initiatives. The Centers for Medicare & Medicaid Services (CMS) requires the measurement of patient experience of care by ACOs involved in three programs: the Medicare Shared Savings Program (MSSP) and the Next Generation ACO Model. The ACO-CAHPS survey captures the core CG-CAHPS Survey plus the Summary Survey Measures that are part of the ACO quality standard and were finalized in the Shared Savings Program final rule.

3. For 2018, over 550 Medicare ACOs are in operation either through the Medicare program or through specific models managed by the Center for Medicare and Medicaid Innovation (CMMI). ACOs are located in almost all states and the District of Columbia.

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.

The first stage of the CAHPS program, known as CAHPS I (1995-2001), focused on the measurement and reporting of consumers' experiences with the care and services provided by health plans. All survey development was conducted by a public-private research collaborative known as the CAHPS Consortium. During this period, the Consortium also developed a CAHPS survey for group practices.

From 2003-2004, the CAHPS team conducted qualitative research to identify the key factors defining the quality of physician care from the patient's and the provider's perspective. Through critical incident interviews with 40 providers and 167 patients, the team collected over 3,000 critical incidents (that is, descriptions of specific behaviors performed by providers or their staff that were directly responsible for the visit being evaluated positively or negatively). These behaviors were organized into a taxonomy, leading to the identification of over 80 patient-staff behavioral domains. Each of these domains was evaluated with respect to its potential value as a patient-reported care measure. This work contributed significantly to the development of the Clinician & Group Survey by confirming and expanding the Consortium's understanding of the most important domains for the instrument. Portions of this work were presented at professional and public meetings. In 2004, the Consortium created a national advisory group to represent these various stakeholder audiences. This Advisory Group has convened at intervals in the development process to provide input on survey content.

Based on this research with users, the CAHPS Consortium embarked on an initiative known as Ambulatory Care CAHPS (A-CAHPS) resulting in two products: CAHPS Health Plan Survey 4.0 and CAHPS Clinician & Group Survey 1.0.

The CAHPS Clinician and Group survey has gone through two more revisions since 1.0, utilizing field and psychometric testing as well as collaboration with the NCQA and other stakeholders to increase the scientific rigor of the tool and the usability of the data for physician practices.

Throughout the development and testing of the Clinician & Group Survey, the CAHPS Consortium solicited input of stakeholders from the public and private sectors. Key stakeholders that assist with development of the CAHPS surveys and provide feedback are organizations that represent medical groups, medical societies, health plans, employers and other purchasers of care, consumers, and federal and State agencies. Other steps that have contributed to the content and design of the CG-CAHPS Survey over time have included:

- Literature review and review of existing measures
- Development and consultation with technical expert panels
- Focus groups with patient
- Cognitive testing
- Field testing
- Public comment
- On-going collaboration and harmonization with key partners and stakeholders

Throughout the development process, the CAHPS Consortium 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 patients and other stakeholders.

Most recently, to support the development of the 3.0 version of CG-CAHPS, AHRQ and the Consortium released a call for public comment on proposed changes to the core survey and PCMH item set. A summary of the feedback was published and then followed by further testing to obtain additional analytic results to more completely address comments received prior to the release of the final 3.0 version in 2015.

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 CG-CAHPS Database Online Reporting System is updated annually with new data submitted by CAHPS Clinician & Group survey users. The most recent CG-CAHPS Database ORS can be found at: https://cahpsdatabase.ahrq.gov/CAHPSIDB/Public/CG/CG_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.

Development

Between version 1.0 and 2.0, medical practices and other stakeholders provided feedback to the CAHPS Consortium. In addition to a focus on physicians, these groups wanted to measure patients' perceptions of care delivered by nurse practitioners, physician assistants and nurse midwives, among others. Therefore, the CAHPS instrument team developed and tested item wording where the referent caregiver is not restricted to physician-level providers.

The Yale and RAND CAHPS teams conducted cognitive interviews in both Spanish and English to test the draft "Provider version" of the CG instrument. The testing confirmed that "this provider" worked in the way it is intended – providing the necessary cognitive link to the named healthcare provider listed in the survey. A report of the methods and findings from the interviews conducted by the Center for Survey Research, University of Massachusetts Boston, a member of the Yale CAHPS team, is available upon request.

As noted above, a similar process of development and testing, solicitation of stakeholder feedback through public comment, consultation with a technical expert panel, and further testing was conducted to inform the transition between the 2.0 and 3.0 versions of CG-CAHPS.

Implementation

The CG-CAHPS Database Online Reporting System is updated annually with new data submitted by CAHPS Clinician & Group survey users. The most recent CG-CAHPS Database ORS can be found at: https://cahpsdatabase.ahrq.gov/CAHPSIDB/Public/CG/CG_About.aspx. User can contact the CAHPS Database team with questions or comments by phone at 888-808-7108 or email at CAHPSDatabase@westat.com. Updates are sent out via email by GovDelivery for those who have signed up to receive updates from the CAHPS Databases.

Educational Offerings

1. Last updated in 2017, the CAHPS Ambulatory Care Improvement Guide provides the following for users for CG-CAHPS and Health Plan CAHPS:
    - Why Improve Patient Experience? A compelling case for health care organizations to focus on improving their patients' experience with care.
    - An overview of behaviors common to health care organizations that have been effective in providing positive experiences with care.
    - A walk through the basic steps of a CAHPS-related quality improvement initiative.
    - A discussion of ways to analyze data from CAHPS surveys in order to identify opportunities to improve and various approaches to gathering additional information to help inform selection of strategies for improvement.
    - Descriptions of strategies that health care organizations can implement in order to help improve consumers' and patients' experiences with care.
2. AHRQ offers a variety of written materials and recordings to inform and support the improvement process accessible at https://www.ahrq.gov/cahps/surveys-guidance/cg/improve/index.html. For example, these presentations are currently available for viewing from AHRQ's Website:
    - Creating an Improvement Culture
    - The Role of Human Resources Staff in Improvement Efforts
    - A Breakthrough Approach to Improving CAHPS Communication
    - Strategies for Improving CAHPS Clinician & Group (CG-CAHPS) Survey Scores (March 15, 2016) accessible at https://www.ahrq.gov/cahps/news-and-events/events/webcast-031516.html

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Between versions 1.0 and 2.0 of Child CG-CAHPS, in response to feedback obtained during the previous NQF review, pediatric experts felt that the Child version of the CG-CAHPS Survey would benefit from a more comprehensive measurement of development and prevention. The CAHPS Consortium worked with the American Academy of Pediatrics and other key stakeholders to develop 11 new items that address development and prevention. These items were grouped into new two composite measures for the Child Survey version 2.0 (Gallagher et al, 2009).

Gallagher P, Ding L, Ham HP, Schor EL, Hays RD, Cleary PD. (2009) Development of a new patient-based measure of pediatric ambulatory care. Pediatrics. Nov;124(5):1348-54.

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

The CG-CAHPS Survey was initially developed with the standard CAHPS 4-point response scale. The field test data that was used for the initial endorsement consisted of several testing sites in the US. One of the larger field tests was conducted in Massachusetts as part of the Massachusetts Health Quality Partners (MHQP) statewide surveying initiative. At that time, MHQP used a 6-point response scale instead of the standard CAHPS 4-point scale. Based on that evidence, NQF endorsed the CG-CAHPS Survey with a 6-point response scale. Affirmed by significant user feedback, additional testing was conducted to add to the other field test data to confirm the properties and function of the standard CAHPS 4-point response scale and the CG-CAHPS Survey was updated to the 4-point response scale so that results could be aligned across other CAHPS surveys (e.g., CAHPS Health Plan Survey, CAHPS Hospital Survey (HCAHPS)). Drake et al (2014) examined how different response scales affect responses to the CG-CAHPS survey among 6,500 patients. They found that compared to the 4-category response options surveys, respondents to the 6-category response options surveys had 41 percent more missing items. There were no significant differences between the 4-category and 6-category response option surveys' average composite score means or provider-level reliability.

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. Aug;49(4):1387-99.

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

Between versions 1.0 and 2.0, users said that the benefits of using the chronic condition items to conduct subgroup analyses on identified individuals with chronic conditions was outweighed by the desire to streamline the core survey.

Feedback obtained in the development of the 3.0 version included extensive comments related to:

- The survey reference time period.
- Access composite measure.
- Communication composite measure.
- Office staff composite measure.
- Care coordination composite measure.
- Measures in the PCMH item set.

A summary of this feedback and the Consortium's responses are available at:

https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/about/public-summary-2015-cg-cahps-fedreg.pdf

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.

In late 2009, version 2.0 addressed the fact that many people receive care in the ambulatory setting from non-physician providers such as nurse practitioners and physicians´ assistants. This change to "this provider" was in response to requests from the medical community for a survey instrument that would allow patients to report on their experiences with all their health care practitioners.

The items designed to identify chronic conditions were moved from the core survey version 2.0 to a supplemental item set.

After extensive testing and receiving feedback from users, with the release of CG-CAHPS Version 2.0, the CAHPS consortium endorsed e-mail notification for web-based surveys as an additional mode of data

collection. The CAHPS consortium recommends a mixed mode that would have two e-mail reminders and a follow-up by mail or telephone to all who are surveyed.

Based on the stakeholder and user feedback obtained through public comment, technical expert review, and further testing, the following key changes were made in the release of the CG-CAHPS 3.0 version:

- One instrument, in contrast to the three instruments available for the 2.0 version.
- Use of a 6-month reference time period rather than a 12-month reference period.
- New and modified composite measures:
  - New composite measure for "Care Coordination."
  - Modified composite measure for "Access."
  - Modified composite measure for "Communication."
- A modified Patient-Centered Medical Home Item Set.
- Shift of development and prevention items from the core Child Survey to the Patient-Centered Medical Home Item Set.
- Overall reduced length.

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.

In 2018, the CAHPS Database released the CAHPS Clinician & Group Survey Database 2017 Chartbook. The Chartbook provides a figure showing trends in top box scores over time for the Adult 3.0 version of the survey. As the Version 3.0 was released in 2015, there are 3 years of data to show for the Adult version and no trends for the Child version. In the 3 years for the Adult version, the mean top box scores for Access has remained flat. However, there has been a slight decrease on Rating of the Provider (84% to 80%), Care Coordination (75% to 73%), and the Office Staff composite (81% to 79%). See Figure 1 and Table 6 in

https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/cahps-database/comparative-data/2017-cg-chartbook.pdf

The Washington Health Alliance conducts a community-wide CG-CAHPS Survey to address underuse of effective care. The Washington Health Alliance (Alliance) administered CG-CAHPS four times since 2011. Between August and November 2017, survey results were collected representing 114 primary care medical groups and 351 clinics across the state of Washington. WHA reports that there is significant room to improve patient experience to achieve national top 10% performance.

The most recent report by the Washington Health Alliance can be found at
https://www.wacommunitycheckup.org/media/47164/2018-patient-experience-survey-report.pdf

In Minnesota, a comparison between the 2014 and 2016 statewide survey results for select composites and question items that are identical across the survey versions used (CG-CAHPS 2.0 and 3.0, respectively) showed that the performance of Minnesota clinics overall in 2016 slightly improved since 2014 on all comparable measures. (Source: “The Patient Experience of Health Care in Minnesota: A Review of Statewide Survey

Results", unpublished draft submitted to the Minnesota Department of Health by Shaller Consulting Group, January 31, 2018.)

In Massachusetts, four-year trend results found at www.healthcarecompassma.org, show increases over the past three years across all topic areas explored in the survey for both adult and pediatric care. This includes communication, integration of care, knowledge of patient, adult behavioral health, organizational access, self-management support, office staff, pediatric preventive care, child development, and willingness to recommend. (Source: http://mhqp.org/highlights/?content_item_id=262)

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 during implementation.

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

NA

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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.

The two related measures -- CAHPS for Accountable Care Organizations (ACO) and CAHPS for MIPS -- both use the core CG-CAHPS items. These versions have been recommended by the NQF Measure Applications Partnership (MAP) Clinician Workgroup for use by CMS in the Medicare Shared Savings Program and the Merit-Based Incentive Payment System, respectively. (See: http://www.qualityforum.org/Publications/2018/03/2018_MAP_Clinicians_-_Final_Report.aspx)

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 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: 2007

Ad.3 Month and Year of most recent revision: 07, 2015

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.7 Disclaimers: None.

Ad.8 Additional Information/Comments: