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 #: 1741

Measure Title: Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey Version 2.0

Measure Steward: American College of Surgeons, Division of Advocacy and Health Policy

Brief Description of Measure: The following 6 composites and 1 single-item measure are generated from the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surgical Care Survey. Each measure is used to assess a particular domain of surgical care quality from the patient's perspective.

Measure 1: Information to help you prepare for surgery (2 items)

Measure 2: How well surgeon communicates with patients before surgery (4 items)

Measure 3: Surgeon's attentiveness on day of surgery (2 items)

Measure 4: Information to help you recover from surgery (4 items)

Measure 5: How well surgeon communicates with patients after surgery (4 items)

Measure 6: Helpful, courteous, and respectful staff at surgeon's office (2 items)

Measure 7: Rating of surgeon (1 item)

The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Surgical Care Survey (S-CAHPS) is a standardized survey instrument that asks patients about their experience before, during and after surgery received from providers and their staff in both inpatient and outpatient (or ambulatory) settings. S-CAHPS is administered to adult patients (age 18 and over) that had an operation as defined by CPT codes (90 day globals) within 3 to 6 months prior to the start of the survey.

The S-CAHPS expands on the CAHPS Clinician & Group Survey (CG-CAHPS), which focuses on primary and specialty medical care, by incorporating domains that are relevant to surgical care, such as sufficient communication to obtain informed consent, anesthesia care, and post-operative follow-up and care coordination. Other questions ask patients to report on their experiences with office staff during visits and to rate the surgeon.

The S-CAHPS survey is sponsored by the American College of Surgeons (ACS). The survey was approved as a CAHPS product in early 2010 and the Agency for Healthcare Research and Quality (AHRQ) released version 1.0 of the survey in the spring of 2010. The S-CAHPS survey Version 2.0 was subsequently endorsed by NQF in June 2012 (NQF #1741). The survey is part of the CAHPS family of patient experience surveys and is available in the public domain at https://cahps.ahrq.gov/surveys-guidance/cg/about/index.html. Surgeons may customize the S-CAHPS survey by adding survey items that are specific to their patients and practice. However, the core survey must be used in its entirety in order to be comparable with other S-CAHPS data. The S-CAHPS survey is available in English and Spanish.

The 6 composite measures are made up of the following items:

The 1 single item measure (Measure 7) is (Q35): Using any number from 0 to 10, where 0 is the worst surgeon possible and 10 is the best surgeon possible, what number would you use to rate all your care from this surgeon?

Measure 1: Information to help you prepare for surgery (2 items)

Q3. Before your surgery, did anyone in this surgeon´s office give you all the information you needed about your surgery?

Q4. Before your surgery, did anyone in this surgeon's office give you easy to understand instructions about getting ready for your surgery?

Measure 2: How well surgeon communicates with patients before surgery (4 items)

Q9. During your office visits before your surgery, did this surgeon listen carefully to you?

Q10. During your office visits before your surgery, did this surgeon spend enough time with you?

Q11. During your office visits before your surgery, did this surgeon encourage you to ask questions?

Q12. During your office visits before your surgery, did this surgeon show respect for what you had to say?

Measure 3: Surgeon's attentiveness on day of surgery (2 items)

Q15. After you arrived at the hospital or surgical facility, did this surgeon visit you before your surgery?

Q17. Before you left the hospital or surgical facility, did this surgeon discuss the outcome of your surgery with you?

Measure 4: Information to help you recover from surgery (4 items)

Q26. Did anyone in this surgeon's office explain what to expect during your recovery period?

Q27. Did anyone in this surgeon's office warn you about any signs or symptoms that would need immediate medical attention during your recovery period?

Q28. Did anyone in this surgeon's office give you easy to understand instructions about what to do during your recovery period?

Q29. Did this surgeon make sure you were physically comfortable or had enough pain relief after you left the hospital or surgical facility where you had your surgery?

Measure 5: How well surgeon communicates with patients after surgery (4 items)

Q31. After your surgery, did this surgeon listen carefully to you?

Q32. After your surgery, did this surgeon spend enough time with you?

Q33. After your surgery, did this surgeon encourage you to ask questions?

Q34. After your surgery, did this surgeon show respect for what you had to say?

Measure 6: Helpful, courteous, and respectful staff at surgeon's office (2 items)

Q36. During these visits, were clerks and receptionists at this surgeon's office as helpful as you thought they should be?

Q37. During these visits, did clerks and receptionists at this surgeon's office treat you with courtesy and respect?

Developer Rationale: All of the measures submitted to NQF for endorsement share the main objective of the S-CAHPS survey, which is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information. At the same time, consistent with the survey, the measures aim to directly benefit a variety of users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care.

The S-CAHPS survey measures patients' perceptions of their experiences with pre-, during and post-surgical care. The survey addresses issues relevant to surgical care, such as informed consent, communication, clear and helpful information, anesthesia care, office staff helpfulness, and post-operative follow-up.

To offer surgical patients and surgeons valid and reliable information on patient experience of care, the American College of Surgeons (ACS), in partnership with other surgical and anesthesia organizations, sponsored the development of the S-CAHPS survey. The S-CAHPS survey is a patient experience-of-care survey measure specifically tailored for

surgical patients. The S-CAHPS survey was developed by working with patients to report on the full experience of surgical care, including their experience with the surgeon, the anesthesiologist, and the facility. Qualitative research (i.e., focus groups, cognitive testing, literature review) prior to field testing and focus groups following the main field test contributed to the development of the final composite measures. The data gathered through S-CAHPS survey data can assist consumers in identifying a high-quality surgeon and help surgeons to better understand and ultimately improve patient care.

To capture and ascertain the major domains surrounding the consumers' view of quality surgical care, the American Institutes for Research performed a literature review prior to developing the surgical CAHPS instrument. Using four literature databases: Medline, PsychInfo, CINAHL, and Evidence-Based Medicine Review, AIR reviewed 930 abstracts. These abstracts were narrowed from certain search terms and limitations. AIR asked the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP) members to provide input and guidance in these terms and limitations. The TAP included 21 members from various surgery and anesthesiology specialty societies, each having technical expertise in the topic of surgical care. The AIR research analyst then classified each abstract as relevant, possibly relevant, and not relevant.

From that point on, only the research analyst's "relevant" and "possibly relevant" classifications were reviewed by the project director. From this abstract review method, 37 abstracts were chosen to review in full. In addition, one article was added from an SQA Technical Advisory Panel (TAP) member. The 38 articles reviewed indicated 14 domains, or primary issues, of surgical patient care experience from a patient's perspective: information/education, interpersonal manner, pain, emotional support, accessibility/convenience, technical quality of care, efficacy/outcomes of care, availability, environment, customization/personalized care, patient involvement in care, continuity of care, overall satisfaction, and finances. These results guided the development of the surgical survey based on relevance to quality and to consumers and the ability of consumers to act as reliable reporters of the domain.

In addition, American Institutes for Research conducted six focus groups to identify important quality issues inherent in patients' experiences of surgical care. The results of the focus groups are included in Attachment D Main 1c5 Surgical CAHPS Focus Groups_2nd Round_2010.pdf. In total, 49 people participated in the focus groups, all of whom were 18 years of age or older and had undergone a surgical procedure billable with a 90-day global fee within the last 7 months. The groups were diverse both demographically and in the type of surgery. For recruiting the additional two groups of patients with more complex surgeries, each participant had their surgery in a hospital and stayed overnight for at least one night.

After the focus groups were conducted, the AIR project director, senior research analyst, and research analyst examined the notes and focused in on each of the above issues. They gathered recurring themes and issues into a report of results, used to develop insights about what patients feel characterizes quality surgical care. The three main domains were surgeon's interpersonal skills and behaviors, surgeon's expertise/technical competence, and surgeon's skill in communicating or providing health information and patient education. The results of the focus groups and the surgical patients' experiences of care ultimately guided development of the surgical survey. Though patients rated technical skill of the surgeon as highly relevant to quality, it was determined that patients are not the best reporters of technical surgical expertise, so this was not included as a domain in the survey instrument. Additionally, based on the psychometric analysis and discussions within the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP), ACS did not recommend the Anesthesia Care items as a reporting composite. While ACS believes these items should remain in the survey because anesthesia care is an important factor in the surgical patient's experience of care, the Anesthesiologist works within the hospital system, and thus, the surgeon cannot adequately control the experiences between patient and anesthesiologist, and should not be measured on such.

Numerator Statement: We recommend that S-CAHPS Survey items and composites be calculated using a top-box scoring method. The top box score refers to the percentage of patients whose responses indicated excellent performance for a given measure. This approach is a kind of categorical scoring because the emphasis is on the score for a specific category of responses.

The top box numerator for the Overall Rating of Surgeon 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 What´s Available for the CAHPS Surgical Care Survey: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/surgical/about/whats-available-surgical-care-survey.pdf

Also see Patient Experience Measures from the CAHPS Surgical Care Survey Document 409 obtained by going to: https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

Also, for more information on the calculation of reporting measures, see How to Report Results of the CAHPS Clinician & Group Survey, available at https://cahps.ahrq.gov/surveys-guidance/cg/cgkit/HowtoReportResultsofCGCAHPS080610FINAL.pdf.

Denominator Statement: The measure's denominator is the number of survey respondents. The target population for the survey is adult patients (age 18 and over) who had a major surgery as defined by Common Procedural Terminology (CPT) codes (90 day globals) within 3 to 6 months prior to the start of the survey.

Results will typically be compiled over a 12-month period.

For more information on the calculation of reporting measures, see Patient Experience Measures from the CAHPS Surgical Care Survey, available at https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html.

Denominator Exclusions: The following are excluded when constructing the sampling frame:

- Surgical patients whose procedure was greater than 6 months or less than 3 months prior to the start of the survey.

- Surgical patients younger than 18 years old.

- Surgical patients who are institutionalized (put in the care of a specialized institution) or deceased.

Measure Type: Outcome: PRO-PM

Data Source: Instrument-Based Data

Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: May 1, 2012 **Most Recent Endorsement Date:** May 1, 2012

Staff 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.

The submission contains information for 7 patient-reported outcome based performance measures (PRO-PMs) that are calculated from data aggregates from responses to the Surgical CAHPS survey. The 7 PRO-PMs include:

1. Measure 1: Information to help you prepare for surgery (2 items)
2. Measure 2: How well surgeon communicates with patients before surgery (4 items)
3. Measure 3: Surgeon's attentiveness on day of surgery (2 items)
4. Measure 4: Information to help you recover from surgery (4 items)
5. Measure 5: How well surgeon communicates with patients after surgery (4 items)
6. Measure 6: Helpful, courteous, and respectful staff at surgeon's office (2 items)
7. Measure 7: Rating of surgeon (1 item)

Summary of evidence from 2012 evaluation:

- Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey Version 2.0 indicates performance on the CAHPS Surgical Care Survey, which measures key components of patient experience, such as provider communication, that are consistent with patient-centered care.
- This is an patient-reported outcome performance measure. The developer provided a conceptual logic model describing the relationship between the Surgical Consumer Assessment of Healthcare Providers and Systems (S-CAHPS) measure and clinical outcomes.
- The developer provides a literature review of 18 studies and articles related to the measure including: a positive association between patient experience and clinical outcomes; a positive effect on key patient survey responses as a result of improved provider/staff communication and increasing patient partnership; and a positive effect of improved quality of medical consultations and patient education on the surgical patients' length of stay, anxiety levels, recovery time and compliance with treatment regimens.

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:

- The developer provides updated evidence of three systematic reviews that have examined the relationship between patient experience, clinical process and patient outcomes.
    - Patient experience is favorably associated with adherence to recommended medications and treatments, preventative care such as screenings and immunizations, patient-reported health outcomes, clinical outcomes, reduced healthcare utilization, and reduced adverse events. (Doyle et al., 2013)
    - 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. (Anhang Price, 2014)
    - Beattie et al. critiqued the utility of published patient-reported experience measures (PREMs) aiming to measure the adult inpatient experience of hospital quality of care based on the PREMs' validity, reliability, cost efficiency, acceptability and educational impact (Beattie et al., 2015). They identified eleven international PREMs and concluded that patient experience data could be used to drive improvements in hospital care at national, local, and healthcare team levels.
- The developer provided six additional sources that were not included in the previous measure submission.
    - Hospitals in the highest quartile of performance on patent satisfaction had length of stay that was on average 0.6 days shorter than those with the lowest patient satisfaction. (Tsai, Orav, Jah, 2015)
    - There is a significant association between patient satisfaction and bot 30-day readmissions and the occurrence of postoperative surgical complications. (Lobo Prabhu, 2017)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

- *If derived from patient report, does the target population value the measured outcome and finds it meaningful?*

Guidance from the Evidence Algorithm

Assess performance on a health outcome or PRO (box 1) → Relationship between PRO and healthcare action (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.

- Current performance data are calculated from a pilot study of the S-CAHPS in 2010-2011, which included data from 2,719 survey results from 32 practices across nine specialty types.

Measures	Top Box Mean	Standard Deviation	Median	Min	Max
Information to Help you Prepare for Surgery (2 items)	90%	0.05	90%	79%	98%
How Well Surgeon Communicates with Patients Before Surgery (4 items)	85%	0.07	84%	67%	98%
Surgeon's Attentiveness on Day of Surgery (2 items)	81%	0.12	84%	42%	97%
Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office (4 items)	82%	0.07	83%	64%	100%
How Well Surgeon Communicates With Patients After Surgery (4 items)	84%	0.06	84%	73%	97%
Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	87%	0.08	88%	58%	100%
Global rating of Surgeon	86%	0.07	88%	70%	98%

- In addition to performance data, the developer cited 13 studies and articles, including eight additional citations that were not included in the broader evidence overview, that indicate an opportunity for improvement or less than optimal performance rates in this healthcare area.

Disparities

- The developer reported disparities data on the factors gender, age, and ethnicity.
  - The developer reported slight differences in the rates by gender. On the item *Surgeon's Attentiveness on Day of Surgery*, male respondents report 82% while female reported 80%. On the item *How Well Surgeon Communicates with Patients After Surgery*, male respondents reported 85% while female respondents reported only 83%.
  - No difference in scores due to ethnicity were reported.

- Older patients generally reported more positive patient experiences.
- Patients age 18-24 (the youngest group) reported the highest top box scores for global rating of surgeons.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary ratings for opportunity for improvement:

1) Information to help you prepare for surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
2) How well surgeon communicates with patients before surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
3) Surgeon's attentiveness on day of surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
4) Information to help you recover from surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
5) How well surgeon communicates with patients after surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
6) Helpful, courteous, and respectful staff at surgeon's office:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
7) Rating of Surgeon:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**

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

1a. Evidence

Comments:

**The evidence for support of this measure is solid with three new systematic reviews and six additional studies being cited since this measure was initially endorsed. I am unaware of new studies that change the base of evidence for this measure.

**This measure does not depart substantially from the constructs already represented in H-CAPHS.

**Developer provided data form 3 systematic reviews supporting that patient-reported experience measures (PREMs) were associated with multiple positive outcomes (reduced complications ,better adherence, improved clinical outcomes, reduced utilization)

1b. Performance Gap

Comments:

**Data are provided that demonstrate a gap in performance and thus a performance measure is warranted. Disparities data was provided demonstrating differences between gender and age, but not ethnicity (however, the respondent sample was overwhelmingly white - 86%).

**I have concerns about: 1) at which level scores (i.e. who is the 'target population'?) are to be reported (site, surgeon); 2) in the data provided, for 11 of the 32 sites, practice and state were completely confounded; and 2) there appear to be ceiling effects in the top box scores 4 of the 7 S-CAHPS constructs based on data from the pilot study (p. 33).

**While all measures are favorably skewed, there is variation. Also, my own experience working with ABIM on physician peer review data indicated that even in assessments with ceiling effects, marked negative performance outliers can still be identified.

The SDS analysis seemed thin, with little ethnic gap demonstrated and no racial analysis (but racial data provided for test sample but not ethnicity data)? This strikes me as a significant gap and I would like to see developer evaluate disparities more comprehensively (particularly as I would presume that ethnic and racial minorities respond less frequently to PREMs than whites/non-Hispanics and they recommend risk adjustment by these factors, which might obscure relevant disparities compared with stratification).

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 Staff

Evaluation of Reliability and Validity:

Link A

Additional Information regarding Scientific Acceptability Evaluation:

While this is an outcome measure and therefore considered complex, the measure was not reviewed by the Scientific Acceptability Methods Panel because no new testing was submitted from previous endorsement review.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Scientific Acceptability

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 subcriteria for both reliability and validity to pass this criterion.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 1741

Measure Title: CAHPS Surgical Care Survey 2.0

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

   TIPS: Consider the following: Are all the data elements clearly defined? Are all appropriate codes included? Is the logic or calculation algorithm clear? Is it likely this measure can be consistently implemented?

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

   ☒Yes (go to Question #4)

   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?

   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

4. Was reliability testing conducted with computed performance measure scores for each measured entity?

   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

   ☒Yes (go to Question #5)

   ☐No (go to Question #8)

5. 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.*

   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☒Yes (go to Question #6)

   ☐No (please explain below then go to Question #8)

6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

   Do the results demonstrate sufficient reliability so that differences in performance can be identified?

   ☐High (go to Question #8)

   ☐Moderate (go to Question #8)

   ☒Low (please explain below then go to Question #7)

   For estimated reliability based on ICCs, .70 often is regarded as a minimum acceptable value

Score Level Reliability Rating	Measures	Average # of Respondents per site	Site-Level Reliability
Low	Information to Help You Prepare for Surgery (2 items)	85	0.52
Low	How Well Surgeon Communicates with Patients Before Surgery	76	0.68
Low	Surgeon Attentiveness on Day of Surgery (2 items)	83	0.50
Moderate	Information to Help You Recover from Surgery - from Surgeon or another Health Provider from the office.	85	0.71
Low	How Well Surgeon Communicates With Patients After Surgery	72	0.48
Moderate	Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	84	0.71
Low	One-item Global Rating of surgeon	82	0.60

In the original measure submission, developer notes that the site-level reliability analysis was conducted on field test data consisting of a relatively small selection of surgeon practices which may have led to reduced variability between sites.

7. Was other reliability testing reported?

   ☒Yes (go to Question #8)

   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

☒Yes (go to Question #9)

☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☒Yes (go to Question #10)

☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

Internal Consistency Reliability (Chronbach's alpha) was used to assess reliability for each of the survey questions.

10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

☒Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

☐Insufficient (go to Question #11)

Data Element Reliability Rating	**Measure and Items**	**Standardized Cronbach's Alpha**
Moderate	1. Information to Help You Prepare for Surgery (2 items)	0.74
Moderate	2. How Well Surgeon Communicates with Patients Before Surgery	0.82
Moderate	3. Surgeon Attentiveness on Day of Surgery (2 items)	0.66
Moderate	4. Information to Help You Recover from Surgery - from Surgeon or another Health Provider from the office.	0.84
Moderate	5. How Well Surgeon Communicates With Patients After Surgery	0.86
Moderate	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	0.85

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?

   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

   ☒Yes (go to Question #2)

   ☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

2. Analysis of potential threats to validity: Any concerns with measure exclusions?

   TIPS: Consider the following: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☐Yes (please explain below then go to Question #3)

   ☒No (go to Question #3)

   ☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

   b. Are social risk factors included in risk model? ☒Yes ☐No

   c. Any concerns regarding the risk-adjustment approach?

   *TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

   ☐Yes (please explain below then go to Question #4)

   ☒No (go to Question #4)

S-CAHPS Analysis program provides optional risk adjustment through case-mix adjusted scores. Case-mix adjustment is determined by users based on decision of what is most appropriate to adjust for to account for case-mix differences.

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

   ☐Yes (please explain below then go to Question #5)

   ☒No (go to Question #5)

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

   ☐Yes (please explain below then go to Question #6)

   ☒No (go to Question #6)

   ☐Not applicable (go to Question #6)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

   ☒Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

   ☐No (please explain below then go to Question #8)

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☐Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

☒Yes (go to Question #11)

☐No (please explain below and go to Question #13)

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

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☒Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

    The developer examined the relationships between each individual item's top box score and the top box score for the global PRO-PM of "How would you rate your surgeon?" using Spearman rank-order correlations at the site level to determine the validity of each of the PRO-PMs.

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☒High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

    The survey results show that each of the PRO-PMs are related to the global PRO-PM (rating of surgeon) at the individual level and five of the six PRO-PMs are related to the global rating at the practice level. The two Communication PRO-PMs and the Recovery Information PRO-PM have the strongest relationship with the global rating of surgeon.

Validity Measure Score Rating	**Measures**
High	1. Information to Help You Prepare for Surgery
High	2. How Well Surgeon Communicates with Patients Before Surgery
Moderate	3. Surgeon Attentiveness on Day of Surgery
High	4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.
High	5. How Well Surgeon Communicates With Patients After Surgery
High	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office
High	7. One-item Global Rating of Surgeon

13. Was other validity testing reported?

    ☒Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

☒Yes (go to Question #15)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

15. 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.*

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☒Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

    The developer examined Spearman rank-order correlations among the PRO-PMs to assess the extent to which they measure different constructs.

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☒Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☐Insufficient (go to Question #17)

    The developer examined Spearman rank-order correlations among the composites to assess the extent to which they measure different constructs. While the composites are correlated with each other, intercorrelations greater than 0.8 may indicate that the composites are not unique enough to be considered separate measures. One intercorrelation, "Communicate post surgery," and "communicate pre surgery" received a score at or above that level, with a score of 0.8.

    Developer states that the relationships met their expectations.

Validity Rating	Measures
Moderate	1. Information to Help You Prepare for Surgery
Moderate	2. How Well Surgeon Communicates with Patients Before Surgery
Moderate	3. Surgeon Attentiveness on Day of Surgery
Moderate	4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.
Moderate	5. How Well Surgeon Communicates With Patients After Surgery
Moderate	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office
Moderate	7. One-item Global Rating of Surgeon

17. OVERALL VALIDITY RATING

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 (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

Overall Validity Rating	**Measures**
High	1. Information to Help You Prepare for Surgery
High	2. How Well Surgeon Communicates with Patients Before Surgery
High	3. Surgeon Attentiveness on Day of Surgery
High	4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.
High	5. How Well Surgeon Communicates With Patients After Surgery
High	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office
High	7. One-item Global Rating of Surgeon

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

2a1: Reliability- Specifications

Comments

** Reliability specifications appear to be defined and clearly described. Consistent implementation may be challenged by resources to deploy the survey (mailing and emailing patients), data entry with accuracy checks (paper surveys), and conducting survey analyses

** I am most concerned about the sampling strategy given the intended attribution to apparently both the surgeon and the hospital. It is not clear from the data provided from the pilot exactly how many patients per surgeon and surgeons per practice were sampled. Score estimates that do not account for the nested nature of the data are problematic.

** The greatest concern is responder bias which seems to be mostly addressed through possible risk adjustment, which seems an inadequate response to this concern.

2a2: Reliability- Testing

Comments

**No.

** Yes. I am concerned about the "site" level analysis performed. Because of the hierarchical nature of the data, in some cases it appears that there may be 1 surgeon for some sites (i.e. complete confound) vs. up to 19 providers at each site. Variation between providers across sites vs within providers at the same site (where the site and surgeon are not completely confounded) did not appear to be provided. Also, the "site" level ICC's are very high. I am concerned that the reliability testing was at the item level within patients averaged for the site, not a true between provider/site vs. within provider site variation. Site level ICC's calculated in that way tend to produce coefficients in the range of .03-.05.

**No

2b1: Validity—Testing

2b4-7: Threats to Validity

2b4: Meaningful Differences

Comments

** The validity appears to be acceptable. In the initial survey of 2719 respondents, only 3.7% of data were missing which should not threaten validity.

** To assess whether differences observed in the quality improvement studies cited are meaningful (vs. within the standard error of measurement), estimates of the error variance at the units being compared are needed.

I am concerned about the inclusion of patients who completed at least one item of the composite (see p. 40). For the "how well surgeon communicates with patients after surgery", missing data was observed to be 16% at the individual level. It is not clear how/what imputation was done.

** I think the validity is acceptability strong. I particular like the developers use of patient focus groups to ensure domains/instrument content were truly patient-centered.

2b2-3: Other Threats to Validity

2b2: Exclusions

2b3: Risk Adjustment

Comments

** Appropriate patient groups are included. Risk adjustment - optional case-mix risk adjustment is available and is supported by available analysis instructions.

** It appears that the majority of patients (48% of survey respondents) from the pilot were 65+; this population typically reports higher scores on satisfaction measures, independent of the quality of care provided.

** The risk adjustment appears to be guidance and detailed information about predictive ability of the recommended risk variables was not found. Also, they do not provide any performance data stratified by SDS factors, despite flagging that these might be reasonable risk adjusters. In all, I found the information about risk adjustment less clear than the other components and I have concerns that stratification or other opportunities to illuminate disparities were not detailed or perhaps even considered.

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.

- This measure is collected via a survey of surgeon's patients.
- CAHPS surveys are primarily delivered via mail. Electronic databases are then created after mailed surveys are returned.
- Email has been added as a mixed mode strategy for surgeon groups with reliable email addresses for all of their population.
- Because the survey instrument, protocol, analysis, and reporting are standardized, surgeons can benchmark and compare their performance with that of their peers within the same practice or outside of their practice.
- There is no fee associated with the CAHPS measure.
- 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.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*

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

RATIONALE:

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility

Comments

** All data are derived from patient responses to a lengthy survey (47 questions). Having an electronic option may reduce the survey burden for patients with access to a computer, may increase data accuracy. and may enhance response rate.

** The response rate on S-CAHPS, as on H-CAHPS is low, raising issues about representativeness of the population of patients seen at sites/by providers.

** Feasibility remains a concern, as with all PROMs/PREMs, but the use of multiple modalities for data collection and I hope movement in the future towards lower burden electronic options will continue to minimize this issue.

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

- The measure is currently used in CMS Quality Payment Program (QPP) Merit-based Incentive Payment System (MIPS) and as part of an Advanced Alternative Payment Model (A-APMs). The measure is included in the CMS Core Quality Measures Orthopedics Set.
- The measure is also used quality improvement programs including the American College of Surgeon's National Surgical Quality Improvement Program.
- In the 2015 Medicare Physician Fee Schedule final rule, CMS agreed that the S-CAHPS survey would be more relevant to a surgical group practice compared to the CG-CAHPS and noted that the majority of commenters supported the use of S-CAHPS in the PQRS program. However, CMS did not accept and finalize the measure. CMS explained "due to the cost and time it would take to find vendors to collect S-CAHPS data, it is not technically feasible to implement the reporting of the S-CAHPS survey measures for the 2017 or 2018 PQRS payment adjustments.
- Pending CMS approval for the 2018 MIPS program, participants in the American College of Surgeons (ACS) Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism.

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 developer noted that it was "too early to broadly determine" how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

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

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.
4b.1 Improvement**.** Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- There is no repository for S-CAHPS data, therefore trend data is not available.

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

- The developer reported no unexpected findings have been uncovered.

Potential harms

- The developer reported no unexpected findings have been uncovered.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

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

4a1. Use- Accountability and Transparency

Comments

** The measure is currently publicly reported and used in accountability programs.
Developers report it is too early to determine the impact of the performance data being reported.
** Uses in MIPS programs raises concerns about reliability and validity, particularly at the surgeon level.
** Currently in QPP and ACS is integrating into APMs and their registry. I think goals for transparency are laudable. I would like to see efforts to make this data more available to patients as well.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0005 : CAHPS Clinician & Group Surveys (CG-CAHPS)-Adult, Child
- 0006: Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey, Version 5.0 (Medicaid and Commercial)
- 0166: HCAHPS
- 0258: CAHPS In-Center Hemodialysis Survey
- 0517: CAHPS Home Health Care Survey (experience with care)
- 2651: CAHPS Hospice Survey (experience with care)
- 2548: Child Hospital CAHPS (HCAHPS)
- 2967: CAHPS Home- and Community-Based Services Measures

Harmonization

- The Surgical Care Survey was updated in 2011 to remain consistent with the Clinician & Group Survey, which was also updated in 2011. The updates from do not affect the ability of survey users to assess trends in performance.

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

N/A

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 18, 2018

- No comments received.

- **Zero NQF members who have submitted a support/non-support choice.**

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

nqf_evidence_attachment_7.1_SCAHPS_FINAL.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*)**:** 1741

Measure Title: CAHPS Surgical Care Survey Version 2.0

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

Date of Submission: 11/8/2017

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.

- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Outcome: PRO-based Performance Measure (PRO-PM)

☒Patient-reported outcome (PRO): Experience with Surgical 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:

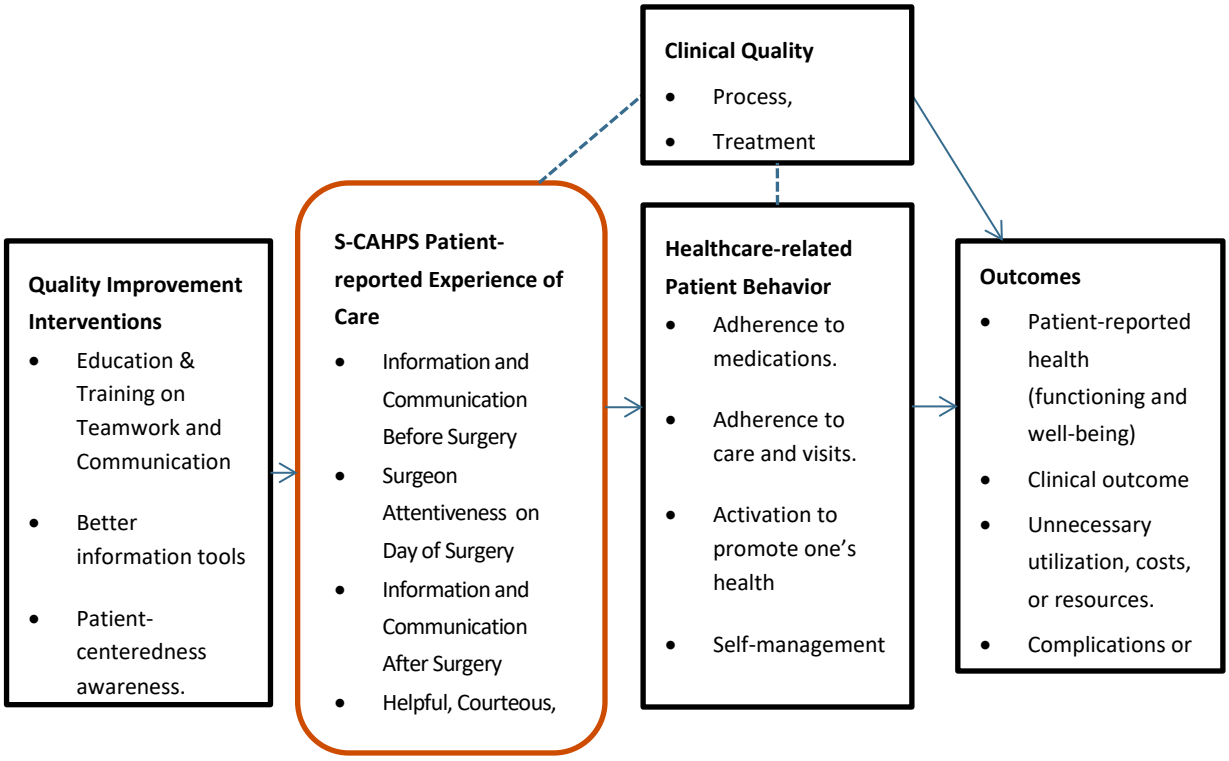
☐ Appropriate use measure:

☐ 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 following framework will be used to describe the patient-reported Surgical Consumer Assessment of Healthcare Providers and Systems (S-CAHPS) measure, the factors that influence it, and the relationship between S-CAHPS and clinical outcomes.



----- Hypothesized Association

___ Hypothesized Casual Association

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.)

**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

The CAHPS Surgical Care Survey measures key components of patient experience, such as provider communication, that are consistent with patient-centered care. All CAHPS surveys focus 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. Patient care experiences are influenced by the quality of patient care and interventions to improve quality of care. In turn, patient perceptions of high quality care improve patient adherence and activation, leading to improved clinical outcomes (See Figure 1.) Surgical CAHPS is an actionable measure that helps surgeons and their practices target interventions that will improve the quality and patient-centeredness of care.

REVIEW OF THE EVIDENCE

Three 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 healthcare utilization, and reduced adverse events (Doyle et al., 2013). More recently, in the U.S., Anhang Price et al. reviewed evidence on the association between patient experiences and other measures of health care quality (Anhang Price, 2014). 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. Beattie et al. critiqued the utility of published patient-reported experience measures (PREMs) aiming to measure the adult inpatient experience of hospital

quality of care based on the PREMs' validity, reliability, cost efficiency, acceptability and educational impact (Beattie et al., 2015). They identified eleven international PREMs and concluded that patient experience data could be used to drive improvements in hospital care at national, local, and healthcare team levels. Most importantly, the authors cautioned that clinicians, managers, policymakers and researchers need to select PREMs that are fit for purpose because only selecting the right PREM for the right purpose can the data aid in quality improvement.

RATIONALE OF INFLUENTIAL RELATIONSHIPS

Quality Improvement Initiatives/Interventions

The Agency for Healthcare Research and Quality (AHRQ) maintains a Quality Improvement (QI) Guide that assists providers and health plans in using CAHPS scores to identify problems, prepare staff for QI initiatives, select and implement interventions, and evaluate intervention effectiveness through changes in CAHPS scores. The guide presents interventions that have been successfully used to improve patient experience of care and other measures of quality relevant to CAHPS dimensions (AHRQ CAHPS QI Guide, 2012).

Relevant to surgical care, the AHRQ QI Guide provides information that can help surgeons to improve shared decision-making. Shared decision-making is a model of patient-centered care that enables and encourages people to play a role in the management of their own health. It operates under the premise that armed with good information, consumers can and will participate in the medical decision-making process by asking informed questions and expressing personal values and opinions about their conditions and treatment options. Devine et al. (1983 and 1988) found that improved quality of medical consultations and patient education had positive effect on the surgical patients' length of stay, anxiety levels, recovery time, and compliance with treatment regimens. Interactive technology and instructional applications are employed to help prepare patients for various procedures, to inform them about their surgery, and to explain what they need to know after surgery. The QI Guide notes the importance of training physicians to help them understand how to facilitate the shared decision-making process and to ensure that they appreciate the importance of respecting patient's values, preferences, and expressed needs. The shared decision-making model should be a practice-wide, team approach so that the surgeon's time is used appropriately.

A common intervention is communication skills training, with the purpose of improving provider communication (Hardee & Kasper, 2008). Five clinics in San Francisco took part in a two-year learning collaborative aimed at improving provide/staff-patient communication and increasing patient partnership. The intervention's focus was to ensure that patients' most important concerns were addressed during their visits. After the intervention, all five clinics showed significant improvement in survey items "doctor spends enough time", "doctor's explanations are understandable", "doctor provides easy-to-understand instructions", and "clerks and receptionists are helpful". Ten months post-intervention, these clinics had sustained statistically significant improvements in eight of 12 measures (Fisher, 2011). Improving staff service is also a common intervention. To address service issues, a surgical practice associated with an academic medical center offered a series of training courses to help staff interact with patients in a more positive manner. The practice's faculty group developed efforts to improve communication with patients. This emphasis helped to support the practice manager by signaling a wider commitment to the goal of improving patient experience. Their CAHPS scores for the helpful office staff composite increased to 87 (based on a 0-100 mean score) from a baseline score of 84, meeting the target set of >85. The overall rating of the doctor score increased from 89 to 95, exceeding the target of >90 (Shaller, 2011).

The CAHPS domain measures are available to assess all of the organizational processes that are addressed by health care organizations' innovations and for which patients are the best source of information. A recent study of AHRQ's Innovations Exchange website was conducted to examine the use of patient experience surveys in assessing the impact of innovations implemented in health care settings. Researchers found that fewer than half of the innovations used a patient experience measure. The authors conclude that there is considerable untapped potential for using CAHPS measures or surveys to assess QI initiatives' effectiveness. Organizations committed to patient-centeredness will benefit by monitoring patient survey data, along with clinical and operational data, to implement and measure quality improvement interventions and/or to evaluate patient reported outcomes associated with new models of care (Weinick, 2014; McWilliams, 2014).

Health-related Patient Behavior

Interventions targeting CAHPS dimensions indirectly improve clinical outcomes by positively influencing patient behavior (Fuertes, Boylan et al. 2009). For example, in a 2009 meta-analysis, Zolnierik and Dimatteo (2009) found evidence that patients' treatment adherence improved significantly more among patients whose physicians participated in communication skills training. 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).

Clinical Quality

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. Patients' overall ratings of their hospitals have been positively associated with hospitals' performance on CMS's process measures for pneumonia, congestive heart failure, AMI and surgical care in the US (Isaac, Zaslavsky, Cleary, & Landon, 2010), and to process indicators relating to 19 different conditions in the UK (Llanwarne, et al., 2013).

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).

Research suggests an association between better patient experiences and lower healthcare utilization. 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).

In surgery, the association between patient experience and quality of care has been investigated. Tsai and colleagues studied more than 2,900 hospitals between 2010 and 2011 by merging 100% Medicare inpatient claims data with the H-CAHPS survey data. Patients were included if they underwent coronary artery bypass grafting, pulmonary lobectomy, endovascular aortic aneurysm repair, open abdominal aortic aneurysm repair, colectomy, and hip replacement. They found that hospitals in the highest quartile of performance on patent satisfaction had length of stay that was on average 0.6 days shorter than those with the lowest patient satisfaction. This was also true for readmission rates and risk-adjusted perioperative mortality rates. More recently, Lobo Prabhu and colleagues examined the association of patient satisfaction and clinical outcomes using the American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP) framework. Using standard ACS NSQIP follow-up procedures, patients were provided a survey asking about their experiences. The investigators found a significant association between patient satisfaction and both 30-day readmission and the occurrence of postoperative surgical complications. They suggested further study is warranted to evaluate patient satisfaction as a healthcare quality indicator.

Since NQF-endorsement, the S-CAHPS Survey has demonstrated uptake on several fronts. First, three peer-reviewed publications have been published. Schmocker et al. used the S-CAHPS survey to determine which aspects of perioperative care were predictive of satisfaction with the surgeon. Their response rate was 45.3%. They found that on multivariable analysis, preoperative communication and attentiveness on the day of surgery were the most important determinants of overall surgeon rating. Lenherr et al. published the first experience and results of using S-CAHPS in urology. With a 33.8% response rate, their data suggested patient satisfaction with the surgeon is more influenced by

postoperative communication and information. Jiang and Malkin used Lean A3 thinking to analyze S-CAHPS survey data to identify quality improvement opportunities. Care processes in their postoperative clinic were modified and they found improvement in their S-CAHPS survey scores on the domains that they targeted.

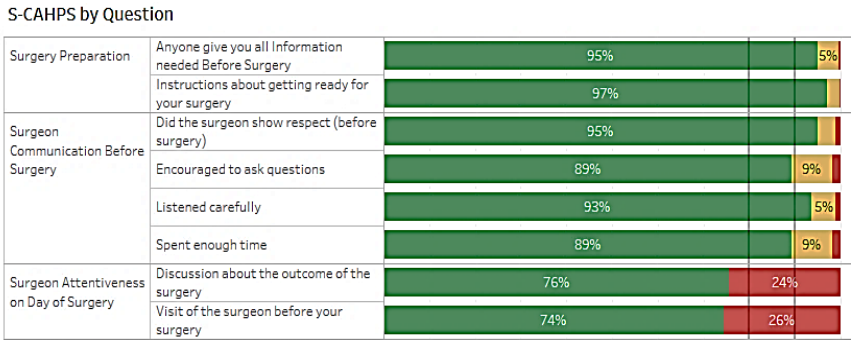


Figure A. Results of S-CAHPS survey administered to patients undergoing ambulatory surgery (n=951). Green = "Yes, definitely." Yellow = "Yes, somewhat." Red = "No."

Second, the S-CAHPS has also been utilized by an institution for internal quality improvement purposes when structuring its new ambulatory surgery center. At this institution, 25% (n=238/951) of ambulatory surgical patients reported using a web-based portal on the S-CAHPS that they were dissatisfied with their surgeon's attentiveness on the day of surgery (Figure A; publication in preparation). Patient throughput and flow management processes were reviewed, which identified an unexpected consequence of a highly efficient ambulatory surgery environment — surgeons were caring for the next patient in the operating room and unable to speak with their previous patient before discharge. Leveraging health information technology, a secure online video streaming solution was developed to facilitate surgeons in the operating room to interact face-to-face with their postoperative patient prior to discharge. The effect of this intervention on S-CAHPS scores is under active investigation.

Third, the American College of Surgeons has begun incorporating the S-CAHPS survey questions into the American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP) to improve quality from both surgeons' and patients' perspectives (i.e., tracking clinical outcomes and patient-reported outcomes). A pilot is currently underway, launched October 16, 2017, to evaluate the feasibility of collecting PROs into the registry – the S-CAHPS is being used for this purpose. The ACS NSQIP leadership intends to continue using S-CAHPS in the future and make it available to all hospitals participating in ACS NSQIP (more than 700 as of 2017) in the first quarter of 2018.

Fourth, pending CMS approval for the 2018 Merit-based Incentive Payment System (MIPS) program, participants in the ACS Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures set for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism. The Phases of Surgical Care Measure set is an episode-based measure framework which follows the five phases of surgical care: preoperative, perioperative, intraoperative, postoperative, and postdischarge. The measure set is composed of surgical outcome measures, high-value process measures, and the appropriate S-CAHPS measures which follow up on key processes within the measure set.

Fifth, the ACS is also currently working with CMS on an Advanced Alternative Payment Model that we anticipate may soon be tested by Center for Medicare and Medicaid Innovation (CMMI). This model incorporates a novel quality measurement framework which measures care around the patient for a given episode and incorporates patient reported experience and patient reported outcomes. As currently proposed, the surgical episodes include measures included in the S- CAHPS survey.

Citations

Jiang N. and B. Malkin. Use of Lean and CAHPS Surgical Care Survey to Improve Patients' Experiences with Surgical Care. Otolaryngol Head Neck Surg. 2016. 155(5): 743-7.

Lenherr SM, DeCicco B, Cameron AP, et al. The S-CAHPS Survey in Urology. Urol Prac. 2015. 2(1):12-6.

Liu JB, Pusic AL, Temple LK, Ko CY. Patient-reported outcomes in surgery: Listening to patients improves quality of care. Bull Am Coll Surg. 2017; 102(3): 19-23.

Lobo Prabhu K, Cleghorn MC, Elnahas A, et al. BMJ Qual Saf Published Online First. doi:10.1136/bmjqs-2017-007071

Schmocker RK, Cherney Stafford LM, Siy AB, et al. Understanding the Determinants of Patient Satisfaction with Surgical Care Using the CAHPS Surgical Care Survey (S-CAHPS). Surgery. 2015. 158(6):1724-33.

Tsai TC, Orav EJ, and AK Jha. Patient satisfaction and quality of surgical care in US hospitals. Ann Surg. 2015; 261(1):2-8

AHRQ CAHPS Quality Improvement Guide (2013). Accessible at https://cahps.ahrq.gov/quality-improvement/improvement-guide/improvement-guide.html.

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

Beattie M, Murphy DJ, Atherton I, et al. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Systematic Reviews*. 2015 4; 97.

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

Devine EC and Cook TD. A meta-analytic analysis of effects of psycho-educational interventions on length of postsurgical hospital stay. *Nurs Res* 1983;32(5): 267-74.

Devine EC, O'Connor FW, Cook TD, Wenk VA, Curtin TR. Clinical and financial effects of psycho-educational care provided by staff nurses to adult surgical patients in the post-DRG environment. *Am J Public Health.* 1988 Oct;78(10):1293-7.

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.

Fisher T and Gatewood H. (2011) Improving patient experience: A hands on guide for safety net clinics. California Healthcare Foundation. Accessible at http://www.chcf.org/publications/2011/10/patient-experience-safety-net-clinics

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.

Isaac, T., Zaslavsky, A. M., Cleary, P. D., & Landon, B. E. (2010). The relationship between patients' perception of care and measures of hospital quality and safety. *Health services research*, 45(4), 1024-1040.

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.

Llanwarne, N. R., Abel, G. A., Elliott, M. N., Paddison, C. A., Lyratzopoulos, G., Campbell, J. L., et al. (2013). Relationship between clinical quality and patient experience: analysis of data from the english quality and outcomes framework and the National GP Patient Survey. *Annals of family medicine*, 11(5), 467-472.

McWilliams JM, Landon BE, Chernew ME, Zaslavsky AM. (2014) Changes in Patients' Experiences in Medicare Accountable Care Organizations. *N Engl J Med* 371:1715-24.

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.

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.

Shaller, D. (2011) A Tale of Three Practices: How Medical Groups are Improving the Patient Experience. RWJF Aligning Forces for Quality Case Study. Accessible at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/quality-improvement/reports-and-case-studies/CaseStudyPatientExperience_July2011FINAL_revised20110728.pdf. .

Weinick RM, Quigley DD, et al. Use of CAHPS Patient Experience Surveys to Assess the Impact of

Health Care Innovations. *The Joint Commission Journal on Quality and Patient Safety*. September 2014. Volume 40(9): 418-427.

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.

All of the measures submitted to NQF for endorsement share the main objective of the S-CAHPS survey, which is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information. At the same time, consistent with the survey, the measures aim to directly benefit a variety of users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care.

The S-CAHPS survey measures patients' perceptions of their experiences with pre-, during and post-surgical care. The survey addresses issues relevant to surgical care, such as informed consent, communication, clear and helpful information, anesthesia care, office staff helpfulness, and post-operative follow-up.

To offer surgical patients and surgeons valid and reliable information on patient experience of care, the American College of Surgeons (ACS), in partnership with other surgical and anesthesia organizations, sponsored the development of the S-CAHPS survey. The S-CAHPS survey is a patient experience-of-care survey measure specifically tailored for surgical patients. The S-CAHPS survey was developed by working with patients to report on the full experience of surgical care, including their experience with the surgeon, the anesthesiologist, and the facility. Qualitative research (i.e., focus groups, cognitive testing, literature review) prior to field testing and focus groups following the main field test contributed to the development of the final composite measures. The data gathered through S-CAHPS survey data can assist consumers in identifying a high-quality surgeon and help surgeons to better understand and ultimately improve patient care.

To capture and ascertain the major domains surrounding the consumers' view of quality surgical care, the American Institutes for Research performed a literature review prior to developing the surgical CAHPS instrument. Using four literature databases: Medline, PsychInfo, CINAHL, and Evidence-Based Medicine Review, AIR reviewed 930 abstracts. These abstracts were narrowed from certain search terms and limitations. AIR asked the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP) members to provide input and guidance in these terms and limitations. The TAP included 21 members from various surgery and anesthesiology specialty societies, each having technical expertise in the topic of surgical care. The AIR research analyst then classified each abstract as relevant, possibly relevant, and not relevant.

From that point on, only the research analyst's "relevant" and "possibly relevant" classifications were reviewed by the project director. From this abstract review method, 37 abstracts were chosen to review in full. In addition, one article was added from an SQA Technical Advisory Panel (TAP) member. The 38 articles reviewed indicated 14 domains, or primary issues, of surgical patient care experience from a patient's perspective: information/education, interpersonal

manner, pain, emotional support, accessibility/convenience, technical quality of care, efficacy/outcomes of care, availability, environment, customization/personalized care, patient involvement in care, continuity of care, overall satisfaction, and finances. These results guided the development of the surgical survey based on relevance to quality and to consumers and the ability of consumers to act as reliable reporters of the domain.

In addition, American Institutes for Research conducted six focus groups to identify important quality issues inherent in patients' experiences of surgical care. The results of the focus groups are included in Attachment D Main 1c5 Surgical CAHPS Focus Groups_2nd Round_2010.pdf. In total, 49 people participated in the focus groups, all of whom were 18 years of age or older and had undergone a surgical procedure billable with a 90-day global fee within the last 7 months. The groups were diverse both demographically and in the type of surgery. For recruiting the additional two groups of patients with more complex surgeries, each participant had their surgery in a hospital and stayed overnight for at least one night.

After the focus groups were conducted, the AIR project director, senior research analyst, and research analyst examined the notes and focused in on each of the above issues. They gathered recurring themes and issues into a report of results, used to develop insights about what patients feel characterizes quality surgical care. The three main domains were surgeon's interpersonal skills and behaviors, surgeon's expertise/technical competence, and surgeon's skill in communicating or providing health information and patient education. The results of the focus groups and the surgical patients' experiences of care ultimately guided development of the surgical survey. Though patients rated technical skill of the surgeon as highly relevant to quality, it was determined that patients are not the best reporters of technical surgical expertise, so this was not included as a domain in the survey instrument. Additionally, based on the psychometric analysis and discussions within the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP), ACS did not recommend the Anesthesia Care items as a reporting composite. While ACS believes these items should remain in the survey because anesthesia care is an important factor in the surgical patient's experience of care, the Anesthesiologist works within the hospital system, and thus, the surgeon cannot adequately control the experiences between patient and anesthesiologist, and should not be measured on such.

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.*

Performance data with the statistics requested are provided in the attached Excel file named "Attachment B Main 1b2 S-CAHPS_score_Tables.xlsx." The summary statistics at the practice site level (mean, SD, min, max, deciles) can be found in the worksheet tab titled "Summary Stats".

Performance data from a pilot study of the S-CAHPS in two surgical subspecialties is shown below. The pilot study was an IRB-approved secondary analysis conducted of S-CAHPS data collected by Duke Otolaryngology Head and Neck Surgery (OHNS, via web-based electronic data capture system) and University of Michigan Urology (GU, via postal mail) from 2011-2013. A total of 2695 adult patients were administered S-CAHPS within 4 weeks of surgery (1424 OHNS, 1271 GU). Survey content was separated into 6 composite scores and analyzed by %-top box scoring for separate and pooled data. A total of 727 patients completed the survey (n=303, 21.3% OHNS and n=424, 33.8% GU). Full survey completion rate for all rated questions was 62% OHNS and 72% GU. Composite top-box scores were similar between OHNS and GU except for Communication Pre-Operatively. As shown in the Table below, pooled overall surgeon rating was high (88% top-box scores, ranked 9-10 out of 10). The overall surgeon rating was most correlated with surgeon communication pre- and post-operatively, followed by information to recover from surgery. Differences in both surgical subspecialty and mode of administration yielded similar responses to S-CAHPS, but mailed GU survey yielded a higher response rate.

S-CAHPS Composite Top-Box Results from Lenherr et al. study

Composite	% Top Box Pooled OHNS and GU (n=727)
Information to prepare for surgery	92%
Communication pre-operatively	92%

Surgeon's attentiveness day of surgery 84%

Information to recover from surgery 83%

Communication post-operatively 91%

Helpful, courteous and respectful staff 92%

Overall surgeon rating 88%

Citations:

Lenherr SM, DeCicco B., Cameron AP, Malaeb BS., Oldendorf AL, Stoffel JT, Karls EM, and Clemens JQ. The S-CAHPS Survey in Urology. (in press) Urology Practice. Vol 2. Available online at http://www.urologypracticejournal.com/article/S2352-0779(14)00109-5/pdf.

Schulz KA, Rhee JS, et al. (2012) Consumer Assessment of Healthcare Providers and Systems Surgical Care Surveys: Benefits and Challenges. Otolaryngology - Head and Neck Surgery. 147(4):671-7.

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.

The CAHPS Surgical Care Survey was implemented by the American Academy of Otolaryngology (AAO) – Head and Neck Surgery from August 2010 through March 2011. The S-CAHPS was administered in four sites with a total of 14 surgeons. Two sites were academic medical centers, and the other two sites were community practices. A total of 354 surveys were collected across the three sites. An overall response rate of 39.9% was achieved with individual site response rates of 37.7%, 35.7% and 47.9% respectively. As can be seen in the table from Schulz et al., minimum scores at the surgeon level were quiet low. The overall scores for Recovery Information and Surgeon Communication have much room for improvement.

Schmocker et al. used the S-CAHPS survey to determine which aspects of perioperative care were predictive of satisfaction with the surgeon. Their response rate was 45.3%. They found that on multivariable analysis, preoperative communication and attentiveness on the day of surgery were the most important determinants of overall surgeon rating. Lenherr et al. published the first experience and results of using S-CAHPS in urology. With a 33.8% response rate, their data suggested patient satisfaction with the surgeon is more influenced by postoperative communication and information. Jiang and Malkin used Lean A3 thinking to analyze S-CAHPS survey data to identify quality improvement opportunities. Care processes in their postoperative clinic were modified and they found improvement in their S-CAHPS survey scores on the domains that they targeted.

Below are several case studies of physician practices that have used the CAHPS survey to focus on and implement successful quality improvement interventions.

Five clinics in San Francisco improved the scores in communication after focusing on patients' most important concerns (Fisher and Gatewood, 2011). In Massachusetts, the Massachusetts General Physicians Organization implemented procedures for informing patients of waits, service recovery, physician communication and coaching, and staff huddles. These approaches coupled with education, training and recognition programs led to improved CG-CAHPS scores.

At the Dean Clinic in Wisconsin (over 800 medical staff in 60 locations), 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. As a result, Dean Clinic's overall performance on the "Helpful, Courteous, and Respectful Office Staff" composite measure increased from 79 percent in 2011 to 83 percent in 2013 (AHRQ CAHPS Website).

In July 2014, The Robert Wood Johnson Foundation updated their inventory of CAHPS quality improvement resources. This document offers tools to support health care organizations in determining what they need to do to improve patient experience and how to implement those improvements. These resources are available for both ambulatory care settings and hospitals and can be found here: http://forces4quality.org/af4q/download-document/6540/Resource-12-125_inventory_of_pat_exp_improvement_resources_-_designed_-_revised_11.3.pdf .

Providers routinely use patient experience measures such as CAHPS to guide quality improvement (QI) efforts (Friedberg et al, 2011; Davies, Shaller et al., 2013). Friedberg et al (2011) found that physician groups commonly targeted improvement at access, communication with patients, and customer service by addressing office workflow, providing additional training for nonclinical staff, and adopting or enhancing an electronic health record.

Citations:

AHRQ CAHPS Website: Quality Improvement Reports and Case Studies accessible at https://www.ahrq.gov/cahps/quality-improvement/reports-and-case-studies/Case-Study_QI-Initiatives.html

Fisher T and Gatewood H. Improving patient experience: A hands on guide for safety net clinics. California Healthcare Foundation. 2011. Accessible at http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20I/PDF%20ImprovingPatientExperienceHandsOnGuide.pdf

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-centred care: lessons from a quality improvement collaborative. Health Expect. Jun;11(2):160-76.

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.

Schulz KA, Rhee JS, et al. (2012) Consumer Assessement of Healthcare Providers and Systems Surgical Care Surveys: Benefits and Challenges. Otolaryngology – Head and Neck Surgery. 147(4):671-677.

Henderson A, Caplan G, Daniel A. Patient satisfaction: the Australian patient perspective. Aust Health Rev. 2004;27(1):73-83.

Hepner DL, Bader AM, Hurwitz S, Gustafson M, Tsen LC. Patient satisfaction with preoperative assessment in a preoperative assessment testing clinic. Anesth Analg. 2004 Apr;98(4):1099-105, table of contents.

Lledó R, Rodriguez-Ros T, Targarona EM, Trias M, Trilla A, Asenjo MA. Perceived quality of care of inguinal hernia repair: assessment before and after the procedure. Int Surg. 2000 Jan-Mar; 85(1):82-7.

Tabolli S, Molino N, Abeni D, Sampogna F. Satisfaction with care in vascular surgery inpatient units. Eur J Vasc Endovasc Surg. 2003 Jan;25(1):48-52.

Wolosin RJ. Patient satisfaction in gastroenterology clinics. Gastroenterol Nurs. 2003 Sep-Oct; 26(5):203-8.

Jiang N. and B. Malkin. Use of Lean and CAHPS Surgical Care Survey to Improve Patients' Experiences with Surgical Care. Otolaryngol Head Neck Surg. 2016. 155(5): 743-7.

Lenherr SM, DeCicco B, Cameron AP, et al. The S-CAHPS Survey in Urology. Urol Prac. 2015. 2(1):12-6.

Schmocker RK, Cherney Stafford LM, Siy AB, et al. Understanding the Determinants of Patient Satisfaction with Surgical Care Using the CAHPS Surgical Care Survey (S-CAHPS). Surgery. 2015. 158(6):1724-33.

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.*

Top box scores from the adult survey by gender, race and by age category are provided in worksheets "Sum Stats Ethnicity", "Sum Stats Age" and "Sum Stats Gender" in the attached Excel file named Attachment B Main 1b2 S-CAHPS_score_Tables.xlsx.

Differences in the top box scores by gender are small. There is also not an apparent difference in scores due to ethnicity.

Older patients generally reported more positive patient experiences. The group that gave the highest top box scores for global rating of surgeon were the youngest group of patients (18-24 years) and patients age 65-74 (93% and 89%, respectively). The scores from the older group is consistent with a study of Hospital CAHPS. In that study, O'Malley et al

(2005) found that younger age patients (18-24) scored significantly lower than patients 25-34 and patients in the eight age categories above 34 all scored higher than those 25-34.

Citations:

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.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

De.5. Subject/Topic Area *(check all the areas that apply):*

Surgery

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

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

Elderly

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/surgical/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)*

No data dictionary **Attachment:**

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

Attachment **Attachment:** surgical_eng-636461813392404132.pdf

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

Patient

S.3.1. 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.

Recommendation to include E-mail as part of Mixed-mode Survey Administration

Based on field test results, the CAHPS team recommends the following modes:

- Mail only.

- Telephone only.

- Mixed mode (mail and telephone, email and mail, or email and telephone).

The addition of e-mail administration (i.e., notification for web-based surveys) as a type of mixed-mode data collection (Drake et al., 2014; McInnes et al., 2012) is a recommendation since last endorsement. The CAHPS Consortium recommends including an option to conduct a mixed mode survey that would have two e-mail reminders and a follow-up by mail or telephone to all who are surveyed. The follow-up to the entire sampling frame is necessary to get a representative sample from a practice that is not based just on e-mail alone. The Consortium does not recommend a mailed hard copy letter with a link to a web survey.

While there are no explicit data on the use of electronic capture for the CAHPS Surgical Care Survey, the psychometric properties of surveys does not appear to change when transition from paper to electronic data capture (Bennett AV, et al. 2014).

Citation:

Bennett AV, Keenoy K, Shouery M, Basch E, Temple LK. Evaluation of mode equivalence of the MSKCC Bowel Function Instrument, LASA Quality of Life, and Subjective Significance Questionnaire items administered by Web, interactive voice response system (IVRS), and paper. Qual Life Res. 2016; 25:1123-30.

Drake KM, Hargraves JL, Lloyd S, Gallagher PM, Cleary PD. The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey. Health Serv Res. 2014 Aug;49(4):1387-99. doi: 10.1111/1475-6773.12160

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. 2012 Nov;50 Suppl:S11-9.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
<u>IF an OUTCOME MEASURE</u>, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

We recommend that S-CAHPS Survey items and composites be calculated using a top-box scoring method. The top box score refers to the percentage of patients whose responses indicated excellent performance for a given measure. This approach is a kind of categorical scoring because the emphasis is on the score for a specific category of responses.

The top box numerator for the Overall Rating of Surgeon 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 What´s Available for the CAHPS Surgical Care Survey: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/surgical/about/whats-available-surgical-care-survey.pdf

Also see Patient Experience Measures from the CAHPS Surgical Care Survey Document 409 obtained by going to: https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

Also, for more information on the calculation of reporting measures, see How to Report Results of the CAHPS Clinician & Group Survey, available at https://cahps.ahrq.gov/surveys-guidance/cg/cgkit/HowtoReportResultsofCGCAHPS080610FINAL.pdf.

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).

This section is used to describe the composite top box score. The composite top box score is the average proportion of respondents who answered the most positive response category across the questions in the composite.

The top box numerators for items within Composite measures 1, 2, 4, 5, and 6 is the number of respondents who answered “Yes, definitely" across the items in each composite. The top box composite score is the average proportion of respondents who answered “Yes, definitely" across the items in the composite.

The top box numerator for items within Composite measure 3 is the number of respondents who answered "Yes" across the items in this composite. The top box composite score is the average proportion of respondents who answered "Yes" across the items in this composite.

The top box numerator for the Measure 7, the Global Rating Item, is the number of respondents who answered 9 or 10 to the Global Rating Item.

EXAMPLE:

Given a composite with four items, where each item has three response options, a practice’s score for that composite is the proportion of responses (excluding missing data) in each response category.

The following steps show how those proportions are calculated:

Step 1 – Calculate the proportion of cases in each response category for the first question:

P11 = Proportion of respondents who answered “yes, definitely”

P12 = Proportion of respondents who answered “yes, somewhat”

P13 = Proportion of respondents who answered “no”

Follow the same steps for the second question:

P21 = Proportion of respondents who answered “yes, definitely”

P22 = Proportion of respondents who answered “yes, somewhat”

P23 = Proportion of respondents who answered “no”

Repeat the same procedure for each of the questions in the composite.

Step 2 – Combine responses from the questions to form the composite.

Calculate the average proportion responding to each category across the questions in the composite. For example, in the “How Well Surgeon Communicates With Patients Before Surgery” composite (four items), the calculations would be as follows:

Measure top box score = proportion who said “yes, definitely” = (P11 + P21 + P31 + P41) / 4

Example results: If P11 = 81% and P21=92% and P31 = 84% and P41 = 95% then the top box score = (81% + 92% + 84% = 95%) / 4 = 88%.

Also see Patient Experience Measures from the CAHPS Surgical Care Survey Document 409 obtained by going to: https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

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 population for the survey is adult patients (age 18 and over) who had a major surgery as defined by Common Procedural Terminology (CPT) codes (90 day globals) within 3 to 6 months prior to the start of the survey.

Results will typically be compiled over a 12-month period.

For more information on the calculation of reporting measures, see Patient Experience Measures from the CAHPS Surgical Care Survey, available at https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html.

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

The survey is sampled at the ambulatory care level. However, there are questions that ask about care received at the hospital or surgical care facility.

The major criterion for selecting patients is having surgery, as defined by Medicare 90-day global surgery codes within 3 to 6 months prior to the start of the survey. Since post-surgical care was an important component of the survey, surveys could not be appropriately administered until an adequate time for experiencing post-surgical care (3 months) had passed. The time frame for the surgery was selected to (1) minimize recall bias and (2) ensure ample time was allowed for follow-up care after surgery. The survey is not administered more than 6 months post-surgery because of concerns about recall bias.

Patients have to be adults and non-institutionalized. Included surgeries should be scheduled and not an emergency procedure. This is because an important component of the survey deals with pre-surgical office visits – a topic which would not be relevant for most emergency surgeries.

The Survey's denominator code table lists 90-day global CPT codes for major surgery, representing over 10,000 possible codes across multiple surgical specialties. The Surgical Quality Alliance felt that specifying only Medicare's 90-day global procedure codes would include appropriate procedures while excluding minor procedures that were not intended to be included.

The attached excel file named "Attachment A Main S7 CY2015-90-day-global codes.xlsx" includes the CPT codes that are currently used to identify the S-CAHPS survey's target population of patient with major surgery (i.e., measure denominator).

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

The following are excluded when constructing the sampling frame:

- Surgical patients whose procedure was greater than 6 months or less than 3 months prior to the start of the survey.

- Surgical patients younger than 18 years old.

- Surgical patients who are institutionalized (put in the care of a specialized institution) or deceased.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

The following patients would be excluded from the measure's denominator:

- Survey users and vendors should exclude surveys where the respondent reports he or she has not had surgery performed on the date listed by the surgeon named. (First question of survey.)

- Surgical patients that had an emergency surgical procedure since emergency procedures are unlikely to have visits with the surgeon before the surgery.

- Individuals from a household that has already been sampled.

- Respondents who did NOT answer at least one item of the measure are NOT included in the denominator.

Instructions on how to transform raw data from a CAHPS survey into data that the CAHPS Analysis Program can use can be found in Preparing and Analyzing Data from the CAHPS Clinician & Group Surveys available at https://www.cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1035_Preparing_analyzing_data_from_cg.pdf

Survey code specifications --- including how to code an appropriately skipped item, multiple marks or blank items --- can be found in the Instructions for Analyzing Data available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf.

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.)*

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

See pages 18-19 of the Instructions for Analyzing Data available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf.

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

Other

If other: Case-mix adjustment

S.12. Type of score:

Other (specify):

If other: Top-box Score; case-mix adjusted 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 a non-emergency surgery within 3 to 6 months prior to the start of the survey.

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 the patient had surgery performed on the date listed by the surgeon named.

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.

Note that for users who want to case-mix adjust their scores, case-mix adjustment can be done using the CAHPS macro and the adjustment is made prior to the calculation of the total score.

Case-mix Adjusted Scores

Case-mix adjustment is done via linear regression. The CAHPS Consortium recommends self-reported overall health, age, and education as adjusters. These items are printed in the "About You" section of the survey, questions 38-45.

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://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf.

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.

The following sampling guidelines are provided to users as part of the "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" document available at

https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1033_CG_Fielding_the_Survey.pdf.

Defining the Sample Frame: Eligibility Guidelines

The sample will be drawn from a list of individuals (adults age 18 and older) who had a planned surgery from the surgeon 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 decision will depend on which organization has the most accurate and complete data. Health plans or purchasers of care may have administrative or billing data to identify individual patients. In some instances, the data to identify individual patients may be found only in the records of medical practices. It may be necessary to pull data from two or more sources in order to have both up-to-date contact information and to be able to connect the procedure to a specific surgeon.

Users should review these guidelines for determining who to include in the sample frame:

- Include only adult patients (age 18 and over) that had a major surgery as defined by CPT codes (90 day globals) within the last 3 to 6 months. This time frame is also known as the look back period.

- The sampling frame is a person-level list and not a procedure list. Therefore, patients should appear only once in the sampling frame regardless of how many surgeries they have had in the look back period. Use the patient's most recent surgery for inclusion in the sampling frame.

- Draw the sample irrespective of reason for surgery 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.

- 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, 2011, include all those who have had surgery during February 1 – May 31.

- Allow the sample frame to include multiple individuals from the same household, but the sample you draw should not have more than one adult 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.

- To administer the survey, the name of the surgeon must be available, even if you are surveying at the practice level. If the sampling frame does not accurately identify the surgeon who performed the surgery, you may want to select a larger sample to account for errors in connecting health care received to a specific provider.

Recommended Number of Completes

In order to determine the size of the sample, you first need to determine the level of sampling and how many completed surveys are required to obtain usable information at that level. The CAHPS Clinician & Group Survey and the CAHPS Surgical Care Survey 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.

- Individual providers: 45 completed surveys per provider. For applications of the survey intended to report or assess performance for individual providers, the Consortium recommends at least 45 completed surveys per provider.

- Practice site: The recommended number of completed surveys is based on the number of surgeons at the site. The site-level sample size recommendations vary by the number of surgeons per location site. The CAHPS Consortium has issued the following recommended sample sizes for collecting CAHPS Clinician and Group (CG-CAHPS) data at the practice site level: 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. 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.

The minimum sample size recommendations are based on extensive research conducted on the CAHPS Clinician & Group Survey. These recommendations are based on data regarding the number of completed questionnaires necessary to achieve adequate physician-level reliability for a measure. That is, how many completed surveys does one need to reliably distinguish among different physicians? To answer this question, CAHPS investigators examined data from multiple field trials.

Sample Size Calculation: To have a sufficient number of responses for analysis and reporting, survey users need to select enough individuals to obtain approximately 45 completed questionnaires per physician. Assuming you achieve the recommended response rate of 40 percent, survey users would need to start with a minimum sample size of 113.

More detail and reasoning behind the recommendations can be found in the 2011 document titled "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1033_CG_Fielding_the_Survey.pdf.

Citation:

Dyer, N., J. S. Sorra, et al. (2012). "Psychometric properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Adult Visit Survey." Medical care 50 Suppl: S28-34.

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.

The measures are dependent on the standardized implementation of the complete CAHPS Surgical Care Survey. Failure to administer the entire survey will compromise the validity and reliability of the measures.

Developing the Sampling Frame

The sampling frame for the CAHPS Surgical Care Survey consists of patients that have had a non-emergency 90-day global procedure in the last 3-6 months from the date that the survey will be administered. The sampling frame is a patient level file – patients with multiple procedures that meet the criteria may not be included multiple times in the sampling. The actual sample will be de-duplicated to ensure that patients are not provided more than one survey. Users of S-CAHPS should follow the recommended guidelines of the CAHPS Clinician and Group Survey found in the "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" document available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1033_CG_Fielding_the_Survey.pdf.

Preparing Sample Files for Data Collection

Once the sample has been selected, 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.

As previously noted, 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 from the plan and contact records from the 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). When you have incomplete address information or reason to believe that this information may be inaccurate, sponsors and/or vendors may be able to use other sources, such as CD-ROM directories, Internet sources, or directory assistance, to clean the sample file.

Data Collection Modes

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 CAHPS team recommends the following modes:

- Mail only.

- Telephone only.

- Mixed mode (mail and telephone, email and mail, or email and telephone).

Survey sponsors that employ one of these modes using the recommended protocols can expect to achieve response rates of approximately 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 followup method is most effective; results from survey research literature indicate that followup by telephone often adds 10 to 15 percentage points to the response rate. A sample telephone script for Surgical CAHPS in both English and Spanish is available at https://www.cahps.ahrq.gov/surveys-guidance/survey4.0-docs/sample_tele_script_surgical_care_survey.pdf.

Note that the addition of e-mail administration (i..e, notification for web-based surveys) as a type of mixed-mode data collection is a recommendation since last endorsement. The CAHPS Consortium recommends including an option to conduct a mixed mode survey that would have two e-mail reminders and a follow up by mail or telephone to all who are surveyed. The follow up to the entire sampling frame is necessary to get a representative sample from a practice that is not based just on e-mail alone. The Consortium does not recommend a mailed hard copy letter with a link to a web survey.

Other Modes

The CAHPS team recognizes that many organizations may already be or are interested in conducting patient surveys using different modes of survey administration and has conducted preliminary testing of other modes, specifically in-office distribution and interactive voice response (IVR, also known as telephone audio computer-assisted self-interviewing, or T-ACASI). Further study is required before either of these modes can be recommended.

A study of in-office distribution found that the survey results were not comparable to those collected with recommended modes. The investigators observed incomplete distribution rates, lower response rates, and declining distribution rates. Finally, there were significant mode-physician interaction effects, which suggests that data cannot be pooled then adjusted to account for the differences. Because the implications of using these modes are not yet fully known, they should be used with caution. If a sponsor uses one of these modes to collect data, the ability to compare survey results across sponsors may be limited.

The Consortium's support of the use of multiple modes of survey administration is intended to minimize disruption to organization's 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. More detail on the protocols per mode, can be found in the full fielding guide. The Consortium is currently collecting and analyzing data in order to assess the need for procedures for data adjustments as a function of each survey mode and to enable the team to develop such procedures.

Tracking Returned Questionnaires

Most vendors have established methods for tracking the sample. You should also set 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.

Citation:

Drake KM, Hargraves JL, Lloyd S, Gallagher PM, Cleary PD. The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey. Health Serv Res. 2014 Aug;49(4):1387-99. doi: 10.1111/1475-6773.12160

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. 2012 Nov;50 Suppl:S11-9.

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.

Consumer Assessment of Health Providers and Systems (CAHPS) Surgical Care Survey Version 2.0

Available in English at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/surgical_eng.pdf

Available in Spanish at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/surgical_span.pdf

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)*

Inpatient/Hospital, Other, Outpatient Services

If other: Hospital Outpatient Surgery Center, Ambulatory Surgical Centers

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

2. Validity – See attached Measure Testing Submission Form

Attachments_1741_11062017-636461679792662086.zip,nqf_testing_attachment_7.1_SCAHPS_FINAL2.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the

most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.2 For maintenance of endorsement

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

No

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Yes - Updated information is included

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

Measure Number (*if previously endorsed*): 1741
Measure Title: CAHPS Surgical Care Survey 2.0
Date of Submission: 11/11/2017

Type of Measure:

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

Instructions

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

- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful [16] differences in performance;

OR

there is evidence of overall less-than-optimal performance.

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

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

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face

validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

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

*Metrics presented throughout are derived from analysis of the CAHPS Surgical Care Survey, Adult Version 2.0 core items.

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The main field test was conducted in the summer of 2008 and included 33 surgical practices across 9 surgical specialties. There were four mailings sent to patients in two waves, or batches. The response rate was 49%. One site only had 5 responses and is excluded from this dataset for testing. Forty-six surveys that answered "NO" to the first question "Did you have surgery from this surgeon on specified date?" were excluded. The final dataset has 2,719 survey results from 32 practices across nine specialty types.

1.3. What are the dates of the data used in testing? February 2007 to April 2008

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*)

The measured entity is referred to as a "practice site." Surgeons in a single practice site should share administrative and clinical support staff. Practice site level survey results are calculated across the respondents within a specific site.

The data for CAHPS Surgical Care (S-CAHPS) Version 2.0 includes survey response data for surgeries that occurred between February 2007 – April 2008. Data are being pooled from investigators that have utilized the S-CAHPS survey since 2008, such that a meta-analysis can be conducted. We anticipate submitting results to the NQF during a future measure update period.

The data in the testing and analysis includes 32 practice sites and 2,719 respondents across nine specialty types. The average number of respondents per site is 85 (standard deviation = 75) ranging from 19 individuals per site to 298 individuals per site. The median number of respondents per site is 47.5. Three of the respondents took the survey using the Spanish version. The total number of unique surgeons in the dataset was 72 resulting in an average of 2.25 surgeons per practice. Table 1.5a below shows the distribution of states within the dataset. Table 1.5b below shows the distribution of specialty type within the dataset.

Tables 1.5a Distribution of States in CAHPS Surgical Care Version 2.0 Dataset

Count of Complete Records & Practice Sites within State		
STATE	Total Complete Records	Total Practices
Arizona	218	1
California	49	1
Florida	82	1
Georgia	45	1
Illinois	61	2
Indiana	117	1
Kansas	50	1
Maryland	67	2
Massachusetts	122	2
Michigan	683	5
Mississippi	35	1
New Jersey	64	2
New York	19	1
North Carolina	217	1
Ohio	517	4
Pennsylvania	143	1
Texas	186	4
Utah	44	1
Total	*2,719*	*32*

Tables 1.5b Distribution of Specialty Type in CAHPS Surgical Care Version 2.0 Dataset

Count of Complete Records & Practice Sites within Specialty Type		
Specialty Type	Total Complete Records Within Specialty	Total Practices
Colon and Rectal	326	5
General Surgery	396	6
Ophthalmology	405	4
Orthopaedic	415	2
Otolaryngology	294	4
Thoracic	235	3
Urology	534	5
Vascular	114	3
Total	*2,719*	*32*

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 non-institutionalized adult patients that had a major scheduled surgery as defined by CPT codes (90 day globals) within 3 to 6 months prior to the start of that practice's survey. The attached excel file named "Table Testing 1.6 – S-CAHPS CPT Codes.xlsx" includes the CPT codes that were used during the main field test to identify the S-CAHPS survey's target population of patients with major surgery (i.e., measure denominator).

Although multiple individuals in a single clinician group may be on the sampling frame, the final sample contained only one respondent per household. Where a duplicate household was sampled, it was 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. Table 1.6 shows descriptive characteristics of the individuals surveyed included in our analyses. Practice sites had adult patients that were predominantly white (86%) and 55 years or older (71%). Eighty-three percent of the surveys were collected through the mail.

Table 1.6. Descriptive Characteristics for S-CAHPS 2.0 Dataset (32 Practices, 2,719 Respondents)

2008 Field Test	
	N and (Percent of Total)
GENDER	
Female	1355 (49.8%)
Male	1261 (46.4%)
Missing	101 (3.8%)
RACE	
White	2346 (86.3%)
Black or African American	156 (5.7%)
Asian	25 (0.9%)
Hawaiian or Pacific Islander	2 (0.1%)
American Indian or Alaskan	10 (0.4%)
Other	50 (1.8%)
Multi-racial	29 (1.1%)
Missing	101 (3.7%)
AGE CATEGORY	
18-24 Years	38 (1.4%)
25-34 Years	100 (3.7%)
35-44 Years	207 (7.6%)
45-54 Years	362 (13.3%)
55-64 Years	600 (22.1%)
65-74 Years	701 (25.8%)
75 Years or older	623 (22.9%)
Missing	88 (3.2%)
HOW SURVEY COLLECTED	
Mail	2253 (83%)
Web	466 (17%)

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?

Items collected in the survey that could be considered social risk factors are Education and Race/Ethnicity. They are not analyzed.

2a2. RELIABILITY TESTING

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

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

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

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

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Cronbach's alpha is the most common measure of internal consistency reliability and commonly used for surveys with scale-type questions. We calculated a Cronbach's alpha for each composite to assess the extent to which respondents consistently answered the items, with a reliability of at least 0.70 considered acceptable (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.

Given that individual responses are nested within practice sites, we measure site reliability on multi-item composite top-box scores and global one-item top-box scores, which partition within- and between-site variance. For this test, we used case-mix adjusted scores generated by following the specifications from the CAHPS Analysis Program Version 4.1 for adjusting on patients' general health rating, education and age. Similar to internal consistency reliability (i.e., Cronbach's alpha), values of 0.70 and higher are considered acceptable for site reliability (Nunnally and Bernstein, 1994; CAHPS Analysis Program, 2017; Instructions for Analyzing Data from CAHPS Surveys, 2017) .

The individual site reliability was calculated using the following formula and then averaged across the practice sites to derive measure-level reliability:

$$\text{Reliability}_g = \frac{\Sigma_B}{\Sigma_B + \frac{\Sigma_W}{N_g}},$$

where $\sum_B$ refers to the between-group variance; $\sum_W$ refers to the within-group variance, and Ng is the sample size for site g (Raudenbush and Bryk, 2002).

Citations:

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

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

The CAHPS Analysis SAS Program Version 4.1c is downloadable from.

https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

Instructions for Analyzing Data from CAHPS® Surveys is accessible at https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html

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*)

Table 2a2.3a. Cronbach's Alpha Individual-level Reliability Coefficient for S-CAHPS Version 2.0 Dataset. 2008 (32 practice sites, 2,719 Respondents)

Measure and Items	Standardized Cronbach's Alpha	Cronbach's Alpha if Item Deleted*
7. Information to Help You Prepare for Surgery (2 items)	0.74	
8. How Well Surgeon Communicates with Patients Before Surgery	**0.82**	
Listened carefully to them (Q9).		0.74
Spent enough time with them (Q10).		0.77
Encouraged them to ask questions (Q11).		0.78
Showed respect for what they had to say (Q12).		0.81
9. Surgeon Attentiveness on Day of Surgery (2 items)	**0.66**	
10. **Information to Help You Recover from Surgery - from Surgeon or another Health Provider from the office.**	**0.84**	
Explained what to expect during recovery (Q26).		0.77
Gave easy to understand instructions about what to do during recovery (Q28).		0.79
Warned them about symptoms that need immediate attention (Q27).		0.77
Made sure they were physically comfortable or had enough pain relief after leaving the hospital or surgery facility (Q29).		0.85
11. How Well Surgeon Communicates With Patients After Surgery	**0.86**	
Listened carefully to them (Q31).		0.81
Spent enough time with them (Q32).		0.82
Encouraged them to ask questions (Q33).		0.81
Showed respect for what they had to say (Q34)		0.86
12. Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	**0.85**	

* Not applicable if less than three items per composite.

Table 2a2.3c. Top Box Practice Site-Level Reliability Statistics for Surgical CAHPS, Version 2.0 (32 Practices)

Measures	Average # of Respondents per site	Site-Level Reliability
Information to Help You Prepare for Surgery	85	0.52
How Well Surgeon Communicates with Patients Before Surgery	76	0.68
Surgeon Attentiveness on Day of Surgery	83	0.50
Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.	85	0.71
How Well Surgeon Communicates With Patients After Surgery	72	0.48
Helpful, Courteous, and Respectful Staff at Surgeon's Office	84	0.71
One-item Global Rating of surgeon	82	0.60

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?*)

Table 2a2.3a shows the Cronbach's alpha for each composite. These individual-level test results show that each composite of the S-CAHPS survey has an acceptable level of reliability. 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 the alpha by dropping an item. The results do not suggest removal of any questions.

Table 2a2.3b shows the mean number of respondents per practice site and practice site-level reliability statistics for the surveys. The dataset excludes sites with less than 10 surveys. The majority of the composites and global ratings exhibit satisfactory site-level reliability with two measures above 0.70 and two between 0.6 and 0.7. Information to Prepare for Surgery and Post-Surgery Communication exhibit lower reliability at the site level than they did at the individual level. It should be noted that the site-level reliability analysis was conducted on field test data consisting of a relatively small selection of surgeon practices which may have led to reduced variability between sites.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

At the practice site level, we examined the relationships between each composite's top box score and the top box score for the global measure of "How would you rate your surgeon?" using Spearman rank-order correlations to determine the validity of the composite measures. For example, the composite measuring how well patients' surgeon communicates with before surgery is expected to be strongly related to the patients' overall rating of their surgeon. Finding such a relationship supports interpretation of the composite as a valid measure of patient experience with a surgeon and that surgeon's office. Specifically, Table 2b1.3b shows validity testing results of the instrument tested across all responses. Tables 2b1.3a and 2b1.3c reflect practice-level scoring validity.

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.

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

Table 2b1.3a. Practice Site-Level Correlation of Composites and Global Rating for S-CAHPS Version 2.0 Sample, 2008 (32 practice sites)

Rating of Surgeon	Info Pre-Surgery	Communicate Pre-Surgery	Attentiveness on Day of Surgery	Recovery Information	Communicate Post-Surgery	Office Staff Service
GLOBAL	0.48	0.59	0.02*	0.68	0.50	0.49

P < 0.01 for all values. *P>0.05.

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

Table 2b1.3b. Individual-Level Correlation of Composites and Global Rating for S-CAHPS Version 2.0 Sample, 2008 (2,719 respondents)

Rating of Surgeon	Info Pre-Surgery	Communicate Pre-Surgery	Attentiveness on Day of Surgery	Recovery Information	Communicate Post-Surgery	Office Staff Service
GLOBAL	0.42	0.49	0.24	0.46	0.47	0.34

P < 0.001 for all values.

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

Table 2b1.3c. Site-Level Top-Box Composite Intercorrelations for CAHPS Surgical Care Version 2.0 Sample, 2008 (32 Practices)

Composites	Info Pre-Surgery	Communicate Pre_Surgery	Attentiveness on Day of Surgery	Recovery Information	Communicate Post-Surgery	Office Staff Service
Info Pre-Surgery	1	0.58***	0.30	0.69***	0.64***	0.54**
Communicate Pre_Surgery		1	0.49**	0.71***	0.80***	0.31
Attentiveness on Day of Surgery			1	0.33	0.46**	-0.07
Recovery Information				1	0.64***	0.55**
Communicate Post-Surgery					1	0.36
Office Staff Service						1

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

Note: Values are Spearman rank-order correlations.

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 survey results show that each of the six composites are related to the global rating scale at the individual level and five of the six composites are related to the global rating at the practice level.. The two Communication composites and the Recovery Information composite have the strongest relationship with the global rating of surgeon.

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). In general, relationships among the composites met our expectations.

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 _risk factors**

☐ **Stratification by _risk categories**

☒ **Other,** Optional risk adjustment*

*The CAHPS analysis program allows users to select adjustment factors of their choosing.

The S-CAHPS measures, like the CG-CAHPS measures (NQF #0005), can be reported as either top-box scores or as case-mix adjusted scores depending on the user's purposes. Note that case-mix adjustment is possible in certain situations and is optional.

Top Box Score Calculation:

1) Target Population: Patients that had at least one visit during the past 12-months

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 12 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 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

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.

Not applicable.

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.

We are not putting forth a risk-adjusted measure. Adjustment is not a necessity because CAHPS surveys are not a clinical outcome or resource use. However, the S-CAHPS measures, like the CG-CAHPS measures (NQF #0005), does give users the option to case-mix adjust scores depending on their specific purposes.

Note: We did use case-mix adjusted scores for reliability testing, adjusting on patients' general health rating, education and age using the CAHPS Macro.

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 recommended to users and the case-mix adjustment that is performed in parallel to the methods used for all CAHPS measures. Specifically, case-mix adjustment methods for the Clinician and Group CAHPS (CG-CAHPS) are applicable to the S-CAHPS.

Case-mix adjustment takes into account respondent characteristics, such as age or educational attainment, which may affect the reports and ratings of care but are unrelated to differences in care quality. In other words, it is important to account for patient characteristics that are not under the control of the group but are related to the patient's experiences and survey responses. For example, several studies have found that younger and more educated patients provide less positive evaluations of health care (Elliott et al. 2001; Zaslavsky et al. 2001). CAHPS data can also be adjusted for other factors such as survey administration mode. Without an adjustment, differences in CAHPS scores between entities could be due to case-mix differences rather than true differences in quality. CAHPS survey results can be case-mix adjusted by users of the CAHPS analysis program.

Each user's project team must determine if it is appropriate to adjust its data to account for case-mix differences and if so, which adjusters to use. Specifically, for the S-CAHPS, available adjusters include overall health, overall mental or emotional health, age, sex, number of previous surgeries, educational attainment, Hispanic ethnicity and race. The project team must also decide whether or not to impute missing data for the adjusters at each adjuster's entity-level mean. The document "Instructions for Analyzing Data from CAHPS Surveys" dated April 2012 (available at: https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html) contains instructions for coding these variables and for including them in analyses using the CAHPS Analysis Program in SAS. Notwithstanding, the CAHPS Consortium only recommends that users use the CAHPS macro to adjust their survey data for respondent age, education, and general health status if they are comparing scores across practices that may differ in the characteristics of patients. The only characteristics that may be considered social that is collected in the survey are Education level and "Race" and "Hispanic or Latino origin/descent."

One of the methodological issues associated with comparison across practice sites and/or individual clinicians is the need to adjust appropriately for differences in case-mix, particularly in situations where there are differences among sites in patient characteristics. Therefore, current CG-CAHPS guidance suggests health status, age, and education as possible case-mix adjusters for users. These patient characteristics are not under the control of the provider. Studies have found that patient health status, education and age are predictors of CAHPS' scores (Kim et al, 2005; O'Malley, Zaslavsky et al. 2005; Elliott, Zaslavsky et al. 2009). A recent study by Drake and colleagues (2014) found that telephone respondents gave more positive responses than mail respondents.

Citations:

CAHPS Clinician & Group Survey. Content last reviewed August 2017. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/cahps/surveys-guidance/cg/index.html

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.

Elliott MN, Swartz R, Adams J, Spritzer KL, Hays RD. Case-mix adjustment of the National CAHPS benchmarking data 1.0: a violation of model assumptions? *Health Serv Res*. 2001 Jul; 36(3):555-73.

Kim M, Zaslavsky AM, Cleary PD. (2005) Adjusting Pediatric Consumer Assessment of Health Plans Study (CAHPS) Scores to Ensure Fair Comparison of Health Plan Performances. *Med Care*. Jan;43(1):44-52.

O'Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L, Cleary PD. (2005) Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv Res.* Dec;40(6 Pt 2):2162-81.

Zaslavsky AM, Zaborski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD. Adjusting Performance Measures to Ensure Equitable Plan Comparisons. *Health Care Financ Rev*. 2001 Spring; 22(3):109-126.

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)**

See 2b3.3a above.

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

See above.

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.**

See above.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable.

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*)
Not applicable.

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)*

Statistical significance is determined based on case-mix adjusted mean scores for individual items, composites and global ratings, and is used to determine whether a practice site mean is statistically different from the mean results for all practice sites. The statistical test used is a t-test of means, with $p<.05$ used as the criterion for determining significance. When there are large sample sizes, relatively small differences between practice sites may be statistically significant.

The CAHPS analysis program allows users to perform testing for both statistical and substantive significance. 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.

More information about Surgical CAHPS can be found at https://www.ahrq.gov/cahps/surveys-guidance/surgical/about/index.html.

More information about how to analyze practice level CAHPS scores can be found on this website: https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html.

For guidance in adapting these analysis instructions, users can contact the CAHPS User Network by e-mail (mailto:mcahps1@ahrq.gov) or telephone (1-800-492-9261).

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*)

T**able 2b5.2. Variability of S-CAHPS Measures' Scores**

	Descriptive at the Site Level (32 Practices)				
Measure	**Mean Top Box**	**SD**	**Min-Max Range**	**Min**	**Max**
Information to Help you Prepare for Surgery (2 items)	90%	0.05	19	79%	98%
How Well Surgeon Communicates Before Surgery (4 items)	85%	0.07	31	67%	98%
Attentiveness on Day of Surgery (2 items)	81%	0.12	55	42%	97%
Information to help you Recover from Surgery (4 items)	82%	0.07	36	64%	100%
How Well Surgeon Communicates After Surgery (4 items)	84%	0.06	24	73%	97%
Helpful, Courteous, and Respectful Staff (2 items)	87%	0.08	42	58%	100%
One-item Global Rating of Surgeon	86%	0.07	28	70%	98%

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The field test data suggests large differences across the 32 practices in surgical performance across many of the topic areas addressed by the survey. For example, with the Recovery Information measure, there was a 36-percentage point difference between the lowest scoring practice (64%) and the highest scoring practice (100%). Each of these are meaningfully different than the mean of 82%.

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

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

Section not applicable – one set of specifications.

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Survey Non-Response

For CAHPS, the recommended or target response rate is 40 percent. Survey users that follow the recommended protocols for sampling and data collection, including followup with nonrespondents, typically report response rates of 40 percent or higher. It is also important to begin with as accurate a sampling frame as possible. This section presents the guidance provided to users for calculating response rates (AHRQ, 2011).

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

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

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

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

Numerator Inclusions:

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

Denominator Inclusions:

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 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 is presumed to be eligible but was never located.

Denominator Exclusions:

- **Deceased**. In some cases, a household or family member may inform you of the death of the sampled individual.
- **Ineligible.** The sampled individual did not have a scheduled surgery from surgeon during the 3 to 6 months prior to the survey.

Users are provided the following advice for improving response rates:

- Improve initial contact rates by making sure that addresses and phone numbers are current and accurate (e.g., identify sources of up-to-date sample information, run a sample file through a national change-of-address database, send a sample to a phone number look-up vendor).
- Use all available tracking methods (e.g., directory assistance, 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 both a mail and telephone data collection procedure. In field tests, the combined approach was more likely to achieve a desired response rate than did either mode alone.
- Train interviewers on how to deal with gatekeepers (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, discussed in sections S4-S11 of the main NQF submission form, maximizes the use of available data by averaging available individual-level responses in construction of an overall score for the practice site. For each individual item, the top box score is percentage of respondents who answered the most positive response for the item. At the site-level, the top box composite score is the average of those percentages across the items. Most often, as a result of this averaging methodology, the percent of missing items at the site level is very small (i.e., < 1%).

We provide the percentage of cases with missing values at the item level below.

Citation:

AHRQ "Fielding the CAHPS® Clinician & Group Surveys. Sampling Guidelines and Protocols. "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.

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 2b7.2a. Surgical CAHPS Version 2.0: Item-level Percent Missing (2,719 respondents)

Survey Item	% Missing at Individual Level	Item is after a Skip Question
1. Information to Help You Prepare for Surgery	**0%**	
Gave all the information needed before surgery (Q3).	1%	Yes
Gave easy to understand instructions (Q4).	1%	Yes
2. How Well Surgeon Communicates with Patients Before Surgery	**11%**	
Listened carefully to them (Q9).	2%	Yes
Spent enough time with them (Q10).	2%	Yes
Encouraged them to ask questions (Q11).	2%	Yes
Showed respect for what they had to say (Q12.)	2%	Yes
3. Attentiveness on Day of Surgeon	**1%**	
Visited them before surgery (Q15.)	3%	No
Discussed the outcome of their surgery (Q17.)	2%	No
4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.	**1%**	
Explained what to expect during recovery (Q26).	2%	No
Gave easy to understand instructions about what to do during recovery (Q28).	2%	No
Warned them about symptoms that need immediate attention (Q27).	2%	No
Made sure they were physically comfortable or had enough pain relief after leaving the hospital or surgery facility (Q29).	2%	No
5. How Well Surgeon Communicates With Patients After Surgery	**16%**	
Listened carefully to them (Q36).	4%	Yes
Spent enough time with them (Q37).	4%	Yes
Encouraged them to ask questions (Q38).	4%	Yes
Showed respect for what they had to say (Q34)	3%	yes
6. Helpful, Courteous, and Respectful Staff at Surgeon's Office	**2%**	
Clerks and receptionists were helpful to patients (Q36).	2%	No
Clerks and receptionists treat patients with courtesy and respect (Q37)	3%	No
7. Single-Item Global Rating of Surgeon (Q34)	**4%**	**No**

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*)

Table 2b7.2a. shows the percentage of cases with missing data on each item and the last column denotes if the item followed a screen/skip item. Less than five percent of cases are missing response scores on the individual items, which suggests that our item-level results are likely not biased by systematic missing data due to item nonresponse.

It is important to note that ten of the items in S-CAHPS were not applicable for patients who were directed to skip the question due to their response to a screen/skip question. For example, the second question of the survey asks, "Before your surgery, how many office visits did you have with this surgeon?" If the taker checks NONE, then they are guided to skip the two questions about information to prepare for surgery (composite measure #1) and the four questions about pre-surgery communication (composite measure #2).

The post- surgical communication items also follow a screen question. The screener item asks: "After your surgery, did you talk with this surgeon by phone or visit the surgeon at his or her office?" If the response is "NO" then the survey taker skips the four questions about post-surgical communication with the surgeon.

Screening questions can result in a high percentage of missing due to appropriate skips. 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)

Citation:

Rodriguez HP, Glahn Tv, Li A, Rogers WH, Safran DG. The effect of item screeners on the quality of patient survey data: a randomized experiment of ambulatory care experience measures. *Patient*. 2009 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 surgeon´s patients

3b. Electronic Sources

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

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

No data elements are in defined fields in electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Though multiple modes of data collection, as well as mixed mode types of administration, are possible and have been tested, CAHPS surveys are primarily delivered via mail. Electronic databases are created after mailed surveys are returned. 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. However, email has been added as a mixed mode strategy for surgeon groups with reliable email addresses for all of their population. This is important as the uptake of electronic devices continues to grow.

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

The CAHPS Surgical Care survey was developed in consultation with patients using the most sophisticated, valid, and reliable methodologies available in survey and measurement science. During development of the survey (both English and Spanish versions), The American Institutes for Research (AIR) conducted 30 cognitive interviews in 2008 with 20 English- and 10 Spanish-speaking patients who had undergone a non-emergency, 90-day global surgery within the past 12 months. Each cognitive interview lasted approximately 2 hours. A trained cognitive interviewer administered the draft survey and conducted each interview using a semi-structured protocol.

During development, Round 1 cognitive testing revealed a variety of problems in the initial survey drafts. Changes were implemented and tested in Round 2, resulting in an improved survey. Examples of the types of revisions made to the draft Surgical CAHPS surveys include the following:

- A chronological ordering of sections was deemed the most intuitive order for respondents.

- Some of the section headings in the survey were simplified. For example, “Your Pre-Operative Care From This Surgeon” was changed to “Before Your Surgery.”

- The introductory sentences at the beginning of sections were eliminated. In some cases, they caused confusion and, with the simplified headings, they were no longer necessary.

- A new screener question for patients with no follow-up office visits was added.

- A definition of “staff at this surgeon’s office” was added to one question to ensure that respondents include all staff in their responses (receptionists, clerks, nurses, etc.).

- A response option of “Don’t know” was added to several items.

- Wording changes were made to various items to simplify and clarify comprehension.

- Several items were eliminated because of limited response variation across rounds or significant respondent interpretation problems.

- Spanish language translation changes were made to some items.

The cognitive testing led to many improvements in the draft Surgical CAHPS survey during development and improved the quality of the field testing and, ultimately, the NQF-endorsed S-CAHPS survey.

An important distinction when comparing patient experience versus patient satisfaction is that patient experience measures aspects of care that are actionable for surgical quality improvement. And because the survey instrument, protocol, analysis, and reporting are standardized, surgeons can benchmark and compare their performance with that of their peers within the same practice or outside of their practice. Surgeons may customize the S-CAHPS survey by adding survey items that are specific to their patients and practice. However, the core survey must be used in its entirety in order to be comparable with other S-CAHPS data. The S-CAHPS survey may be used in both the inpatient and outpatient setting.

A barrier identified as a result of operational use for the measure has been patient confidentiality when attempting to administer the survey using email. Because of guidance published within the Federal Register as part of the 2013 HIPAA

rules, initial emails sent to patients cannot contain any Protected Health Information unless sent via encrypted email. To comply, these initial emails must be sent in a generic manner that may have lead patients to believe they are spam, resulting in low response rates. Despite the desperate need for patient input to evaluate quality, US security laws are in effect that prevent the patient perspective from being easily incorporated into national quality improvement strategies.

More importantly, the S-CAHPS survey was designed by surgical care professionals (i.e., surgeons and anesthesiologists) for surgical patients. The survey measures aspects of patients´ surgical care that are important to them. The S-CAHPS provides more actionable data specific to surgery as compared to other available patient experience and CAHPS measures, such as the Hospital CAHPS, which focuses primarily on the facility rather than the surgeon. Studies have been published demonstrating their use for quality improvement (see Evidence attachment for more details).

Sources:

Levine, Burling, Huberman, and Hurtado from the American Institutes for Research. (2008) Surgical CAHPS: Cognitive Testing of the English and Spanish Survey Instruments. Report prepared for the American College of Surgeons. (Report was Attachment D of S-CAHPS 2011 NQF Submission).

Sage, J. (2013) What Surgeons Should Know About Using S-CAHPS. Bulletin of the American College of Surgeons. Accessible at http://bulletin.facs.org/2013/08/using-s-cahps/

Liu JB, Pusic AL, Temple LK, Ko CY. Patient-reported outcomes in surgery: Listening to patients improves quality of care. Bull Am Coll Surg. 2017; 102(3): 19-23.

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 Surgical Care 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. All of these tools are available at: https://www.cahps.ahrq.gov/surveys-guidance/cg/instructions/index.html. Requirements 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 Professional Certification or Recognition Program	Payment Program https://qpp.cms.gov/ CMS Quality Payment Program Quality Improvement (external benchmarking to organizations) American College of Surgeons National Surgical Quality Improvement Program https://www.facs.org/quality-programs/acs-nsqip CMS Core Quality Measures Orthopedics Set https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures.html Quality Improvement (Internal to the specific organization) Surgical Practices N/A

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

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

A. Payment Program

1. Name of Program and Sponsor: CMS Quality Payment Program (QPP) Merit-based Incentive Payment System (MIPS) and as part of an Advanced Alternative Payment Model (A-APMs)
2. Purpose: National pay-for-reporting program.
3. Current and Planned Use:

3.1 Current use: QPP Improvement Activity: Participation in a CAHPS survey, such as the S-CAHPS, is a high-weighted patient safety and practice assessment activity within the Improvement Activities component of MIPS for performance years 2017 and 2018. MIPS eligible clinicians who administer the S-CAHPS survey can attest to the completion of this activity and earn points toward their MIPS Final Score for the Improvement Activity component of MIPS.

3.2 Planned use:

3.2.1 2017 MIPS Quality Component: Pending CMS approval for the 2018 MIPS program, participants in the American College of Surgeons (ACS) Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism. The Phases of Surgical Care Measure set is an episode-based measure framework which follows the five phases of surgical care: preoperative, perioperative, intraoperative, postoperative, and postdischarge. The measure set is composed of surgical outcome measures, high-value process measures, and the appropriate S-CAHPS measures which follow up on key processes within the measure set.

3.2.2 Advanced Alternative Payment Model: The ACS is currently working with CMS on an Advanced Alternative Payment Model that we anticipate may soon be tested by Centers for Medicare and Medicaid Innovation (CMMI). This model incorporates a novel quality measurement framework which measures care around the patient for a given episode and incorporates patient reported experience and patient reported outcomes. As currently proposed, the surgical episodes include measures included in the S- CAHPS survey.

4. Geographic area and number and percentage of accountable entities and patients included: National Program. Percentage of entities and patients included will be determined in 2018 or 2019, depending on the CMS program.
5. Level of measurement and setting: Level of measurement will vary based on QPP program:

- Improvement Activities in MIPS: provider-level or group level;
- Quality Component of MIPS: provider-level;
- Advanced Alternative Payment Model: will be at either the provider level or at the level of the APM entity.

B. Quality Improvement

1. Name of Program and Sponsor: American College of Surgeons´ National Surgical Quality Improvement Program
2. Purpose: National Surgical Quality Improvement
3. Current Use: S-CAHPS is currently used by the ACS NSQIP to measure PROs alongside its clinical data in a pilot. Complementing the ACS NSQIP with PROs represents an opportunity to improve those outcomes that matter most to patients. None of the currently measured clinical outcomes give any insight as to whether the outcome aligned with the patients´ views and goals. Currently, more than 50 hospitals in the United States and internationally participate in this PRO pilot, which was launched in October 2017. All patients accrued into the ACS NSQIP are asked to complete the S-CAHPS survey as part of their follow-up. Data collection is ongoing, and we expect to provide annual measure updates to the NQF.
4. Geographic area and number and percentage of accountable entities and patients included: the ACS NSQIP currently includes more than 700 hospitals in the USA and internationally, representing more than 990,000 operations accrued into the registry annually.
5. Level of measurement and setting: hospitals in the ACS NSQIP
C. Quality Improvement
1. Name of Program and Sponsor: CMS Core Quality Measures Orthopedics Set
2. Purpose: The Core Quality Measure Collaborative, led by the America's Health Insurance Plans (AHIP) and its member plans' Chief Medical Officers, leaders from CMS and the National Quality Forum (NQF), as well as national physician organizations, employers and consumers, worked hard to reach consensus on core performance measures. Through the use of a multi-stakeholder process, the Collaborative promotes alignment and harmonization of measure use and collection across payers in both the public and private sectors. For more information, visit https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures.html.
3.Current use: The S-CAHPS survey is included in the Orthopedics Core measure set, which is available here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Downloads/Orthopedic-Measures.pdf
4. Geographic area and number and percentage of accountable entities and patients included: national; number of patients and accountable entities unknown
5. Level of measurement and setting: provider-level

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?*)

There has long been support by multi-stakeholder groups that the S-CAHPS survey should have been included in the Physician Compare and PQRS programs as a stand-alone measure in addition to the CG-CAHPS because the CG-CAHPS does not accurately reflect care provided to the surgical patient. The support for the S-CAHPS has been acknowledged by CMS as well as supported by the NQF Measures Applications Partnership (MAP). The S-CAHPS survey´s initial NQF endorsement is also a testament to this.
For CY 2014-2016 proposed regulations, CMS received public comments to the PQRS program supporting the inclusion of the S-CAHPS as a stand-alone measure. Comments from The American College of Surgeons stated that the CG-CAHPS survey would not accurately reflect the care provided by single or multispecialty surgical or anesthesia groups. ACS also noted that S-CAHPS has been tested by the same standards as CG-CAHPS and follows the same collection mechanism as CG-CAHPS. In the 2014 Medicare Physician Fee Schedule final rule, several commenters opposed the publication of CG-CAHPS measures citing that the measures are not relevant to their particular specialty. They requested that CMS allow physicians the flexibility to select the survey instruments and patient satisfaction measures most appropriate for their practices, and many of the commenters recommended CMS use S-CAHPS as an optional patient experience of care measure. CMS responded that the Agency understands that CG-CAHPS is not the most applicable CAHPS survey for all specialties and service settings represented by groups on Physician Compare. Therefore, they explained that the Agency will evaluate the feasibility of including additional CAHPS surveys, such as S-CAHPS, on the site in the future.
In the 2015 Medicare Physician Fee Schedule final rule, CMS agreed that the S-CAHPS survey would be more relevant to a surgical group practice compared to the CG-CAHPS and noted that the majority of commenters supported the use of S-CAHPS in the PQRS program. However, CMS did not accept and finalize the measure. CMS explained "due to the cost

and time it would take to find vendors to collect S-CAHPS data, it is not technically feasible to implement the reporting of the S-CAHPS survey measures for the 2017 or 2018 PQRS payment adjustments." They also note that Qualified Clinical Data Registries (QCDRs) will have the option to administer the S-CAHPS as a non-PQRS measure for the 2017 or 2018 PQRS payment adjustments (Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, Access to Identifiable Data for the Center for Medicare and Medicaid Innovation Models & Other Revisions to Part B for CY 2015; Final Rule, 79 Federal Register 219 November 13, 2014, page 67795).
Also in the 2015 Medicare Physician Fee Schedule final rule, several commenters noted the limitations of CAHPS for PQRS measures for some health care professionals and supported adding other types of patient experience data to Physician Compare, including the Surgical CAHPS and experience data collected via other sources. CMS agreed that Surgical CAHPS data is useful to consumers and explained that the Agency is exploring how it can incorporate this information into Physician Compare (Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, Access to Identifiable Data for the Center for Medicare and Medicaid Innovation Models & Other Revisions to Part B for CY 2015; Final Rule, 79 Federal Register 219 November 13, 2014, page 67777).
Similar comments were made by stakeholders in 2016 Medicare Physician Fee Schedule final rule, and similarly CMS explained that they understand that not all measures equally apply to all types of professionals included in Physician Compare, however, they believe the CAHPS for PQRS measures (ie. CG-CAHPS) apply to a large majority of professionals on the Physician Compare site (Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY2016, 80 Federal Register 220 November 16, 2015, page 71129).
Lastly, the NQF's MAP has recommended the inclusion of S-CAHPS for two consecutive years. In the MAP Pre-Rulemaking Report: 2013 Recommendations on Measures Under Consideration by HHS report the NQF MAP recommended the inclusion of S-CAHPS in the PQRS program. In the MAP 2014 Recommendations on Measures for More Than 20 Federal Programs, the MAP recommended the inclusion of the S-CAHPS measure in PQRS, Meaningful Use, the Value-based Payment Modifier, and Physician Compare. These reports can be found here:
MAP Pre-Rulemaking Report: 2013 Recommendations on Measures Under Consideration by HHS. Final report available at: http://www.qualityforum.org/Publications/2013/02/MAP_Pre-Rulemaking_Report_-_February_2013.aspx
MAP Pre-Rulemaking Report: 2014 Recommendations on Measures for More than 20 Federal Programs. Final report available at: http://www.qualityforum.org/Publications/2014/01/MAP_Pre-Rulemaking_Report__2014_Recommendations_on_Measures_for_More_than_20_Federal_Programs.aspx

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.*)

A. Payment Program and Public Reporting

1. Name of program and sponsor: CMS Physician Compare

2. Purpose: Physician Compare is a national public reporting website which provides consumers with quality of care information to make informed health care decisions. Physician Compare is also intended to encourage clinicians to improve the quality of care they provide to their patients and create incentives to maximize performance.

3. Planned Use: Pending CMS approval for the 2018 MIPS program, participants in the American College of Surgeons (ACS) Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism. The Phases of Surgical Care Measures are an episode-based measure framework which follows the five phases of surgical care: preoperative, perioperative, intraoperative, postoperative, and postdischarge. The measure set is composed of high-value process measures, appropriate S-CAHPS measures which correlates to the process measures, and surgical outcome measures. As part of this measure set, the S-CAHPS measures will be reported on the Physician Compare downloadable database if they meet CMS statistical public reporting standards which require that measures be valid, reliable, and accurate.

4. Geographic area and number and percentage of accountable entities and patients included: National Program. Percentage of entities and patients included will be determined in 2018.

B. Payment Program. Please see above in 4a1.1 for the use of S-CAHPS in the Quality Payment Program.

C. Professional Certification or Recognition Program

1. Name of program and sponsor: American Board of Surgery (ABS) Maintenance of Certification® (MOC) Part IV (Practice Assessment Resources) URL: http://www.absurgery.org/default.jsp?exam-mocpa

2. Purpose: The American Board of Surgery Maintenance of Certification Part IV can be satisfied by ongoing participation in a local, regional or national outcomes registry or quality assessment program. Components of the S-CAHPS will be available via the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP), and the ACS Surgeon Specific Registry (SSR), starting in 2018. Both the SSR and NSQIP registries can meet the MOC Part IV MOC component, and are listed on the ABS website as examples of good programs for meeting Part 4, http://www.absurgery.org/default.jsp?exam-mocpa.

3. Geographic area and number and percentage of accountable entities and patients included: National Program. Percentage of entities and patients included will be determined in 2018.

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.

Too early to broadly determine. However, as discussed in the Evidence form, several peer-reviewed studies have demonstrated the use of the S-CAHPS survey for local quality improvement purposes with good success.

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.

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.

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

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

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.

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.

There is no repository for S-CAHPS data. Trend data is not available. Surgeons use S-CAHPS to identify their strengths and weaknesses and to help develop strategies for improving patients' experiences with care delivered surrounding a surgery.

The objective of the S-CAHPS survey is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information. At the same time, consistent with the survey, the measures aim to

directly benefit a variety of users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care.

1. Patients will use information from the measures to help make better and more informed choices about their surgical care.

2. Practices, health plans, and insurers will use the measure results for quality improvement initiatives and incentives.

3. Specialty boards may use the measure results for maintenance of certification purposes.

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 have been uncovered.

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

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)

0005 : CAHPS Clinician & Group Surveys (CG-CAHPS)-Adult, Child

0007 : NCQA Supplemental items for CAHPS® 4.0 Adult Questionnaire (CAHPS 4.0H)

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

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

In December 2011, the Agency for Healthcare Research and Quality released the 2.0 version of the Surgical Care Survey. This update keeps the Surgical Care Survey consistent with the Clinician & Group Survey, which was updated in October 2011. Because the changes that led to the 2.0 designation do not represent a significant digression from the 1.0 version of this survey, the shift from 1.0 to 2.0 does not affect the ability of survey users to assess trends in performance.

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.)

The Surgical Quality Alliance (SQA; https://www.facs.org/advocacy/quality/surgical-quality-alliance) reviewed the current CAHPS® Clinician and Group Survey and identified critical gaps in content and approach related to the assessment of surgical care. For example, some critical gaps in the survey include informed consent, shared decision making, anesthesia care, and post-operative instructions and access, all of which are issues consumers find to be very important in surgery. Therefore, the SQA felt that the development of a surgical patient experience survey fit well into the mission to improve the quality of healthcare delivered to surgical patients.

Surgeries have high resource use, and poor quality can have serious consequences for patients, including death. Hospital stays that involve operating room (OR) procedures are more costly, on average, than stays that do not involve OR procedures. Therefore, improving the quality of surgical care is of paramount importance to patients and the healthcare system alike. To measure the quality of surgical care in the U.S., a survey that measures the patient-reported outcomes associated with care provided by single- or multispecialty surgical or anesthesia groups is necessary. As shown in the attached support letter from 2012, there is much support from the surgical specialty societies for NQF endorsement and use of the S-CAHPS survey. (See attached "Attachment C Main 1c3 Letter to NQF 2012_surgeryspecialty_1741.pdf".)

The Centers for Medicare and Medicaid Services (CMS) acknowledged the importance of allowing for the administration of S-CAHPS reporting in addition to the Clinician and Group CAHPS for Physician Quality Reporting System (PQRS). The ACS has emphasized to CMS that it is critical that the measures included in the quality-tiering composite are valid, reliable, and applicable to all health care professionals, to avert the unintended consequence of misclassifying a physician's care and unfairly affecting payment. Accordingly, the S-CAHPS measures have been submitted for consideration for the Merit-based Incentive Payment System of the Quality Payment Program.

Though the main emphasis of the survey is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information, the development of measures specific to surgical care benefits many users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care. Patients will use information from the survey to help make better and more informed choices about their surgical care. Practices, health plans, and insurers will use the surgical survey results for quality improvement initiatives and incentives. Specialty boards may use the survey's measure results for maintenance of certification purposes.

CAHPS surveys were originally developed to meet the need of consumers for usable, relevant information on quality of care from the patient's perspective. But they also play an important role as a quality improvement (QI) tool for health care organizations, which can use the standardized data to identify relative strengths and weaknesses in their performance, determine where they need to improve and track their progress over time.

Cultivating the use of CAHPS surveys for QI purposes is one of the key objectives for the CAHPS grants. The CAHPS Improvement Guide is a comprehensive resource for health plans, medical groups, and other providers seeking to improve their performance in the domains of quality measured by CAHPS surveys. The guide may be used to:

- Cultivate an environment that encourages and sustains quality improvement;
- Analyze the results of CAHPS surveys to identify strengths and weaknesses; and
- Develop strategies for improving performance.

Appendix

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

Attachment Attachment: Attachment_B_Main_1b2_S-CAHPS_score_Tables-636461813747558109.xlsx

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Surgeons, Division of Advocacy and Health Policy

Co.2 Point of Contact: Jill, Sage, jsage@facs.org, 202-672-1507-

Co.3 Measure Developer if different from Measure Steward: American College of Surgeons, Division of Advocacy and Health Policy

Co.4 Point of Contact: Jill, Sage, jsage@facs.org, 202-672-1507-

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.

Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP):

1. Priscilla Arnold, MD FACS - American Society of Cataract and Refractive Surgery
2. Larissa Temple, MD FACS - American Society of Colon and Rectal Surgeons
3. Laura King, MD – American Academy of Ophthalmology
4. Robin Brody, MD – American Academy of Otolaryngology – Head and Neck Surgery
5. Lee Eisenberg, MD FACS - American Academy of Otolaryngology – Head and Neck Surgery
6. Rahul Shah, MD FACS - American Academy of Otolaryngology – Head and Neck Surgery
7. Robert Haralson, MD FACS – American Academy of Orthopaedic Surgeons
8. Frank Opelka, MD FACS – American College of Surgeons
9. Antony Sidawy, MD FACS - Society for Vascular Surgery
10. Loren Hiratzka, MD FACS – Society for Thoracic Surgery
11. James Hicks, MD – American Society of Anesthesiologists
12. Andrea Pusic, MD FACS - American Society of Plastic Surgeons
13. Sharon Merrick, Staff - American Society of Anesthesiologists
14. Chip Amoe, Staff - American Society of Anesthesiologists
15. Jason Byrd, Staff - American Society of Anesthesiologists
16. Cathy Cohen, Staff - American Academy of Ophthalmology
17. Cherie McNett, Staff - American Academy of Ophthalmology
18. Kristine Schulz, Staff - American Academy of Otolaryngology - Head and Neck Surgery
19. Stephanie Jones, Staff - American Academy of Otolaryngology - Head and Neck Surgery
20. Beth Kosiak, Staff - American Urological Association
21. Suzanne Pope, Staff - American Urological Association
22. Nancey McCann, Staff - American Society of Cataract and Refractive Surgery
23. DeLaine Schmitz, Staff - American Society of Plastic Surgeons

24. Guy Beaumont, Staff – American College of Osteopathic Surgeons

25. Cynthia Shewan, Staff – Society for Thoracic Surgery

26. Elizabeth Hoy, Staff – American College of Surgeons

27. Caitlin Burley, Staff - American College of Surgeons

28. Valerie Oster, Staff - American College of Surgeons

29. Andrea Burling – American Institutes for Research

30. Roger Levine - American Institutes for Research

31. Samantha Sheridan - Westat

32. John Rauch – Westat

The Surgical quality Alliance (SQA) of the American College of Surgeons includes a Technical Advisory Panel (TAP) which provided continuous support to the development of the Surgical care Survey from the literature review to final testing. The TAP included 32 members from various surgery and anesthesiology specialty societies, each having technical expertise in the topic of surgical care.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2010

Ad.3 Month and Year of most recent revision: 06, 2012

Ad.4 What is your frequency for review/update of this measure? Periodic, as needed

Ad.5 When is the next scheduled review/update for this measure? 11, 2018

Ad.6 Copyright statement: "CAHPS" is a registered trademark of the Agency for Healthcare Research and Quality (AHRQ) but CAHPS surveys themselves are not copyrighted and in the public domain.

Ad.7 Disclaimers: None.

Ad.8 Additional Information/Comments: List of Attachments:

Attachments located in zip file named: Attachments 1741 11062017.zip

Attachment A Main S9 CY2015-90-day-global codes.xlsx

Attachment B Main 1b2 S-CAHPS_score_Tables.xlsx

Attachment C Main 1c3 Letter to NQF 2012_surgeryspecialty_1741.pdf

Attachment D Main 1c5 Surgical CAHPS Focus Groups_2nd Round_2010.pdf