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 #: 3357

Measure Title: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

Brief Description of Measure: Facility-level risk-standardized rate of acute, unplanned hospital visits within 7 days of a general surgery procedure performed at an ambulatory surgical center (ASC) among Medicare Fee-For-Service (FFS) patients aged 65 years and older. An unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission.

Developer Rationale: This measure aims to reduce adverse patient outcomes associated with ASC surgeries and improve follow-up care by capturing and illuminating, for providers and patients, post-surgery unplanned hospital visits that are often not visible to providers at ASCs. The measure score will assess quality and inform quality improvement.

Numerator Statement: The outcome being measured is acute, unplanned hospital visits (ED visit, observation stay, or unplanned inpatient admission) occurring within 7 days of a general surgery procedure performed at an ASC.

Denominator Statement: Target Population

Included patients:

The target population for this measure is Medicare FFS patients aged 65 years and older, who are undergoing outpatient general surgery procedures in ASCs that are within the scope of general surgery training. Specifically, the cohort of procedures includes the following types of surgeries: abdominal, alimentary tract, breast, skin/soft tissue, wound, and varicose vein.

The Medicare FFS population was chosen because of the availability of a national dataset (Medicare claims) that could be used to develop, test, and publicly report the measure. We limit the measure to patients who have been enrolled in Medicare FFS Parts A and B for the 12 months prior to the date of surgery to ensure that we have adequate data for identifying comorbidities for risk adjustment.

Included procedures:

The target group of procedures is surgical procedures that (1) are routinely performed at ASCs, (2) involve risk of post-surgery hospital visits, and (3) are within the scope of general surgery training. The scope of general surgery overlaps with that of other specialties (for example, vascular surgery and, plastic surgery). For this measure, we targeted surgeries that general surgeons are trained to perform with the understanding that other subspecialists may also be performing many of these surgeries at ASCs. Since the type of surgeon performing a particular procedure may vary across ASCs in ways that affect quality, the measure is neutral to surgeons' specialty training.

To identify eligible ASC general surgery procedures, we first identified a list of procedures from Medicare's 2014 and 2015 ASC lists of covered procedures, which include procedures for which ASCs can be reimbursed under the ASC payment system. This lists of surgeries is publicly available at: https://www.cms.gov/medicare/medicare-fee-for-service-payment/ascpayment/11_addenda_updates.html (download January 2014 and January 2015 ASC Approved HCPCS Code and Payment Rates, Addendum AA). Surgeries on the ASC list of covered procedures do not involve or require: major or prolonged invasion of body cavities, extensive blood loss, major blood vessels, or care that is either emergent or life-threatening. The ASC list is annually reviewed and updated by Medicare, and includes a transparent public comment submission and review process for addition and/or removal of procedure codes. Using an existing, defined list of surgeries, rather than defining surgeries de novo, is useful for long-term measure maintenance. Procedures listed in Medicare's list of covered ASC procedures are defined using Healthcare Common Procedure Coding System (HCPCS) and Common Procedural Terminology (CPT®) codes.

Ambulatory procedures include a heterogeneous mix of non-surgical procedures, minor surgeries, and more substantive surgeries. The measure is not intended to include very low-risk (minor) surgeries or non-surgical procedures, which typically have a high volume and a very low outcome rate. Therefore, to focus the measure only on the subset of surgeries on Medicare's list of covered ASC procedures that impose a meaningful risk of post-procedure hospital visits, the measure includes only "major" and "minor" procedures, as indicated by the Medicare Physician Fee Schedule global surgery indicator (GSI) values of 090 and 010, respectively. The GSI code reflects the number of post-operative days that are included in a given procedure's global surgical payment and identifies surgical procedures of greater complexity and follow-up care. This list of GSI values is publicly available for calendar year (CY) 2014 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1600-FC.html and for CY 2015 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1612-FC.html (download PFS Addenda, Addendum B).

Finally, to identify the subset of general surgery ASC procedures, we reviewed with consultants and Technical Expert Panel (TEP) members the Clinical Classifications Software (CCS) categories of procedures developed by the Agency for Healthcare Research and Quality (AHRQ). We identified and included CCS categories within the scope of general surgery, and only included individual procedures within the CCS categories at the procedure (CPT® code) level if they were within the scope of general surgery practice. We did not include in the measure gastrointestinal endoscopy, endocrine, or vascular procedures, other than varicose vein procedures, because reasons for hospital visits are typically related to patients' underlying comorbidities.

See the attached Data Dictionary, sheet S.9 "Codes Used to Define Cohort" for a complete list of all CPT procedure codes included in the measure cohort.

Denominator Exclusions: The measure excludes surgeries for patients without 7 or more days of continuous enrollment in Medicare FFS Parts A and B after the surgery. The measure excludes these patients to ensure all patients have full data available for outcome assessment.

Measure Type: Outcome

Data Source: Claims, Enrollment Data

Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: N/A Most Recent Endorsement Date: N/A

Criteria 1: Importance to Measure and Report

1a. Evidence

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

Evidence Summary – This measure will identify ambulatory surgical centers (ASC) that have significantly higher rates of unplanned hospital visits related to other ASCs performing the same types of procedures on similar patients. In the literature, hospital visit rates following outpatient surgery vary from 0.5-9.0%. This measure is based on literature suggesting that patient selection and preparation, post-operative care, and post-discharge planning can affect the rate of adverse events and unplanned admissions following outpatient surgery.

Empirical data demonstrating a relationship between the outcome to at least one healthcare process is now required. NQF guidance states that a 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.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *Is the performance data from the literature sufficient, in size and variance, to demonstrate that some ASC facilities are engaging in quality improvement activities to decrease unplanned hospital admissions after surgery better than others?*

Guidance from the Evidence Algorithm: Measure assesses performance on a health outcome (Box 1) → There is a relationship between the health outcome and one healthcare action (Box 2) → Pass

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

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

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

- The developer assessed ASC-level variation in performance scores using Medicare FFS claims data for fiscal years 2014 and 2015, which included 236,999 general surgeries from 1,642 ASCs.
- The developer reports variation in the risk adjusted measure scores, ranging from 0.42 to 2.13.

Disparities

- The developer evaluated disparities with the observed rate and then evaluated the magnitude of association of three risk factors (dual eligible, race, SES) with the outcome after adjustment. Dual eligible, African Americans, and those with AHRQ SES Index scores below 42.7 had higher observed rates.
- The developer concluded that the risk factors have a modest but statistically significant association with the risk of a hospital visit.

Disparity Marker	Observed Rate (%)
Dual Eligible	3.7
Non dual eligible	2.2
African American	3.1
Non African American	2.2
AHRQ SES Index <42.7	2.7
AHRQ SES Index >42.7	2.2

Disparity Marker	Odds Ratio	Confidence Interval (95%)	*p* value
Dual Eligible	1.34	1.22-1.48	<0.0001
Race	1.23	1.06-1.42	0.005
AHRQ SES Index	1.14	1.06-1.22	0.0004

Questions for the Committee:

- *Is there a gap in unplanned hospital visits following ambulatory surgical visits that warrants a national performance measure?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

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

1a. Evidence to Support Measure Focus:
**Sufficient data are presented to judge the measure performance (i.e. Medicare claims data nationwide)
**Is there at least one thing that the provider can do to achieve ... Yes I believe there is both in the literature and in my experience as a surgeon working in ASCs
Is the performance data from the literature sufficient, ... yes I believe the evidence is solid.
**The project is very logical and in many ways follows the methodology of 30 day readmission measures that we have evaluated. This will capture unplanned in-patient hospital admissions that occur within 7 days of general surgical procedures that are performed in ambulatory surgery centers and will risk adjust based on administrative data and some socioeconomic and racial factors. Centers will be evaluated for their relative performance assuming that those with lower rates of unplanned admissions are either having superior operative results with less complications and/or better processes of care in patient education and follow up communication than those that have higher rates, assuming very importantly that the patient's relative risk is indeed being accurately captured, which I have some concern about.
**Evidence well supported
**Evidence is sufficient
**Outcome measure with good data to support it
1b. Performance Gap
**Yes, substantially variability was demonstrated.
**Yes. While the performance gap is less than I expected, it is statistically significant. I believe that widespread reporting of and attention to this measure will likely improve quality of care in ASCs and consequently, patient outcomes.
I'm not aware of disparities in this area above and beyond those presented by the developers."
**This is a relatively new area being developed and an important one because surgical centers probably are not as closely scrutinized as hospitals leaving a considerable gap in evaluation of outcomes and processes, so this measure is needed to fill that gap. This should help to define the degree of gap, as I assume there probably is, and then to hopefully

close that, since most surgeons respond to this type of comparative data. The fact that there is considerable variation in the incidence of 7 day admissions indicates that there is a need to close the gap between the centers.
**PG exists
**Performance gap is well-described
**Performance gap is provided and demonstrates a gap

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: Sherrie Kaplan, Christie Tiegland, Laurent Glance

Evaluation of Reliability and Validity:

Evaluation A

Evaluation B

Evaluation C

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 Scientific Methods Panel 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 Scientific Methods Panel 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**

Evaluation A: 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: 3357

Measure Title: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

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.

   The time period for measurement is not entirely clear to me. They indicate they are using 2 years of data consistent with CMS new practices, but it seems that the initial 12 months is required to gather the risk factor variables and that any procedures during first 12 months will not be included in measure because those patients would not have prior 12 months of data? And the period ends at least 7 days before the end of the measurement year? I think that is what they are saying but it is not clearly stated as to what the exact measurement period is.

   Also, I do not see any exclusions from reporting for ASCs that may have a minimal number of the type of procedures included in the measure and rates may thus be unstable or not true indicator of quality for that ASC. The algorithm did control for differences in numbers of procedures but not sure that is sufficient for small denominators.

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)

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)

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)

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]

Empirical testing for data element validity did not cover ALL critical data elements in my mind, specifically the SES proxy used which is measured using a very small survey sample in most geographic areas and aggregated at ZIP code level which can cover widely disparate populations in many geographic ZIPs thus averaging out the social risk factors and resulting in little/no impact on the outcome.

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)

The social risk factor data elements did not have sufficient reliability testing at the patient level. Though a validated SES composite score was used, it was calculated using ACS block level data (a very small sample) at the 5 digit ZIP level, which comprises a wide population that can have widely varying SES within the ZIP area, resulting in "averaging out" and thus showing little impact overall. In addition, race/ethnicity was define as African American vs. Other, thus Hispanics, Asians and other race/ethnic groups are lumped in with "White" which also can skew any impact of race/ethnicity. Given they found effects of SES using the crude ZIP level survey sample data, using social risk factor data collected at a more granular level could very likely show more significant differences that would also impact the outcome rates significantly after controlling for other risk factors included in the models.

I also have concerns about using HCCs as the data level for chronic conditions. Certain individual conditions within an HCC are often more highly associated with the outcome but that relationship gets lost the hierarchical category. An HCC may not be highly associated but individual conditions may be very highly associated. The HCCs also do not necessarily capture the impact of having multiple conditions that may be combined in one HCC.

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)

Concerns about the percentages shown for the quartile cut-offs and how well they truly capture the intent of the variable at the ASC level. For example, the 1st quartile cut-off for the proportion of Medicaid dual eligible patients at the ASC level is <=1.82% and for the 4th quartile >= 7.06%. 7% is still a VERY LOW proportion of dual eligible patients, indicating the distribution of dual eligible patients having one of the outpatient surgeries at an ASC seems to be unexpectedly low at almost all the ASCs included in the sample. The 4th quartile includes ASCs with only 7% duals up to ASCs with 100% duals potentially, which could be why we didn't see this contributing to ASC level differences in rates when including this as a social risk factor adjuster.

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)

   Face validity only (however, they refer to prior empirical validity testing "For several other NQF-endorsed measures, our team has demonstrated the validity of using claims data for risk adjustment in lieu of medical record data in estimating facility-level measure scores.")

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)

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

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)

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)

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.]

See comments above re SES data used and use of HCCs; feel the issues are strong enough to require further empirical validation. Would like to see results further stratified by percent dual population in ASC for example, comparing not all ASCs with 7% or more of patients served having dual status but rates for ASCs with 80% or more of population served having dual status. I am unconvinced the data used for race/ethnicity and SES is granular and accurate enough to actually capture the impact of those risk factors on the outcome.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Evaluation B: 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: 3357

Measure Title:

Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

RELIABILITY

11. 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.

12. 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)

13. 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)

14. 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)

15. 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)

    The ICC – used to examine measure reliability – was 0.51. Values less than 0.5 are indicative of poor agreement, and values between 0.5 and 0.75 are consistent with moderate agreement. A value of 051 is right at the margin – and hence more consistent with poor-to-moderate agreement.

16. **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)

The ICC – used to examine measure reliability – was 0.51. Values less than 0.5 are indicative of poor agreement, and values between 0.5 and 0.75 are consistent with moderate agreement. A value of 051 is right at the margin – and hence more consistent with poor-to-moderate agreement.

17. Was other reliability testing reported?

☐Yes (go to Question #8)

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

18. 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)

19. 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)

20. **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)

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

17. 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]

18. 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)

19. 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)

20. 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)

21. 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)

22. 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

23. 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)

24. 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)

25. **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)

26. 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)

27. 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)

28. **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

29. Was other validity testing reported?

    ☐Yes (go to Question #14)

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

30. 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)

    The measure developers did not specifically test the validity of patient-data elements for this specific measure. As per the measure developers, "While the applicability of these findings to our measure may be limited because these medical record validations medical record evaluations were focused on patients admitted for specific medical conditions, they nevertheless suggest claims data generally have an acceptable degree of agreement with clinical data at a facility level." However, since data element validation on prior measures is also based on a look-back period of 12 months, I believe that prior validation of data elements is generally applicable to this measure.

31. 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)

    I rated this as a "yes" – even though the measure developers do not report the validation results for this specific measure. But I think that the measure developers should summarize the results of data validation for the previous NQF-endorse readmission measure since both metrics capture similar outcomes.

32. **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)

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.]

I scored this as "moderate" because measure reliability was empirically assessed as moderate.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Evaluation C: 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: 3357

Measure Title: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

RELIABILITY

21. 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.

22. 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)

23. 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)

24. 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)

25. 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)

    But they used only split-half reliability. What is really needed is the ICC for between vs. within variance by facility, not agreement between samples.

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

26. **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)

27. Was other reliability testing reported?

☐Yes (go to Question #8)

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

28. 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)

29. 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)

30. **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)

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

33. 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]

34. 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)

35. 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)

36. 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)

37. 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)

38. 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

39. 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)

Only face validity was assessed

40. 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)

41. **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)

42. 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)

43. 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)

44. **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

45. Was other validity testing reported?

    ☐Yes (go to Question #14)

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

46. 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)

    Face validity was assessed as was discriminate validity (for disparities)

47. 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)

    Using current NQF standards

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

48. **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)

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.]

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

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

2a1. Reliability-Specifications

**No issues.
**Evaluation A raises some excellent points regarding the limitations of the disparities analysis (i.e. limiting race to two categories, analyzing SES at the zip code level etc.). However, I find the rationale provided in the measure submission adequately responds to these concerns.
I do believe the measure could be successfully implemented and I have no major concerns with the reliability or validity testing.
I do not see a need to discuss or vote on validity
**I believe the developers have made a reasonably good case for the reproducibility and reliability of collecting this data.
**Reliable
**Well-defined
All clearly defined
**Despite the reply from the stewards, the C-stasitics offered are on the margin. Overall community effect of the population of the ASC and its effect on performance within the measure across the the three given SES parameters is not clear, and its dismissal is not as well. The stewards might want to consider the incorporation of some of the specific individual ICD 9 (now 10) codes that were brought into the Risk Stratified Episode of Care Cost Measure for THA/TKA that CORE developed previously, especially the neuro-degenerative/neuro-cognitive codes and the more specific codes re: obesity. The patients with higher HCC risk factors might be over populating the return to hospital statistics because of returns unrelated to the surgery; perhaps a the longitudinal rate of hospital encounters pretending the index event could be used for a separate risk factor?

2a2. Reliability - Testing

**No
**no
**I would consider the reliability at least as moderate, being concerned somewhat about the accuracy of administrative data which is generally not audited to the extent that clinical databases audit.
**No
**No concerns
**No
**The concern is the risk adjustment, especially if used across small populations and small percentile differences in a payment program.

2b1. Validity
2b4-7. Threats to Validity
2b4. Meaningful Differences
**No issues
**no
**I have some concerns about the validity related to the risk adjustment process, both social and co-morbidities, etc. I realize that there are several papers from centers and individuals that I respect comparing administrative to chart abstracted data but I still have concerns about the degree of severity of various co-morbidities or the lack of that data with the administrative data. Exclusions do not seem to be an issue in this protocol.
**Valid
**No
**Not a substantial threat
**This measure borrows on previous validation work regarding the validity of the administrative data set and real chart review. It is not clear that that conclusion can be assumed.
2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)
**Minimal risk adjustment (other than socioeconomic) applied.
**no
**In regard to the social risks, zip codes have short comings because a zip code area can have a mixture of socioeconomic neighborhoods, and the racial diversity of the US is not captured in the data presented. I probably would rate the validity as moderate. It is somewhat reassuring that the C-index is 0.69, not great, but reasonably good.
Risk adjusted via admin DB
**There are substantial differences among AA, dual-eligible and low SES populations which CMS plans to adjust for.
**Reasonable risk adjustment as much as possible with a measure specified in claims
**Please see comments under reliability. Concerns exist, especially if the measure is used to adjust payments based on small differences in percentile performance.

Criterion 3. Feasibility

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 is a claims based measure. No data elements are in electronic sources.
- There are no fees, licensing, or other requirements to use this measure as specified.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility
**Can be captured in national CMS billing data and calculated remotely.
**I agree with NQF staff this this is high.
**I believe that most of the proposed data points are reasonably straight forward and that the project is very feasible.
**Feasible
**No concerns - this measure can be specified in claims
**No concerns
** Please see comments under reliability. Concerns exist, especially if the measure is used to adjust payments based on small differences in percentile performance.

Criterion 4: Usability and Use

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 developer reports that this measure may ultimately be used in one or more Centers for Medicare & Medicaid Services' (CMS) programs, such as the Ambulatory Surgical Center Quality Reporting Program (ASCQR).
    - This measure was approved for consideration under the ASCQR program and was discussed by the MAP Hospital Workgroup in December 2017. MAP conditionally supported this measure for the ASCQR program pending NQF review and endorsement. MAP recognized that this measure assesses an important outcome for patients receiving care at ambulatory surgery centers and addresses crucial safety concerns by tracking if a patient requires treatment at an acute care hospital (including emergency department (ED) visits, observation stays, and unplanned inpatient admissions) within 7 days of the procedure performed at an ASC. MAP noted this measure could help balance incentives to perform more procedures on an outpatient basis. However, MAP acknowledged a number of concerns raised in public comments about the measure. Commenters raised concerns about the attribution model of measure, noting that these are relatively rare events and could disproportionately impact low-volume ASCs, and that the measure may need risk adjustment for social risk factors. MAP noted this measure should be submitted for NQF endorsement to assess the potential impact of these concerns on the reliability and validity of the measure.

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 reports that they recruited a national TEP during measure development and hosted a public comment period. TEP members and commenters included representatives of the ASCs.
- Data and results were provided to the TEP and members of the TEP were able to give input on five occasions during the measure development process.
- Revisions made to the measure based on feedback included: renaming the measure to reflect the procedures included in the measure cohort; removal of 15 individual CPT codes that were outside the scope of general surgery practice; and a review of variables for the final risk model where one was retained (opioid use) since experts felt it was an important risk predictor.

Additional Feedback: Not applicable

***Questions for the Committee**:*

- *How 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

- The developer indicated that the question was not applicable since the measure is not yet in use.

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 indicated that the question was not applicable since the measure is not yet in use.

Potential harms:

- The developer indicated that the question was not applicable since the measure is not yet in use.

Additional Feedback:

- The developer indicated that the question was not applicable since the measure is not yet in use.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

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

4a1. Use - Accountability and Transparency

**Not yet in use

**The committee asked how measure can be used to further the goal of high quality healthcare? This is a ubiquitous wide-lens view of ASC performance. A focus on reducing readmissions is likely to increase attention to detail in the pre, intra, and post op care of ASC patients. With a national trend towards pushing procedures more into the outpatient realm, this measure can be an important check to be sure we are not pushing too hard to the detriment of patients.

**There is significant variation in rates of 7 day admission across the various surgery centers which allows the centers to identify opportunities for improvement after seeing the data analysis.

**Usable

**New measure - but MAP approved it for use going forward

**Not currently

**Please see comments under reliability. Concerns exist, especially if the measure is used to adjust payments based on small differences in percentile performance.

4b1. Usability – Improvement

**This measure should be a good one to drive performance.

**I do not understand staff comments in this area

**In general, responsible surgeons and other clinicians want to perform to the best of their abilities and if data shows them to not be performing well this should stimulate them to try to understand why and make some changes to improve the outcomes. This has worked well in other settings, but only if they have confidence in the data and its analysis.

**No concerns

**Unknown at this point. IF the risk adjustment is insufficient to account for the substantial differences in outcome among low SES populations, public reporting of this measure could result in poor access to outpatient care in that population
**Benefits likely outweigh harms
**The measure would be best used first in public reporting so that potential deficits in risk adjustment and SES risk adjustment could be assessed before moving to payment adjustments.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2539 Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy
- 2687 Hospital Visits after Hospital Outpatient Surgery
- 3366 Hospital Visits after Urology Ambulatory Surgical Center Procedures (currently under consideration by the Surgery Standing Committee)

Harmonization

- The developer reports that the measure specification are harmonized with the above listed related measures.

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

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 9, 2018

- No NQF members have submitted support/non-support choices as of this date. No comments have been submitted as of this date.

Developer Submission

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 3357

Corresponding Measures:

De.2. Measure Title: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

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

De.3. Brief Description of Measure: Facility-level risk-standardized rate of acute, unplanned hospital visits within 7 days of a general surgery procedure performed at an ambulatory surgical center (ASC) among Medicare Fee-For-Service (FFS) patients aged 65 years and older. An unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission.

1b.1. Developer Rationale: This measure aims to reduce adverse patient outcomes associated with ASC surgeries and improve follow-up care by capturing and illuminating, for providers and patients, post-surgery unplanned hospital visits that are often not visible to providers at ASCs. The measure score will assess quality and inform quality improvement.

S.4. Numerator Statement: The outcome being measured is acute, unplanned hospital visits (ED visit, observation stay, or unplanned inpatient admission) occurring within 7 days of a general surgery procedure performed at an ASC.

S.6. Denominator Statement: Target Population

Included patients:

The target population for this measure is Medicare FFS patients aged 65 years and older, who are undergoing outpatient general surgery procedures in ASCs that are within the scope of general surgery training. Specifically, the cohort of procedures includes the following types of surgeries: abdominal, alimentary tract, breast, skin/soft tissue, wound, and varicose vein.

The Medicare FFS population was chosen because of the availability of a national dataset (Medicare claims) that could be used to develop, test, and publicly report the measure. We limit the measure to patients who have been enrolled in Medicare FFS Parts A and B for the 12 months prior to the date of surgery to ensure that we have adequate data for identifying comorbidities for risk adjustment.

Included procedures:

The target group of procedures is surgical procedures that (1) are routinely performed at ASCs, (2) involve risk of post-surgery hospital visits, and (3) are within the scope of general surgery training. The scope of general surgery overlaps with that of other specialties (for example, vascular surgery and, plastic surgery). For this measure, we targeted surgeries that general surgeons are trained to perform with the understanding that other subspecialists may also be performing many of these surgeries at ASCs. Since the type of surgeon performing a particular procedure may vary across ASCs in ways that affect quality, the measure is neutral to surgeons' specialty training.

To identify eligible ASC general surgery procedures, we first identified a list of procedures from Medicare's 2014 and 2015 ASC lists of covered procedures, which include procedures for which ASCs can be reimbursed under the ASC payment system. This lists of surgeries is publicly available at: https://www.cms.gov/medicare/medicare-fee-for-service-payment/ascpayment/11_addenda_updates.html (download January 2014 and January 2015 ASC Approved HCPCS Code and Payment Rates, Addendum AA). Surgeries on the ASC list of covered procedures do not involve or require: major or

prolonged invasion of body cavities, extensive blood loss, major blood vessels, or care that is either emergent or life-threatening. The ASC list is annually reviewed and updated by Medicare, and includes a transparent public comment submission and review process for addition and/or removal of procedure codes. Using an existing, defined list of surgeries, rather than defining surgeries de novo, is useful for long-term measure maintenance. Procedures listed in Medicare's list of covered ASC procedures are defined using Healthcare Common Procedure Coding System (HCPCS) and Common Procedural Terminology (CPT®) codes.

Ambulatory procedures include a heterogeneous mix of non-surgical procedures, minor surgeries, and more substantive surgeries. The measure is not intended to include very low-risk (minor) surgeries or non-surgical procedures, which typically have a high volume and a very low outcome rate. Therefore, to focus the measure only on the subset of surgeries on Medicare's list of covered ASC procedures that impose a meaningful risk of post-procedure hospital visits, the measure includes only "major" and "minor" procedures, as indicated by the Medicare Physician Fee Schedule global surgery indicator (GSI) values of 090 and 010, respectively. The GSI code reflects the number of post-operative days that are included in a given procedure's global surgical payment and identifies surgical procedures of greater complexity and follow-up care. This list of GSI values is publicly available for calendar year (CY) 2014 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1600-FC.html and for CY 2015 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1612-FC.html (download PFS Addenda, Addendum B).

Finally, to identify the subset of general surgery ASC procedures, we reviewed with consultants and Technical Expert Panel (TEP) members the Clinical Classifications Software (CCS) categories of procedures developed by the Agency for Healthcare Research and Quality (AHRQ). We identified and included CCS categories within the scope of general surgery, and only included individual procedures within the CCS categories at the procedure (CPT® code) level if they were within the scope of general surgery practice. We did not include in the measure gastrointestinal endoscopy, endocrine, or vascular procedures, other than varicose vein procedures, because reasons for hospital visits are typically related to patients' underlying comorbidities.

See the attached Data Dictionary, sheet S.9 "Codes Used to Define Cohort" for a complete list of all CPT procedure codes included in the measure cohort.

S.8. Denominator Exclusions: The measure excludes surgeries for patients without 7 or more days of continuous enrollment in Medicare FFS Parts A and B after the surgery. The measure excludes these patients to ensure all patients have full data available for outcome assessment.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable

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

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

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

Gen_Surg_ASC__NQF_Evidence_Attachment_FINAL_111417.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.

1a Evidence (subcriterion 1a)

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

Measure Title: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

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

Date of Submission: Click here to enter a date

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:

☐ Patient-reported outcome (PRO):

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.

Unplanned hospital visits following ambulatory surgical center (ASC) surgical procedures often reflect procedure-related adverse events and quality issues. Strategies and interventions that have been shown to reduce unplanned hospital visits after outpatient surgical procedures include:

1) Appropriate patient selection for surgical procedures [1];

2) Appropriate patient education on preparation prior to procedures [2];

3) Improving the technical quality of the surgery, including the choice of procedural technique and anesthesia [3];

4) Prevention of surgical site infections through evidence-based guideline-concordant care [4,5]; and

5) Prevention of adverse drug events through medication reconciliation [6].

The measure will identify ASCs that have significantly higher rates of unplanned hospital visits relative to other ASCs performing the same types of surgical procedures on similar patients and will prompt ASCs to evaluate care processes and implement quality improvement strategies.

Citations:

1. Fleisher LA, Pasternak LR, Lyles A. A novel index of elevated risk of inpatient hospital admission immediately following outpatient surgery. *Arch Surg.* 2007;142(3):263-268.

2. Romero A, Joshi GP. Adult Patient for Ambulatory Surgery: Are There Any Limits? *ASA Newsletter.* 2014;78(9):18-20.

3. Whippey A, Kostandoff G, Paul J, Ma J, Thabane L, Ma HK. Predictors of unanticipated admission following ambulatory surgery: a retrospective case-control study. *Can J Anaesth .* 2013;60(7):675-683.

4. Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR, Committee HICPA. Guideline for prevention of surgical site infection, 1999. *Am J Infect Control.* 1999;27(2):97-134.

5. Agency for Healthcare Research and Quality. Proactive Risk Assessment of Surgical Site Infection in Ambulatory Surgery Centers: Final Contract Report. Chapter 3: Risk-Informed Interventions. April 2013. Available at: http://www.ahrq.gov/research/findings/final-reports/stpra/stpra3.html. Accessed July 18, 2016.

6. Joint Commission. Joint Commission National Patient Safety Goals: Practical Strategies and Helpful Solutions for Meeting these Goals. 2005; http://teacherweb.com/NY/StBarnabas/Law-PublicPolicy/JCINT-2005.pdf. Accessed June 8, 2016.

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.)

Not applicable. This measure is not derived from patient report.

****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.

The outcome of unplanned hospital visits following outpatient surgery is an accepted measure of outpatient surgical care quality and reflects important features of healthcare structure, process, and service. These features include patient selection and management, technical aspects of the surgery, and delivery of guideline-concordant care. This measure will provide the opportunity for ASCs to become aware of and to lower rates of adverse events leading to hospital visits after general surgery procedures performed at ASCs.

A hospital visit after outpatient surgery is unexpected, and many of the reasons for such hospital visits are preventable. In the literature, hospital visit rates following outpatient surgery vary from 0.5-9.0%, based on the type of surgery, outcome measured (admissions alone or admissions and emergency department [ED] visits), and timeframe for measurement after surgery [1-10]. These hospital visits can occur due to a range of adverse events, including major adverse events, such as infection, post-operative bleeding, and urinary retention. Patients also frequently report minor adverse events – for example, uncontrolled pain, nausea, and vomiting – that may result in unplanned acute care visits following surgery.

There is literature providing evidence that interventions can improve patient outcomes after outpatient surgery. Studies, many focusing on surgeries in the hospital outpatient department setting, point to the importance of post-discharge factors, such as ability to manage pain and availability of a responsible caregiver, in reducing poor outcomes [3, 11-15]. The quality of patient selection, patient preparation, post-operative care and post-discharge planning can affect the rate of adverse events and unplanned hospital visits following outpatient surgery [3, 11-12]. The risk of unplanned hospital visits is influenced by various technical aspects of the surgery, including anesthetic technique [11-13] and length of surgery [12]. Although there is limited evidence in the ASC context, these interventions should be applicable in both settings.

Additionally, there are growing efforts to enable ASC providers to systematically address issues of complications of surgical care and communication between providers of adverse events when they occur [16-18]. For example, the Agency for Healthcare Research and Quality (AHRQ) developed a quality improvement collaborative for the 65 ambulatory surgery facilities in 47 states to reduce healthcare-associated infections and surgical harms in ASCs through 1) the use of a surgical safety checklist curriculum, and 2) improved safety culture through teamwork and communication [18]. ASC providers involved in the collaborative concluded that efforts to increase the availability of meaningful data would be beneficial to the accurate assessment of outcomes in the ASC setting, reduce admissions, and would facilitate ASC's ability to follow patients after discharge.

Citations:

1. Majholm BB. Is day surgery safe? A Danish multicentre study of morbidity after 57,709 day surgery procedures. Acta anaesthesiologica Scandinavica. 2012;56(3):323-331.

2. Whippey A, Kostandoff G, Paul J, Ma J, Thabane L, Ma HK. Predictors of unanticipated admission following ambulatory surgery: a retrospective case-control study. Canadian Journal of Anesthesia/Journal canadien d'anesthésie. 2013;60(7):675-683.

3. Fleisher LA, Pasternak LR, Herbert R, Anderson GF. Inpatient hospital admission and death after outpatient surgery in elderly patients: importance of patient and system characteristics and location of care. Arch Surg. 2004;139(1):67-72.

4. Coley KC, Williams BA, DaPos SV, Chen C, Smith RB. Retrospective evaluation of unanticipated admissions and readmissions after same day surgery and associated costs. Journal of clinical anesthesia. 2002;14(5):349-353.

5. Bain J, Kelly H, Snadden D, Staines H. Day surgery in Scotland: patient satisfaction and outcomes. Quality in Health Care. 1999;8(2):86-91.

6. Fortier J, Chung F, Su J. Unanticipated admission after ambulatory surgery--a prospective study. Canadian journal of anaesthesia = Journal canadien d'anesthesie. 1998;45(7):612-619.

7. Aldwinckle R, Montgomery J. Unplanned admission rates and postdischarge complications in patients over the age of 70 following day case surgery. Anaesthesia. 2004;59(1):57-59.

8. Owens PL, Barrett ML, Raetzman S, Maggard-Gibbons M, Steiner CA. Surgical site infections following ambulatory surgery procedures. *JAMA.* 2014;311 (7): 709-716.

9. Mioton LM, Buck DW, 2nd, Rambachan A, Ver Halen J, Dumanian GA, Kim JY. Predictors of readmission after outpatient plastic surgery. *Plastic & Reconstructive Surgery.* 2014; 133(1):173-180.

10. Bhattacharyya N. Unplanned revisits and readmissions after ambulatory sinonasal surgery. *Laryngoscope.* 2014; 124(9):1983-1987.

11. Romero A, Joshi GP. Adult Patient for Ambulatory Surgery: Are There Any Limits? *ASA Newsletter.* 2014;78(9):18-20.

12. Whippey A, Kostandoff G, Paul J, Ma J, Thabane L, Ma HK. Predictors of unanticipated admission following ambulatory surgery: a retrospective case-control study. *Can J Anaesth.* 2013;60(7):675-683.

13. Perrot DH. Anesthesia Outside the Operating Room in the Office-Based Setting. *Curr Opin Anaesthesiol*. 2008; 21L 480-485

14. Pearson, Alan, Marilyn Richardson, and Michelle Cairns. ""Best practice" in day surgery units: a review of the evidence." *Ambulatory Surgery* 11.1 (2004): 49-54.

15. Allison, Jan, and Michelle George. "Using preoperative assessment and patient instruction to improve patient safety." *AORN journal* 99.3 (2014): 364-375.

16. Agency for Healthcare Research and Quality. Proactive Risk Assessment of Surgical Site Infection in Ambulatory Surgery Centers: Final Contract Report. Chapter 3: Risk-Informed Interventions. April 2013. Available at: http://www.ahrq.gov/research/findings/final-reports/stpra/stpra3.html. Accessed July 18, 2016.

17. Joint Commission. Joint Commission National Patient Safety Goals: Practical Strategies and Helpful Solutions for Meeting these Goals. 2005; http://teacherweb.com/NY/StBarnabas/Law-PublicPolicy/JCINT-2005.pdf. Accessed June 8, 2016.

18. The Health Research & Educational Trust of the American Hospital Association. AHRQ Safety Program for Ambulatory Surgery, May 2017. Available at: https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/hais/tools/ambulatory-surgery/sections/ambulatory-surgery-report.pdf. Accessed June 23, 2017.

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.

This measure aims to reduce adverse patient outcomes associated with ASC surgeries and improve follow-up care by capturing and illuminating, for providers and patients, post-surgery unplanned hospital visits that are often not visible to providers at ASCs. The measure score will assess quality and inform quality improvement.

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.*

We assessed ASC-level variation in performance scores using 100% Medicare FFS claims data for 2014-2015 (please see Measure Testing Form Section 1.2 and Section 1.7 for full description of the dataset). Using the 2014-2015 data (which included 286,999 general surgeries from 1,642 ASCs meeting a minimum volume threshold of at least 25 cases), we found variation in the risk-adjusted measure scores among ASCs. The median RSHVR was 0.97, ranging from 0.42 to 2.13 (the 25th and 75th percentiles were 0.90 and 1.10, respectively).

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Not applicable. We provide performance scores in 1b.2. See Evidence Form for summary of data from the literature that further indicates opportunity for improvement.

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.*

To examine the impact of social risk factors on the measure calculation we evaluated three indicators of social risk: 1) Medicare-Medicaid dual eligibility (yes vs. no), 2) race (African American vs. all others), and 3) the AHRQ SES Index (explained in Section 2.b.3 of the Measure Testing Form). For these analyses we used 100% Medicare FFS claims data from CYs 2014-2015. These data included 3,653 ASC facilities and 303,220 general surgery procedures. Our goal for these analyses were twofold: 1) to examine whether these factors were associated with increased risk in hospital visits after adjusting for other risk factors and 2) to evaluate the impact of social risk factors on ASC-level measure scores.

We present these analyses and results in greater detail in Section 2b3.4b of the Measure Testing Form. In brief, to evaluate the association of these risk factors with the outcome, we first quantified the observed rate. We then evaluated the magnitude of association of these social risk factors with the outcome after adjustment for clinical comorbidities, procedure type, and age by including each individual indicator as a variable in our risk-adjustment model. Each factor's effect was quantified using odds ratios (ORs) and tested for significance. In addition, we evaluated the change in each model's predictive ability (c-statistic).

To evaluate the impact of social risk factors on the ASC-level measure scores, we compared RSHVRs calculated with and without each disparity marker included in the model. For these analyses, we calculated the RSHVR difference for each ASC (RSVHR with the social risk variable and RSHVR without the social risk variable) and calculated Pearson correlation coefficients for the paired scores.

We further examined the potential impact of these social risk factors on measure scores by comparing RSHVR distributions using current specifications. ASCs were stratified by the proportion of patients at the ASC with each social risk factor, and placed into quartiles based on these proportions. These stratified distributions were examined for systematic differences in RSHVR across quartiles.

Results

Observed hospital visit rates were higher for patients with each disparity marker: 3.7% for dual-eligible patients compared to 2.2% for non-dual-eligible patients, 3.1% for African-American patients compared to 2.2% for non-African-American patients, and 2.7% for low SES patients (scores below 42.7 on the AHRQ SES Index) compared to 2.2% for higher SES patients (scores above 42.7 on the AHRQ SES index). Furthermore, inclusion of each of these risk factors in our models indicated a statistically significant association after controlling for other risk adjusters in our model (dual-eligible: OR: 1.34, 95% CI: 1.22 -1.48, p < 0.0001; race: OR: 1.23, 95% CI: 1.06-1.42, p=0.005; AHRQ SES Index: OR: 1.14, 95% CI: 1.06-1.22, p=0.0004).

However, entering these variables into the risk-adjustment model did not improve model performance (c-statistics remained unchanged) and did not substantially change ASC-level measure scores. Correlation coefficients between risk-standardized hospital visit ratios with and without adjustment for these factors were near 1.

Further, the analyses of ASCs stratified into quartiles based on proportions of dual-eligible, African-American, and low SES patients (as identified by the AHRQ SES Index) showed largely overlapping distributions of the RSHVRs by quartile, although longer tails at the upper ends of the distributions were observed for ASCs with the highest percent of patients with the social risk factor (4th quartile). Distributions for low % of social risk factor ASCs (1st quartile) and high % social risk factor ASCs (4th quartile) by each social risk factor are shown in Table 2, Section 2b3.4b, of the Measure Testing Form.

Based on these analyses we conclude that although the three social risk factors we examined have a modest but statistically significant association with the risk of a hospital visit, these patient-level factors have a limited effect on the ASC-level measure scores. We did not adjust the models for these social risk factors since the association of these factors with the outcome may be quality related, and since these factors have a limited relationship to the facility-level scores.

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

Not applicable. Disparities data and results are discussed above in Section 1b.4.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** Gen_Surg_ASC_NQF_Data_Dictionary_v1.0.xlsx

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

No, this is not an instrument-based measure **Attachment:**

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.

Not an instrument-based measure

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.

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

Not applicable.

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

The outcome being measured is acute, unplanned hospital visits (ED visit, observation stay, or unplanned inpatient admission) occurring within 7 days of a general surgery procedure performed at an ASC.

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).

Outcome Definition

The outcome is unplanned hospital visits, defined as an ED visit, observation stay, or unplanned inpatient admission, occurring within 7 days of the general surgery procedure performed at an ASC identified using Centers for Medicare & Medicaid Services (CMS) Medicare administrative claims data.

Time Period for Data

Numerator time window: 7 days after ASC procedures for unplanned hospital visits.

Denominator time window: General surgery ASC procedures performed during the measurement period.

Identification of Planned Admissions

The measure outcome includes hospital visits within the first 7 days following the procedure, unless that inpatient admission is deemed a “planned” admission. We applied CMS’s Planned Readmission Algorithm Version 4.0 to identified planned admissions [1]. Planned admissions are defined as those planned by providers for anticipated medical treatment or procedures that must be provided in the inpatient setting. CMS seeks to count only unplanned admissions in the measure outcome because variation in planned admissions does not reflect quality differences. The algorithm (see the flowchart in the Data Dictionary, first tab, “S.6 Planned Adm Alg Flowchart”) identifies inpatient admissions that are typically planned and may occur after the patient’s index general surgery procedure, considering a few specific, limited types of care as “planned”(e.g., major organ transplant, rehabilitation, or maintenance chemotherapy). Otherwise, the algorithm defines a planned admission as a non-acute inpatient admission for a scheduled procedure (e.g., total hip replacement or cholecystectomy), and the algorithm never considers inpatient admissions for acute illness or for

complications of care planned. The algorithm considers inpatient admissions that include potentially planned procedures with acute diagnoses, or with diagnoses that might represent complications of a surgery, as "unplanned" and thus counts these inpatient admissions in the measure outcome.

Details of the planned admission algorithm and codes to identify planned admissions are in the attached Data Dictionary sheet labeled "S.6 Planned Adm Alg."

Definition of ED Visits and Observation Stay

The measure defines ED visits and observation stays using one of the specified billing codes or revenue center codes identified in Medicare Part B Outpatient hospital claims.

The codes used to define ED visits and observation stays are in the attached Data Dictionary sheet labeled "S.6 Numerator-ED Obs Def."

Citations

1. Horwitz L, Grady J, Cohen D, et al. Development and validation of an algorithm to identify planned readmissions from claims data. Journal of Hospital Medicine. Oct 2015;10(10):670-677.

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

Target Population

Included patients:

The target population for this measure is Medicare FFS patients aged 65 years and older, who are undergoing outpatient general surgery procedures in ASCs that are within the scope of general surgery training. Specifically, the cohort of procedures includes the following types of surgeries: abdominal, alimentary tract, breast, skin/soft tissue, wound, and varicose vein.

The Medicare FFS population was chosen because of the availability of a national dataset (Medicare claims) that could be used to develop, test, and publicly report the measure. We limit the measure to patients who have been enrolled in Medicare FFS Parts A and B for the 12 months prior to the date of surgery to ensure that we have adequate data for identifying comorbidities for risk adjustment.

Included procedures:

The target group of procedures is surgical procedures that (1) are routinely performed at ASCs, (2) involve risk of post-surgery hospital visits, and (3) are within the scope of general surgery training. The scope of general surgery overlaps with that of other specialties (for example, vascular surgery and, plastic surgery). For this measure, we targeted surgeries that general surgeons are trained to perform with the understanding that other subspecialists may also be performing many of these surgeries at ASCs. Since the type of surgeon performing a particular procedure may vary across ASCs in ways that affect quality, the measure is neutral to surgeons' specialty training.

To identify eligible ASC general surgery procedures, we first identified a list of procedures from Medicare's 2014 and 2015 ASC lists of covered procedures, which include procedures for which ASCs can be reimbursed under the ASC payment system. This lists of surgeries is publicly available at: https://www.cms.gov/medicare/medicare-fee-for-service-payment/ascpayment/11_addenda_updates.html (download January 2014 and January 2015 ASC Approved HCPCS Code and Payment Rates, Addendum AA). Surgeries on the ASC list of covered procedures do not involve or require: major or prolonged invasion of body cavities, extensive blood loss, major blood vessels, or care that is either emergent or life-threatening. The ASC list is annually reviewed and updated by Medicare, and includes a transparent public comment submission and review process for addition and/or removal of procedure codes. Using an existing, defined list of surgeries, rather than defining surgeries de novo, is useful for long-term measure maintenance. Procedures listed in Medicare's list of covered ASC procedures are defined using Healthcare Common Procedure Coding System (HCPCS) and Common Procedural Terminology (CPT®) codes.

Ambulatory procedures include a heterogeneous mix of non-surgical procedures, minor surgeries, and more substantive surgeries. The measure is not intended to include very low-risk (minor) surgeries or non-surgical procedures, which typically have a high volume and a very low outcome rate. Therefore, to focus the measure only on the subset of surgeries on Medicare's list of covered ASC procedures that impose a meaningful risk of post-procedure hospital visits,

the measure includes only "major" and "minor" procedures, as indicated by the Medicare Physician Fee Schedule global surgery indicator (GSI) values of 090 and 010, respectively. The GSI code reflects the number of post-operative days that are included in a given procedure's global surgical payment and identifies surgical procedures of greater complexity and follow-up care. This list of GSI values is publicly available for calendar year (CY) 2014 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1600-FC.html and for CY 2015 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1612-FC.html (download PFS Addenda, Addendum B).

Finally, to identify the subset of general surgery ASC procedures, we reviewed with consultants and Technical Expert Panel (TEP) members the Clinical Classifications Software (CCS) categories of procedures developed by the Agency for Healthcare Research and Quality (AHRQ). We identified and included CCS categories within the scope of general surgery, and only included individual procedures within the CCS categories at the procedure (CPT® code) level if they were within the scope of general surgery practice. We did not include in the measure gastrointestinal endoscopy, endocrine, or vascular procedures, other than varicose vein procedures, because reasons for hospital visits are typically related to patients' underlying comorbidities.

See the attached Data Dictionary, sheet S.9 "Codes Used to Define Cohort" for a complete list of all CPT procedure codes included in the measure cohort.

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).*

Target Population

Included patients:

The target population for this measure is Medicare FFS patients aged 65 years and older, who are undergoing outpatient general surgery procedures in ASCs that are within the scope of general surgery training. Specifically, the cohort of procedures includes the following types of surgeries: abdominal, alimentary tract, breast, skin/soft tissue, wound, and varicose vein.

The Medicare FFS population was chosen because of the availability of a national dataset (Medicare claims) that could be used to develop, test, and publicly report the measure. We limit the measure to patients who have been enrolled in Medicare FFS Parts A and B for the 12 months prior to the date of surgery to ensure that we have adequate data for identifying comorbidities for risk adjustment.

Included procedures:

The target group of procedures is surgical procedures that (1) are routinely performed at ASCs, (2) involve risk of post-surgery hospital visits, and (3) are within the scope of general surgery training. The scope of general surgery overlaps with that of other specialties (for example, vascular surgery and, plastic surgery). For this measure, we targeted surgeries that general surgeons are trained to perform with the understanding that other subspecialists may also be performing many of these surgeries at ASCs. Since the type of surgeon performing a particular procedure may vary across ASCs in ways that affect quality, the measure is neutral to surgeons' specialty training.

To identify eligible ASC general surgery procedures, we first identified a list of procedures from Medicare's 2014 and 2015 ASC lists of covered procedures, which include procedures for which ASCs can be reimbursed under the ASC payment system. This lists of surgeries is publicly available at: https://www.cms.gov/medicare/medicare-fee-for-service-payment/ascpayment/11_addenda_updates.html (download January 2014 and January 2015 ASC Approved HCPCS Code and Payment Rates, Addendum AA). Surgeries on the ASC list of covered procedures do not involve or require: major or prolonged invasion of body cavities, extensive blood loss, major blood vessels, or care that is either emergent or life-threatening. The ASC list is annually reviewed and updated by Medicare, and includes a transparent public comment submission and review process for addition and/or removal of procedure codes. Using an existing, defined list of

surgeries, rather than defining surgeries de novo, is useful for long-term measure maintenance. Procedures listed in Medicare's list of covered ASC procedures are defined using Healthcare Common Procedure Coding System (HCPCS) and Common Procedural Terminology (CPT®) codes.

Ambulatory procedures include a heterogeneous mix of non-surgical procedures, minor surgeries, and more substantive surgeries. The measure is not intended to include very low-risk (minor) surgeries or non-surgical procedures, which typically have a high volume and a very low outcome rate. Therefore, to focus the measure only on the subset of surgeries on Medicare's list of covered ASC procedures that impose a meaningful risk of post-procedure hospital visits, the measure includes only "major" and "minor" procedures, as indicated by the Medicare Physician Fee Schedule global surgery indicator (GSI) values of 090 and 010, respectively. The GSI code reflects the number of post-operative days that are included in a given procedure's global surgical payment and identifies surgical procedures of greater complexity and follow-up care. This list of GSI values is publicly available for calendar year (CY) 2014 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1600-FC.html and for CY 2015 at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1612-FC.html (download PFS Addenda, Addendum B).

Finally, to identify the subset of general surgery ASC procedures, we reviewed with consultants and Technical Expert Panel (TEP) members the Clinical Classifications Software (CCS) categories of procedures developed by the Agency for Healthcare Research and Quality (AHRQ). We identified and included CCS categories within the scope of general surgery, and only included individual procedures within the CCS categories at the procedure (CPT® code) level if they were within the scope of general surgery practice. We did not include in the measure gastrointestinal endoscopy, endocrine, or vascular procedures, other than varicose vein procedures, because reasons for hospital visits are typically related to patients' underlying comorbidities.

See the attached Data Dictionary, sheet S.9 "Codes Used to Define Cohort" for a complete list of all CPT procedure codes included in the measure cohort.

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

The measure excludes surgeries for patients without 7 or more days of continuous enrollment in Medicare FFS Parts A and B after the surgery. The measure excludes these patients to ensure all patients have full data available for outcome assessment.

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.)*

Lack of 7 or more days of continuous enrollment in Medicare FFS after the ASC surgery is determined by patient enrollment status in FFS Parts A and B using the Medicare enrollment file (unless lack of enrollment was due to death). The procedure must be 7 or more days from the end of the month or the enrollment indicators must be appropriately marked for the month that falls within 7 days of the procedure date (unless disenrollment is due to death), otherwise the procedure is excluded.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Not applicable.

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

Statistical risk model

If other:

S.12. Type of score:

Ratio

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Lower 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.)*

The measure uses a two-level hierarchical logistic regression model to estimate ASC-level risk-standardized hospital visit ratios (RSHVRs). This approach accounts for the clustering of patients within ASCs and variation in sample size across ASCs. The RSHVR is calculated as the ratio of the predicted to the expected number of post-surgical unplanned hospital visits among ASC's patients. For each ASC, the numerator of the ratio is the number of hospital visits predicted for the ASC's patients, accounting for its observed rate, the number and complexity of general surgery procedures performed at the ASC, and the case mix. The denominator is the number of hospital visits expected nationally for the ASC's case/procedure mix. To calculate an ASC's predicted-to-expected (P/E) ratio, the measure uses a two-level hierarchical logistic regression model. The log-odds of the outcome for an index procedure is modeled as a function of the patient demographic, comorbidity, procedure characteristics, and a random ASC-specific intercept. A ratio greater than one indicates that the ASC's patients have more visits than expected, compared to an average ASC with similar patient and procedural complexity. A ratio less than one indicates that the ASC's patients have fewer post-surgical visits than expected, compared to an average ASC with similar patient and procedural complexity. This approach is analogous to an observed-to-expected ratio, but accounts for within-facility correlation of the observed outcome and sample size differences and accommodates the assumption that underlying differences in quality across ASCs lead to systematic differences in outcomes, and is tailored to and appropriate for a publicly reported outcome measure as articulated in published scientific guidelines [1-3].

Please see Appendix D of the attached technical report for details.

Citations

1. Normand S-LT, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. Statistical Science. 2007;22(2):206-226.

2. Krumholz HM, Brindis RG, Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. 2006;113(3):456-462.

3. National Quality Forum. Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. 2015; http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/2015_Measure_Evaluation_Criteria.aspx. Accessed July 26, 2016.

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.

Not applicable. This measure is not based on a sample or survey.

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.

Not applicable. This measure is not based on a sample or survey.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.

Claims, Enrollment 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.

Medicare administrative claims and enrollment data.

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)*

No data collection instrument provided

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

Facility

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

Outpatient Services

If other:

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.)*

Not applicable.

2. Validity – See attached Measure Testing Submission Form

Gen_Surg_ASC_NQF_Testing_Attachment_FINAL2_111917.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.

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.

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.

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

Measure Number (*if previously endorsed*)**:**
Measure Title: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers
Date of Submission: Click here to enter a date

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: Enrollment database and denominator files	☒ other: Enrollment database and denominator files

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 measure requires a data source that allows us to link patient data across care settings to identify appropriate surgical procedures for inclusion, comorbidities for risk adjustment, and the outcome of hospital visits [1-3]. Therefore, we used claims data, as they support these linkages and were available for the population of interest.

1. To develop and test the patient-level model, we used a national dataset of Calendar Year (CY) 2015 Medicare claims data from Health Account Joint Information (HAJI) database that included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims.

a. Datasets used to <u>define the cohort</u>:

-Outpatient general surgery procedures performed at Ambulatory Surgical Centers (ASCs) were identified using the full set of Medicare beneficiaries' claims from the CY 2015 Carrier non-institutional claims, which included the ASC facility claim (with a unique facility identifier).
-Enrollment database and denominator files: These datasets contain Medicare Fee-For-Service (FFS) enrollment, demographic, and death information for Medicare beneficiaries, which is used to determine inclusion criteria.

b. Datasets used to <u>capture the outcome</u> (hospital visits):

-The outcomes of emergency department (ED) visits and observation stays after general surgery ASC procedures were identified from the CY 2015 hospital outpatient institutional claims and inpatient hospital admissions from the CY 2015 inpatient institutional claims.

c. Datasets used to <u>identify comorbidities for risk adjustment</u>:

-Inpatient and outpatient claims (institutional and non-institutional carrier) data from the year prior (CY 2014) were used to identify comorbidities for risk adjustment for these patients.

2. To align with the Center for Medicare & Medicaid Services' (CMS's) intention to use more than 1 year of data for public reporting to ensure reliable estimates, we calculated ASCs' measure scores and the measure score reliability for a 2-year reporting period. Specifically, we used 2 years of claims data, which included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims for CY 2014 and CY 2015 from the HAJI database, to calculate ASCs' measure scores.

3. We used the American Community Survey data from the United States (US) Census Bureau (years 2009-2013) to derive the Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index for each zip code in the US. Other social risk factors were identified using enrollment and denominator files described above.

4. To calculate measure score reliability for a 2-year reporting period, we used a 4-year cohort of Medicare claims data from the HAJI database for CYs 2012-2015 (January 1, 2012 – December 31, 2015). We created two patient samples per facility that were equivalent in size to 2 years of data.

The datasets used for testing vary by testing type; see Section 1.7 for details.

References

1. Hosmer DW, Lemeshow S. Introduction to the logistic regression model. *Applied Logistic Regression, Second Edition.* 2000:1-30.
2. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988:837-845.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

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

We used Medicare FFS data from CYs 2011-2015. Years of data vary by testing type.

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 number of measured entities varied by testing type; see Section 1.7 for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of patients varied by testing type; see Section 1.7 for details.

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.

As described in Section 1.2, we used CY 2015 Medicare claims data from the HAJI database that included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims to develop the patient-level model, and CYs 2014-2015 to perform facility-level testing. The measure cohort inclusion and exclusion criteria are specified in the Measure Submission Form, Sections S.7 to S.9. The datasets, number of measured entities, number of general surgery procedures, and demographic profile for the patients used in each type of testing are as follows: 1. Medicare FFS CY 2015 Dataset -Dates: January 1, 2015 – December 31, 2015 -Number of facilities: 3,251 ASCs

-Number of general surgery procedures: 149,468
-Demographic characteristics: average age of 76.3 years; 45.73% female
-Dataset used for: defining the cohort, testing the exclusion criteria (Section 2b2.2), disparities testing (Section 2b3.4b)

2. Development Sample and Validation Sample

The 2015 Development and Validation Samples were derived by selecting two random samples from the Medicare FFS CY 2015 Dataset. The Development Sample included 50% of the general surgery ASC procedures in the Medicare FFS CY 2015 Dataset, and the Validation Sample included 50% of the general surgery ASC procedures in the Medicare FFS CY 2015 Dataset.

Development Sample

-Dates: January 1, 2015 – December 31, 2015

-Number of facilities: 2,966 ASCs
-Number of general surgery procedures: 74,734
-Demographic characteristics: average age of 76.3 years; 45.83% female
-Dataset used for: testing data element reliability (Section 2a2.3), testing the patient-level risk-adjustment model (Section 2b3.4a)

Validation Sample

-Dates: January 1, 2015 – December 31, 2015

-Number of facilities: 2,961 ASCs
-Number of general surgery procedures: 74,734
-Demographic characteristics: average age of 76.3 years; 45.62% female
-Dataset used for: testing data element reliability (Section 2a2.3), validating the patient-level risk adjustment model (Section 2a2.3), internal validation of the model (see Section 2b1.3)

3. Medicare FFS CYs 2014-2015 Dataset

-Dates: January 1, 2014 – December 31, 2015

-Number of facilities (with at least 25 cases): 1,642 ASCs

-Number of general surgery procedures (across ASCs with at least 25 cases): 286,999

-Demographic characteristics: average age of 76.4 years; 45.57% female

-Dataset used for: testing facility-level score distribution

4. Medicare FFS CYs 2012-2015 Dataset

-Dates: January 1, 2012 – December 31, 2015

-Number of facilities: 4,177 ASCs

-Number of general surgery procedures: 619,499

-Demographic characteristics: average age of 76.5 years; 46.18 % female

-Dataset used for: testing facility-level reliability

Note: For all cohorts defined above, we use 1 additional year of data (the year prior to the first year) to gather risk-adjustment variables for the patients undergoing procedures in the first year of the cohort (example: for dataset #4, we use calendar year 2011 data to gather risk factors for patients undergoing procedures in 2012).

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

As detailed below and in Section 2b3.4b, we considered two patient-level social risk factor variables (Medicaid dual-eligibility status and African-American race) and a composite measure of low SES (the AHRQ-SES index score). In

addition, we examined the facility-level proportions of dual-eligible patients, of African-American patients, and of low-SES patients. These analyses were performed with the Medicare FFS CYs 2014-2015 Dataset and data from the Census Bureau's American Community Survey.

We selected social risk factors to analyze after reviewing the literature and examining available national data sources. In the ambulatory surgery setting, studies have demonstrated higher risk of post-procedure hospital visits for African-American and Hispanic patients and for patients residing in lower-income households [1-4].

Potential pathways for SES and race variables' effects are described below in Section 2b3.3a.

The SES and race variables that we examined are:

- Dual-eligible status
- African-American race
- AHRQ-validated SES index score (summarizing the information from the following variables: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room)

In selecting variables, our intent was to be responsive to the National Quality Forum (NQF) guidelines for measure developers and the findings of recent work funded by the Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014 [3, 4]. Our approach was to examine patient-level indicators of both SES and race that are reliably available for all Medicare beneficiaries and linkable to claims data and to select those that have established validity.

Previous studies examining the validity of data on patients' race collected by CMS have shown that only the data identifying African-American beneficiaries have adequate sensitivity and specificity to be applied broadly in research or measures of quality. While this variable is not ideal because it groups all non-African-American beneficiaries together, it is currently the only race variable available on all beneficiaries across the nation that is linkable to claims data.

Similarly, we recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous measure. However, the eligibility threshold for over 65-year-old Medicare patients is valuable, as it considers both income and assets and is consistently applied across states. For both our race and dual-eligible variables, there is a body of literature demonstrating differential health care and health outcomes among beneficiaries indicating that these variables, while not ideal, allow us to examine some of the pathways of interest [3].

Finally, we selected the AHRQ-validated SES Index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas [5]. Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities in which patients live. We used data from the American Community Survey to create AHRQ SES Index scores at the census block group level and then mapped them to 9-digit ZIP codes via vendor software. The patient-level Medicare FFS claims data were then linked to the AHRQ SES Index scores by patients' ZIP codes. Given the variation in cost of living across the country, we adjusted the median income and median property value components of the AHRQ SES Index by regional price parity values published by the Bureau of Economic Analysis. This provided a better marker of low-SES neighborhoods in high-expense geographic areas.

Citations

1. Bhattacharyya N. Healthcare disparities in revisits for complications after adult tonsillectomy. *Am J Otolaryngol*. 2015 Mar-Apr;36(2):249-253.
2. Menachemi N, Chukmaitov A, Brown LS, et al. Quality of care differs by patient characteristics: outcome disparities after ambulatory surgical procedures. *Am J Med Qual*. 2007 Nov-Dec;22(6):395-401.
3. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation. Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2017.

4. National Academies of Sciences, Engineering, and Medicine (NASEM); *Accounting for Social Risk Factors in Medicare Payment: Data.* Washington DC: National Academies Press; 2016.

5. Bonito A, Bann C, Eicheldinger C, et al. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final report, sub-task. 2008;2.

2a2. RELIABILITY TESTING

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

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

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

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

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

Data Element Reliability

In constructing the measure in Medicare FFS patients, we aim to utilize only those data elements from claims data that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across ASCs. Specifically, we used fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies, and we seek to avoid variables which do not meet this standard.

In addition, CMS has in place several auditing programs used to assess overall claims coding accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential for payment.

Measure Score Reliability

We tested the reliability of the facility measure score by calculating the intra-class correlation coefficient (ICC) of the measure score. To calculate the ICC, we used the Medicare FFS CYs 2012-2015 Dataset. For ASCs with two or more general surgery procedures, these procedures were randomly split into the two samples within each facility. The ASCs with one procedure were randomly split into the two samples. The ICC evaluated the agreement between the risk-standardized hospital visit ratios (RSHVRs) calculated in the two randomly selected samples [1].

Citations

1. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

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*)

Data Element Reliability

Table 1: Risk Variable Frequencies, Development and Validation Samples (Medicare 100% FFS Cohort)

Variable (definition)	Development Sample (50%)		Validation Sample (50%)	
	#	%	#	%
N	-	-	-	-
Age: mean (standard deviation [SD])	76.3	7.2	76.3	7.2
Procedure type: Abdomen and its contents	9,506	12.7%	9,474	12.7%
Procedure type: Alimentary tract	4,941	6.6%	5,143	6.9%
Procedure type: Breast	5,089	6.8%	5,094	6.8%
Procedure type: Skin/soft tissue	41,334	55.3%	41,357	55.3%

Variable (definition)	Development Sample (50%)		Validation Sample (50%)	
	#	%	#	%
Procedure type: Wound	13,277	17.8%	13,087	17.5%
Procedure type: Vascular	587	0.8%	579	0.8%
Work Relative Value Units: mean (SD)	7	3.9	7	3.9
Comorbidities	-	-	-	-
Other benign tumors (CC 15, 16)	59,878	80.1%	59,906	80.2%
Liver or biliary disease (CC 27, 28, 29, 30, 31, 32)	6,621	8.9%	6,650	8.9%
Intestinal obstruction or perforation (CC 33)	1,482	2.0%	1,446	1.9%
Dementia or senility (CC 51, 52, 53)	5,611	7.5%	5,697	7.6%
Psychiatric disorders (CC 57, 58, 59, 60, 61, 62, 63)	15,913	21.3%	15,877	21.2%
Other significant central nervous system (CNS) disease (CC 77, 78, 79, 80)	2,698	3.6%	2,745	3.7%
Ischemic heart disease (CC 86, 87, 88, 89)	21,613	28.9%	21,373	28.6%
Specified arrhythmias and other heart rhythm disorders (CC 96, 97)	21,055	28.2%	21,047	28.2%
Stroke (CC 99, 100)	3,215	4.3%	3,273	4.4%
Chronic lung disease (CC 110, 111, 112, 113)	15,192	20.3%	14,976	20.0%
Pneumonia (CC 114, 115, 116)	4,910	6.6%	4,816	6.4%
Dialysis or sever chronic kidney disease (CC 134, 136, 137)	2,122	2.8%	1,990	2.7%
Benign prostatic hyperplasia (ICD-9 codes: 60000, 60001, 60020, 60021, 60090, 6091; ICD-10 codes: N40.0, N40.1, N40.2, N40.3)	14,499	19.4%	14,846	19.9%
Cellulitis, local skin infection (CC 164)	10,371	13.9%	10,541	14.1%
Major traumatic fracture or internal injury (CC 169, 170, 171, 172, 173, 174)	25,337	33.9%	25,389	34.0%
Complications of care (CC 176, 177)	6,083	8.1%	6,179	8.3%
Chronic anticoagulant use (ICD-9 code: V5861; ICD-10 code: Z7901 [long-term [19] use of anticoagulants])	7,671	10.3%	7,653	10.2%
Opioid abuse (ICD-9 codes: 30400, 30401, 30402, 30403, 30470, 30471, 30472, 30403, 30550, 30551, 30552, 30553; ICD-10: codes: F11.10, F11.120, F11.121, F11.122, F11.129, F11.14, F11.150, F11.151, F11.159, F11.181, F11.182, F11.188, F11.19, F11.20, F11.21, F11.220, F11.221, F11.222, F11.229, F11.23, F11.24, F11.250, F11.251, F11.259, F11.281, F11.282, F11.288, F11.29)	386	0.5%	345	0.5%

Measure Score Reliability

Testing measure score reliability yielded an ICC [2,1] of 0.530.

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?*)

Data Element Reliability Results

Table 1 above shows the frequencies across the two split samples for all variables included in the final model. As the results in Table 1 show, the frequencies of the risk variables were similar in the Development and Validation Samples, indicating good variable consistency and data element reliability.

Measure Score Reliability Results

The ICC [2,1] score of 0.530, calculated for four years of data, indicates moderate measure score reliability.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing**

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

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

We demonstrated measure validity through relevant prior validity testing we conducted for other claims-based measures, through the application of established measure development guidelines, and through assessment by external groups.

Validity of Claims-Based Measures

For several other NQF-endorsed measures, our team has demonstrated the validity of using claims data for risk adjustment in lieu of medical record data in estimating facility-level measure scores. CMS has validated six NQF-endorsed measures currently in public reporting (acute myocardial infarction [AMI], heart failure, and pneumonia mortality and readmission measures) with models that used medical record-abstracted data for risk adjustment. Specifically, we conducted claims model validation by building comparable models using abstracted medical record data for risk adjustment for AMI patients (Cooperative Cardiovascular Project data), heart failure patients (National Heart Failure data), and pneumonia patients (National Pneumonia Project dataset). When both models were applied to the same patient population, the hospital risk-standardized rates estimated using the claims-based risk-adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based models for public reporting. Our group has reported these findings in the peer-reviewed literature [1-6]. While the applicability of these findings to our measure may be limited because these medical record validations were focused on patients admitted for specific medical conditions, they nevertheless suggest claims data generally have an acceptable degree of agreement with clinical data at a facility level.

Validity Indicated by Established Measure Development Guidelines:

We developed this measure in consultation with national guidelines for publicly reported outcome measures, with input from outside experts and the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcome measures [7], CMS Measure Management System (MMS) guidance, and guidance articulated in the American Heart Association scientific statement entitled, “Standards for Statistical Models Used for Public Reporting of Health Outcomes” [8].

Validity as Assessed by External Groups:

Throughout the measure development process, we obtained expert and stakeholder input through holding regular discussions with external clinical consultants, consulting our national Technical Expert Panel (TEP), and holding a 20-day public comment period.

Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE) clinicians, as well as clinical experts in the field of surgery, met regularly to discuss all aspects of measure development, including the cohort, outcome definition, and risk adjustment.

In addition to the consultations and in alignment with CMS MMS guidance, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives, including clinicians,

patients, and individuals with expertise in quality improvement and performance measurement. We held two structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We made modifications to the measure specifications (e.g., cohort definition, risk adjustment) based on TEP feedback on the measure.

Additionally, we held a three-week public comment period to solicit input on the measure's methodology and preliminary specifications. We revised the measure in response to public comment and posted a summary of the comments received as well as the updates made to the measure (available in the Downloads section at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/PC-Updates-on-Previous-Comment-Periods.html). This NQF application includes the measure's final specifications, inclusive of the revisions after consideration of the public comments.

Face Validity as Determined by the TEP:

We systematically assessed the face validity of the measure score as an indicator of quality by confidentially soliciting the TEP members' agreement with the following two statements via an online survey following the final TEP meeting:

1. Please rate the following statement on a scale of 1 (strongly agree) to 6 (strongly disagree): "The risk-standardized hospital visit rates obtained from the 'Hospital Visits after General Surgery Ambulatory Surgical Center Procedures' measure as specified are valid and useful measures of ASC general surgical quality of care."

2. Please rate the following statement on a scale of 1 (strongly agree) to 6 (strongly disagree): "The risk-standardized hospital visit rates obtained from the 'Hospital Visits after General Surgery Ambulatory Surgical Center Procedures' measure as specified will provide ASCs with information that can be used to improve their quality of care."

List of TEP Members

1) Robin Blomberg, BA, MA – National Forum of End-Stage Renal Disease, Network 16 (Representative for Kidney Patient Advisory Council); Seattle, WA
2) Kirk Campbell, MD – New York University Hospital for Joint Diseases (Clinical Assistant Professor of Orthopedic Surgery); New York, NY
3) Gary Culbertson, MD, FACS – Iris Surgery Center (Surgeon; Medical Director); Sumter, SC
4) Martha Deed, PhD – Consumers Union Safe Patient Project (Patient Safety Advocate); Austin, TX
5) James Dupree, MD, MPH – University of Michigan (Urologist; Health Services Researcher); Ann Arbor, MI
6) Nester Esnaola, MD, MPH, MBA – Fox Chase Cancer Center (Professor of Surgery; Associate Director for Cancer Health Disparities and Community Engagement); Philadelphia, PA
7) John Gore, MD, MS – University of Washington (Associate Professor of Urology); Seattle, WA
8) Lisa Ishii, MD, MHS – Johns Hopkins School of Medicine (Associate Professor); American Academy of Otolaryngology-Head and Neck Surgery (Coordinator for Research and Quality); Baltimore, MD; Alexandria, VA
9) Atul Kamath, MD – Perelman School of Medicine, University of Pennsylvania (Assistant Professor and Clinical Educator Director of Orthopedic Surgery); Hospital of the University of Pennsylvania (Attending Surgeon); Philadelphia, PA
10) Tricia Meyer, PharmD, MS, FASHP – Scott & White Medical Center (Regional Director of Pharmacy); Texas A&M University College of Medicine (Associate Professor of Anesthesiology); Temple, TX
11) Linda Radach, BA – Consumers Union Safe Patient Project (Patient Safety Advocate); Austin, TX
12) Amita Rastogi, MD, MHA, CHE, MS – Health Care Incentives Improvement Institute (Chief Medical Officer); Newtown, CT
13) Donna Slosburg, RN, BSN, LHRM, CASC – ASC Quality Collaboration (Executive Director); St. Pete Beach, FL
14) Julie Thacker, MD, FACS – Duke Health and Hospital System (Medical Director of Evidence-Based Perioperative Care); Duke School of Medicine Clinical Research Unit (Medical Director, Department of Surgery); Durham, NC
15) Thomas Tsai, MD, MPH – Brigham and Women's Hospital (General Surgeon); Harvard School of Public Health (Research Associate); Boston, MA

Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes

This application includes ICD-10 codes that correspond to all International Classification of Diseases, Ninth Revision (ICD-9) codes included in the specifications. The goal was to convert this measure into a new code set, fully consistent with the intent of the original measure.

ICD-10 diagnosis and procedure codes used to define the Planned Admission Algorithm were identified from the 2015 version of the AHRQ Clinical Classification Software (CCS) categories specified for ICD-10, followed by clinician review. The algorithm also includes some individual ICD-9 codes. To create the crosswalk for the ICD-9-level codes, we used the 2015 ICD-9-CM to ICD-10-CM General Equivalence Mappings tool, made available by CMS, followed by team review.

Citations

1. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*. 2006 Apr 4;113(13):1683-92.
2. Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. 2011 Mar 1;4(2):243-52.
3. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*. 2006 Apr 4;113(13):1693-701.
4. Keenan PS, Normand S-LT, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2008 Sep;1(1):29-37.
5. Bratzler DW, Normand S-LT, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLoS One*. 2011 Apr 12;6(4):e17401.
6. Lindenauer PK, Normand S-LT, Drye EE, et al. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. *J Hosp Med*. 2011 Mar;6(3):142-50.
7. National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.
8. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: An American Heart Association scientific statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council endorsed by the American College of Cardiology Foundation. *Circulation.* 2006;113(3):456-462.

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

Face Validity as Determined by the TEP:

14 out of the 15 TEP members responded to the face validity survey. Of the 14 respondents, 12 respondents indicated that they somewhat, moderately, or strongly agreed; and two respondents moderately disagreed with the following two statements:

1. "The risk-standardized hospital visit rates obtained from the Hospital Visits after General Surgery Ambulatory Surgical Center Procedures ASC measure, as specified, are valid and useful measures of ASC general surgical quality of care."
2. "The risk-standardized hospital visit rates obtained from the Hospital Visits after General Surgery Ambulatory Surgical Center Procedures' measure, as specified, will provide ASCs with information that can be used to improve their quality of care."

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?*)

These validity testing results demonstrate TEP agreement with the overall face validity of the measure.

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*)

We determined the single exclusion criterion to be appropriate based on clinical considerations. We examined the overall frequency and proportion of the total cohort excluded for the single exclusion criterion.

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*)

Applying our inclusion criteria (general surgery procedures, including abdomen and its content, alimentary tract, breast, skin/soft tissue, wound, and varicose vein procedures performed on patients aged ≥65 enrolled in Medicare FFS Parts A and B in the 12 months prior to the date of surgery) to the Medicare FFS CY 2015 Dataset resulted in an initial cohort of 149,512 ASC general surgery procedures. We then applied the following exclusion criterion (see the Measure Submission Form, Sections S.8 and S.9, for exclusion rationale): Excluded surgeries for patients who survived at least 7 days, but were not continuously enrolled in Medicare FFS Parts A and B within 7 days of the general surgery ASC procedure.

This resulted in excluding 44 (0.03%) general surgery procedures performed at ASCs. Thus, the final Medicare FFS CY 2015 Dataset included 149,468 general surgery ASC procedures performed at 3,251 ASCs. Given the few cases affected, we did not examine the distribution of cases across ASCs or the effect of the exclusion on the 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*)

We exclude surgeries for patients without continuous enrollment in Medicare FFS Parts A and B within 7 days of the general surgery ASC procedure. This exclusion is narrowly targeted and necessary to ensure all patients have full data available for outcome assessment. This exclusion criterion removes a small number (0.03%) of general surgery ASC procedures.

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 21 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

The measure uses a two-level hierarchical logistic regression model to estimate ASC-level RSHVRs. This approach accounts for the clustering of patients within ASCs and variation in sample size across ASCs.

The stepwise selection procedure identified age, 17 comorbidities, work Relative Value Units (RVUs) to adjust for surgical procedural complexity, procedure type (abdomen vs. alimentary tract vs. breast vs. skin/soft tissue vs. wound vs. varicose vein), and one interaction term. For the final model, we retained these variables and one variable (opioid use) that had a p-value of 0.0917 because experts advised it was an important risk predictor and expressed a strong preference for including it in the model. Work RVUs are assigned to each Current Procedural Terminology (CPT®) procedure code and approximate surgical procedural complexity by incorporating elements of physician time and effort. For patients with multiple concurrent CPT procedure codes, we risk adjust for the CPT code with the highest Work RVU value.

Model Variables:

1. Age
2. Procedure Type: Abdomen and its contents
3. Procedure Type: Alimentary tract
4. Procedure Type: Breast
5. Procedure Type: Skin/soft tissue

6. Procedure Type: Wound
7. Procedure Type: Varicose vein
8. Work Relative Value Units
9. Other benign tumors (CC 15, 16)
10. Liver or biliary disease (CC 27, 28, 29, 30, 31)
11. Intestinal obstruction or perforation (CC 33)
12. Dementia or senility (CC 51, 52, 53)
13. Psychiatric disorders (CC 57, 58, 59, 60, 61, 62, 63)
14. Other significant central nervous system (CNS) disease (CC 77, 78, 79, 80)
15. Ischemic heart disease (CC 86, 87, 88, 89)
16. Specified arrhythmias and other heart rhythm disorders (CC 96, 97)
17. Stroke (CC 99, 100)
18. Chronic lung disease (CC 110, 111, 112, 113)
19. Pneumonia (CC 114, 115, 116)
20. Dialysis or sever chronic kidney disease (CC 134, 136, 137)
21. Benign prostatic hyperplasia (ICD-9 codes: 60000, 60001, 60020, 60021, 60090, 6091; ICD-10 codes: N40.0, N40.1, N40.2, N40.3)
22. Cellulitis, local skin infection (CC 164)
23. Major traumatic fracture or internal injury (CC 169, 170, 171, 172, 173, 174)
24. Complications of care (CC 176, 177)
25. Chronic anticoagulant use (ICD-9 code: V5861; ICD-10 code: Z7901 [long-term {current} use of anticoagulants])
26. Opioid abuse (ICD-9 codes: 30400, 30401, 30402, 30403, 30470, 30471, 30472, 30403, 30550, 30551, 30552, 30553; ICD-10: codes: F11.10, F11.120, F11.121, F11.122, F11.129, F11.14, F11.150, F11.151, F11.159, F11.181, F11.182, F11.188, F11.19, F11.20, F11.21, F11.220, F11.221, F11.222, F11.229, F11.23, F11.24, F11.250, F11.251, F11.259, F11.281, F11.282, F11.288, F11.29)27
27. Procedure type*RVU

2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable. This measure is risk adjusted.

2b3.3a. Describe the conceptual/clinical <u>and</u> 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?

Our approach to risk adjustment is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines [1,2]. For example, we only adjust for risk factors that are present at the start of care. We do not risk adjust for conditions that are possible adverse events of care and that are only recorded at the time of the surgery (see Data Dictionary, Sheet 2b4.3 Risk Model Specs). We do not adjust for factors related to the delivery of care that may reflect care quality.

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create an ASC-level 7-day RSHVR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within facilities), the underlying risk due to patients' procedures/comorbidities, and sample size at a given ASC when estimating hospital visit ratios. In brief, the approach simultaneously models two levels (patient and facility) to account for the variance in patient outcomes within and between facilities [2]. At the patient level, the model adjusts the log-odds of hospital visits within 7 days after the procedure for selected demographic, clinical, and procedure risk variables. The second level models the facility-specific intercepts as arising from a normal distribution.

The facility intercept, or facility-specific effect, represents the ASC contribution to the risk of 7-day hospital visits, after accounting for patient risk and sample size, and can be inferred as a measure of quality. If there were no differences among ASCs, then after adjusting for patient risk, the facility intercepts would be identical across all ASCs.

Candidate Risk-Adjustment Variables:

The measure adjusts for differences in patient comorbidities, demographics, and in procedure-related differences in risk across ASCs. We identified potential candidate risk factors through: 1) prior work on related quality measures (including the related urology and orthopedic ASC measures); 2) a focused literature review; and 3) TEP and expert input.

Candidate risk factors identified from work on related measures included opioid abuse, chronic anticoagulant use, tobacco use disorder, benign prostatic hyperplasia, morbid obesity, Work RVU, number of qualifying procedures, and procedure type. We used work RVU of the procedure to address surgical procedural complexity, an approach employed by the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) [3].

To identify additional clinical and procedural risk factors, we searched the literature for relevant peer-reviewed publications of variables that predicted hospital visits after outpatient general surgery procedures using Ovid MEDLINE® and PubMed. The search yielded a total of 138 studies potentially relevant to the general surgery measure. Of these studies, 131 were excluded after review of the abstract, and 3 were excluded after full-text review. We added variables identified in the literature to our list of candidate risk factors if they were significantly associated with unplanned hospital visits in bivariate or multivariable analyses at the 0.05 level. From the 4 studies, we identified two variables not already included: anesthesia type and operating time [4-5]. However, we did not include anesthesia type or operating time because we do not risk adjust for discretionary procedure differences (such as approach to anesthesia or surgical techniques).

To define the clinical risk factors in claims data, we used CMS's Version 22 Hierarchical Condition Categories (HCCs) to operationalize the candidate clinical comorbidities. The HCCs classify 68,000 ICD-10-CM and over 15,000 ICD-9-CM diagnosis codes into clinically coherent condition categories. Then, to consolidate similar risk factors into fewer, broader risk variables, we first examined their frequency, bivariate direction and strength of association with the outcome of the individual risk factors defined by condition categories or ICD-10-CM codes, and then combined risk factor diagnoses into clinically coherent comorbidity variables. For example, we created a "cancer" variable that combined several individual cancer diagnoses.

Our expert clinical consultants and the TEP reviewed this preliminary list of risk variables and suggested additional variables: failure to thrive (poor nutritional status), history of falling, sleep apnea, and history of steroid use. We added all suggested candidate variables; the final list included 80 candidate risk variables.

Variable Selection

To select the final set of variables to include in the risk-adjustment model, risk variables were entered into logistic regression analyses predicting the outcome of hospital visits within 7 days in the Development Sample. The Development Sample is a randomly selected 50% sample of our CY 2015 Medicare cohort. To develop a parsimonious risk model, non-significant variables were iteratively removed from the model using a stepwise, purposeful selection approach described by Hosmer and Lemeshow [6]. All variables significant at $p<0.05$ were retained in the final model. The attached Data Dictionary sheet labeled "2b4.3 Risk Model Specs" indicates the final risk variables selected, the codes used to define the risk variables for our statistical model, and their odds ratios and 95% confidence intervals (CIs).

Social Risk Factors for supplementary disparities analyses

We selected variables representing SES factors and race based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we considered and analyzed based on this review. Below, we describe the pathways by which SES and race may influence risk of hospital visits following outpatient surgical procedures.

Our conceptualization of the pathways by which patient SES or race affects the outcome is informed by the literature [7-12] and IMPACT Act–funded work by the National Academies of Sciences, Engineering and Medicine (NASEM) and the Department of Health and Human Services Assistant Secretary for Planning and Evaluation (ASPE) [13-15].

Literature Review of SES and Race Variables and Ambulatory Surgery Post-Procedure Hospital Visits

To examine the relationship between SES and race variables and risk of hospital visits following outpatient surgical procedures, a literature search was performed with the following exclusion criteria: non-English language articles, articles published more than 10 years ago, articles without primary data, articles focused on pediatric patient population, and articles not explicitly focused on SES or race and hospital visits after ambulatory surgery. A total of 176 studies were reviewed by title and abstract, and all but two studies were excluded from full-text review based on the above criteria. The two studies indicated that African-American and Hispanic patients and patients from lower-income households were at increased risk of post-procedure hospital visits in the ambulatory surgery setting [7-8]. No studies were found that suggested that variation in patients' SES and race affected variation in outcome risk across facilities performing ambulatory surgical procedures.

Conceptual Pathways for SES and Race Variable Selection

Although there is limited literature linking social risk factors and adverse outcomes, potential pathways may include:

1. **Differential care within an ASC or unmet differential needs.** One pathway by which SES factors or race may contribute to hospital visit risk is that patients may not receive equivalent care within a facility. In the hospital setting, African-American patients have been shown to experience differential, lower quality, or discriminatory care [9]. Alternatively, patients with SES risk factors, such as lower education, may require differentiated care – e.g., provision of information at a lower health literacy level – that they do not receive.
2. **Use of lower-quality facilities.** Patients may differentially obtain care in lower quality ASCs. With respect to hospital care, patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high-quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients. Thus, patients with low income are more likely to be seen in lower-quality hospitals, which can contribute to increased risk of adverse outcomes following hospitalization [10-11]. Similarly, African-American patients have been shown to have less access to high-quality hospitals compared with white patients [12]. It is unknown to what extent this may be true in the ambulatory surgery setting.
3. **Influence of SES on hospital visit risk outside of ASC quality.** Some SES risk factors, such as income or wealth, may affect the likelihood of post-procedure hospital visits without directly being associated with the quality of care received at the ASC. For instance, while an ASC may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-procedure due to a limited understanding of the discharge plan or a lack of home support, transportation or other resources for following it fully.

As indicated in Section 1.8, the SES and race variables that we examined are:

- Dual-eligible status
- African-American race
- AHRQ-validated SES index score

The description of the analyses related to social risk factors can be found in Section 2b3.4b below.

Citations

1. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: An American Heart Association scientific statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council endorsed by the American College of Cardiology Foundation. *Circulation*. 2006;113(3):456-462.
2. Normand S-LT, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci.* 2007;22(2):206-226.
3. Raval MV, Cohen ME, Ingraham AM, et al. Improving American College of Surgeons National Surgical Quality Improvement Program risk adjustment: incorporation of a novel procedure risk score. *J Am Coll Surg.* 2010;211(6):715-723
4. Fleisher LA, Pasternak LR, Lyles A. A novel index of elevated risk of inpatient hospital admission immediately following outpatient surgery. *Arch Surg.* 2007;142(3):263-268.
5. Mioton LM, Buck DW, 2nd, Rambachan A, Ver Halen J, Dumanian GA, Kim JY. Predictors of readmission after outpatient plastic surgery. *Plast Reconstr Surg.* 2014;133(1):173-180.
6. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley; 2000.

7. Bhattacharyya N. Healthcare disparities in revisits for complications after adult tonsillectomy. *Am J Otolaryngol*. 2015 Mar-Apr;36(2):249-253.
8. Menachemi N, Chukmaitov A, Brown LS, et al. Quality of care differs by patient characteristics: outcome disparities after ambulatory surgical procedures. *Am J Med Qual*. 2007 Nov-Dec;22(6):395-401.
9. Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. *New Engl J Med*. 2014; 371:2298-2308.
10. Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and Medicaid patients. *Health Aff*. 2011; 30:1904-1911.
11. Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. *JAMA Surg.* 2014; 149:475-481.
12. Skinner J, Chandra A, Staiger D, et al. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. *Circulation* 2005; 112:2634-2641.
13. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation. Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2017.
14. National Academies of Sciences, Engineering, and Medicine (NASEM); *Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors.* Washington DC: National Academies Press; 2016.
15. National Academies of Sciences, Engineering, and Medicine (NASEM);. *Accounting for Social Risk Factors in Medicare Payment: Data.* Washington DC: National Academies Press; 2016.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

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

The following candidate variables were significant at $p<0.05$ and were retained in the final model:
1. Age
2. Procedure Type: Abdomen and its contents
3. Procedure Type: Alimentary tract
4. Procedure Type: Breast
5. Procedure Type: Skin/soft tissue
6. Procedure Type: Wound
7. Procedure Type: Varicose vein
8. Work Relative Value Units
9. Other benign tumors (CC 15, 16)
10. Liver or biliary disease (CC 27, 28, 29, 30, 31)
11. Intestinal obstruction or perforation (CC 33)
12. Dementia or senility (CC 51, 52, 53)
13. Psychiatric disorders (CC 57, 58, 59, 60, 61, 62, 63)
14. Other significant central nervous system (CNS) disease (CC 77, 78, 79, 80)
15. Ischemic heart disease (CC 86, 87, 88, 89)
16. Specified arrhythmias and other heart rhythm disorders (CC 96, 97)
17. Stroke (CC 99, 100)
18. Chronic lung disease (CC 110, 111, 112, 113)
19. Pneumonia (CC 114, 115, 116)

20. Dialysis or sever chronic kidney disease (CC 134, 136, 137)
21. Benign prostatic hyperplasia (ICD-9 codes: 60000, 60001, 60020, 60021, 60090, 6091; ICD-10 codes: N40.0, N40.1, N40.2, N40.3)
22. Cellulitis, local skin infection (CC 164)
23. Major traumatic fracture or internal injury (CC 169, 170, 171, 172, 173, 174)
24. Complications of care (CC 176, 177)
25. Chronic anticoagulant use (ICD-9 code: V5861; ICD-10 code: Z7901 [long-term {current} use of anticoagulants])
26. Opioid abuse (ICD-9 codes: 30400, 30401, 30402, 30403, 30470, 30471, 30472, 30403, 30550, 30551, 30552, 30553; ICD-10: codes: F11.10, F11.120, F11.121, F11.122, F11.129, F11.14, F11.150, F11.151, F11.159, F11.181, F11.182, F11.188, F11.19, F11.20, F11.21, F11.220, F11.221, F11.222, F11.229, F11.23, F11.24, F11.250, F11.251, F11.259, F11.281, F11.282, F11.288, F11.29)
27. Procedure type*RVU

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.**

Methods

To examine the impact of social risk factors on the measure calculation, we evaluated three indicators of social risk: 1) Medicare-Medicaid dual eligibility, 2) race, and 3) the AHRQ SES Index. For these analyses we used 100% Medicare FFS claims data from CYs 2014-2015. These data included 3,653 ASC facilities and 303,220 general surgery procedures. Our goal for these analyses were twofold: 1) to examine whether these factors were associated with increased risk in hospital visits after adjusting for other risk factors and 2) to evaluate the impact of social risk factors on ASC-level measure scores.

To evaluate the association of these risk factors with the outcome, we first quantified the observed rate by each group (dual-eligible: yes vs. no, race: African-American vs. all others, AHRQ SES Index: lowest quartile of SES Index vs. all others). We next evaluated the magnitude of association of these social risk factors with the outcome after adjustment for clinical comorbidities, procedure type, and age by including each individual indicator as a variable in our risk-adjustment model. Each factor's effect was quantified using odds ratios (ORs) and tested for significance. In addition, we evaluated the change in each model's predictive ability (c-statistic).

To evaluate the impact of social risk factors on the ASC-level measure scores, we compared RSHVRs calculated with and without each disparity marker included in the model. For these analyses, we calculated the RSHVR difference for each ASC (RSVHR with social risk variable – RSHVR without social risk variable) and calculated Pearson correlation coefficients for the paired scores.

We further examined the potential impact of these social risk factors on measure scores by comparing RSHVR distributions using current specifications. ASCs were stratified by the proportion of patients at the ASC with each factor, and placed into quartiles based on these proportions. For example, ASCs with few dual-eligible beneficiaries in their sample would be in the first quartile while ASCs seeing high numbers of dual-eligible beneficiaries would be in the fourth quartile. These stratified distributions were examined for systematic differences in RSHVR across quartiles.

Results

Observed hospital visit rates were higher for patients with each disparity marker: 3.7% for dual-eligible patients compared to 2.2% for non-dual-eligible patients, 3.1% for African-American patients compared to 2.2% for non-African-American patients, and 2.7% for low SES patients (scores below 42.7 on the AHRQ SES Index) compared to 2.2% for higher SES patients (scores above 42.7 on the AHRQ SES index). Furthermore, inclusion of each of these risk factors in our models indicated a statistically significant association after controlling for other risk-adjusters in our model (dual-eligible: OR: 1.34, 95% CI: 1.22 -1.48, $p < 0.0001$; race: OR: 1.23, 95% CI: 1.06-1.42, $p=0.005$; AHRQ SES Index: OR: 1.14, 95% CI: 1.06-1.22, $p=0.0004$).

However, results of examining the impact of social risk factors on the ASC-level measure scores indicated that entering these variables into the risk-adjustment model did not improve model performance (c-statistics remained unchanged) and did not substantially change ASC-level measure scores. Correlation coefficients between RSHVRs with and without adjustment for these factors were near 1 (0.998, 1.000, and 0.999 for dual-eligible, African-American, and low SES patients, respectively) and mean differences in RSHVRs were near zero (0.0000, -0.0001, and -0.0002 for dual-eligible, African-American, and low SES patients, respectively).

Further, the analyses of ASCs stratified into quartiles based on proportions of dual-eligible, African-American, and low SES patients (as identified by the AHRQ SES Index) showed largely overlapping distributions of the RSHVRs by quartile. The median RSHVR was 1.0 for all three variables except for ASCs with a low % of dual-eligible patients (1st quartile) whose median RSHVR was 0.9. Longer tails at the upper ends of the distributions were observed for ASCs with the highest percent of patients with the social risk factor (4th quartile). Distributions for low % of social risk factor ASCs (1st quartile) and high % social risk factor ASCs (4th quartile) by each social risk factor are shown in below in Table 2.

Based on these analyses, we conclude that although the three social risk factors we examined have a modest but statistically significant association with the risk of a hospital visit, these patient-level factors have a limited effect on the ASC-level measure scores. We did not adjust the models for these social risk factors since the association of these factors with the outcome may be quality-related, and since these factors have a limited relationship to the facility-level scores.

Table 2. Variation in RSHVRs across ASCs grouped into quartiles by proportion of Medicaid dual-eligible, African-American race, and Low SES patients

Social Risk Factor	Medicaid Dual Eligible		African-American Race		Low AHRQ SES Index Score	
Proportion of ASC patients	1st Quartile (≤1.82%)	4th Quartile (≥7.06%)	1st Quartile (0%)	4th Quartile (≥3.95%)	1st Quartile (≤4.04%)	4th Quartile (≥17.17%)
# of ASCs	409	411	599	410	410	410
# of patients	83,214	48,895	79,947	48,318	71,841	63,945
100% Max	1.6	1.9	1.9	2.1	1.8	1.9
90%	1.2	1.3	1.2	1.3	1.2	1.3
75% Q3	1.1	1.1	1.1	1.1	1.1	1.1
50% Median	0.9	1.0	1.0	1.0	1.0	1.0
25% Q1	0.9	0.9	0.9	0.9	0.9	0.9
10%	0.8	0.8	0.8	0.9	0.8	0.8
0% Min	0.5	0.4	0.6	0.4	0.5	0.6

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

To assess performance of the patient-level risk-adjustment model in the Development Sample, the area under the receiver operating characteristic curve as measured by the c-statistic was calculated. Observed hospital visit rates were compared to predicted hospital visit probabilities across predicted risk deciles to assess calibration, and the range of

observed hospital visit rates between the lowest and highest predicted risk deciles was also calculated to assess model discrimination.

Several analyses to validate the patient-level risk-adjustment model were performed. First, we compared model performance in the Development Sample with its performance in the Validation Sample. The c-statistic, and model predictive ability (discrimination) were compared. Second, we examined the stability of the risk variable frequencies and regression coefficients across the Development and Validation Samples. Third, we calculated over-fitting indices in the Validation Sample. Over-fitting refers to the phenomenon in which a model describes the relationship between predictive variables and outcome well in the development dataset but fails to provide valid predictions in new patients in the validation dataset. Estimated calibration values of γ0 far from 0 and estimated values of γ1 far from 1 provide evidence of over-fitting.

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

Development Sample results:

c-statistic=0.699
Predictive ability (hospital visit % in lowest decile, hospital visit % in highest decile): 0.79%-6.39%

Validation Sample results:

c-statistic=0.700
Predictive ability (hospital visit % in lowest decile, hospital visit % in highest decile): 0.71%-6.44%

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

Development Sample results:

Calibration: (0, 1)

Validation Sample results:

Calibration: (-0.08, 0.98)

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

Below are plots of observed vs. predicted values for the hospital visit rate across deciles of patient risk in the Development Sample (Figure 1) and Validation Sample (Figure 2). The plots, which showed that the predicted risk closely approximated the observed risk in most deciles, suggest reasonable calibration.

Figure 1. Calibration plot of predicted versus observed outcomes across deciles of patient risk in the Development Sample (data source: Medicare FFS CY 2015 dataset)

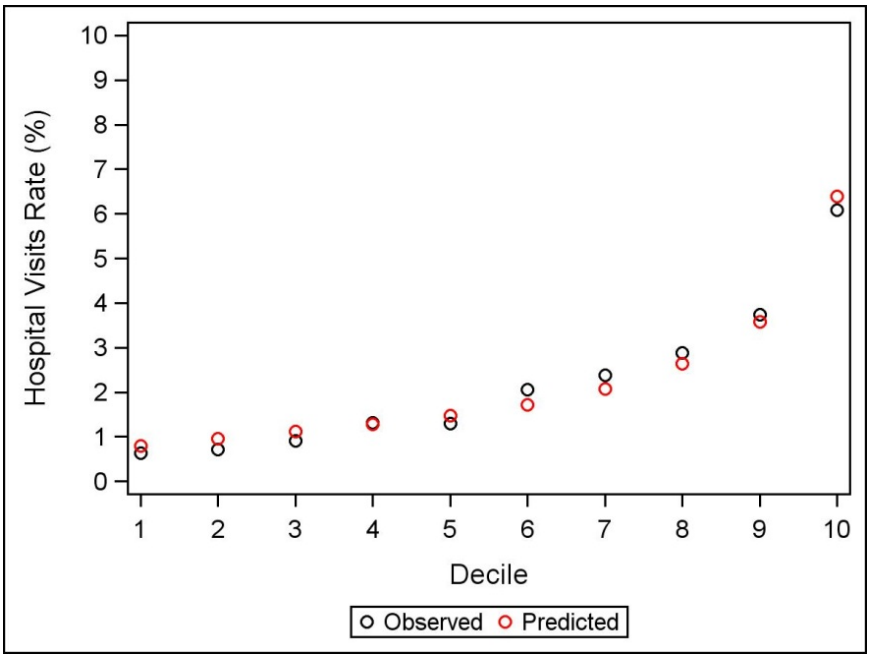
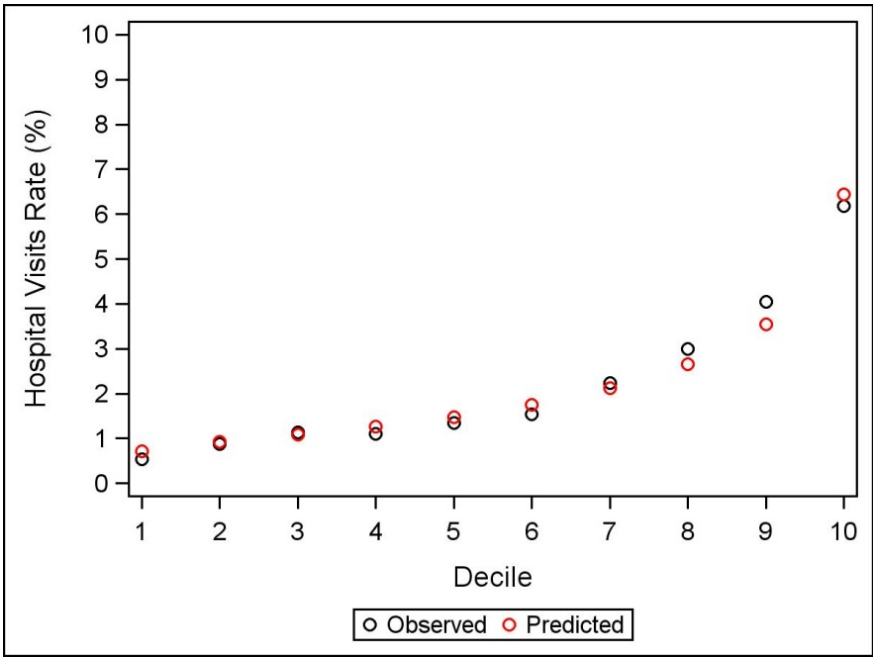


Figure 2. Calibration plot of predicted versus observed outcomes across deciles of patient risk in the 2015 Validation Sample (data source: Medicare FFS CY 2015 dataset)



2b3.9. Results of Risk Stratification Analysis:

Not applicable. This measure is not risk-stratified.

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*)

The c-statistics in the Development Sample and the Validation Sample were 0.699 and 0.700, respectively, showing good discrimination. The risk decile plots, which showed that the predicted risk closely approximated the observed risk across deciles, suggest good model calibration. The predicted unplanned hospital visit rate in the Development Sample ranged from 0.79% in the lowest decile to 6.39% in the highest predicted risk decile, a range of 5.6%; comparable results were

found in the Validation Sample. In addition, the regression coefficients of the model variables were stable across the Development and Validation Samples.

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*)

We tested interaction terms and retained those that were both significant at p<0.01 and clinically correlated with the outcome.

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)*

The measure score is an ASC-level RSHVR. The RSHVR is calculated as the ratio of the predicted to the expected number of post-surgical unplanned hospital visits among an ASC's patients. For each ASC, the numerator of the ratio is the number of hospital visits predicted for the ASC's patients, accounting for its observed rate, the number and complexity of general surgery procedures performed at the ASC, and the patient mix. The denominator is the number of hospital visits expected nationally for the ASC's case/procedure mix. To calculate an ASC's predicted-to-expected (P/E) ratio, the measure uses a two-level hierarchical logistic regression model. The log-odds of the outcome for an index procedure is modeled as a function of the patient demographic, comorbidity, procedure characteristics, and a random ASC-specific intercept. A ratio >1 indicates that the ASC's patients have more hospital visits than expected, compared to an average ASC with similar patient and procedural complexity. A ratio <1 indicates that the ASC's patients have fewer post-surgical visits than expected, compared to an average ASC with similar patient and procedural complexity.

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*)

The risk-standardized measure scores estimated using two full years of Medicare FFS data (2014 and 2015) showed variation across ASCs (Range: Min 0.42 to Max 2.13).

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?*)

These results suggest there are meaningful differences in the quality of care provided to patients undergoing general surgery procedures at ASCs.

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

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

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

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*)

Items 2b5.1-2b5.3 are not applicable, as this measure has only one set of specifications.

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*)

Items 2b5.1-2b5.3 are not applicable, as this measure has only one set of specifications.

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*)

Items 2b5.1-2b5.3 are not applicable, as this measure has only one set of specifications.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

Not applicable.

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*)

Not applicable.

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

Not applicable.

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.

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

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).

Not applicable.

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.

Not applicable. Measure is not currently in use. However, measure development and testing show that the measure cohort can be defined and outcomes can be reported using routinely collected Medicare claims data.

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)*.

Not applicable. There are no fees, licensing, or other requirements to use any aspect of the measure as specified.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

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

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Payment Program Not in use	

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

Not applicable. Measure is not yet in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because the measure is still under development and is now being submitted to the National Quality Forum (NQF) for initial endorsement.

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.*)

The measure may ultimately be used in one or more CMS programs, such as the Ambulatory Surgical Center Quality Reporting program.

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.

During development of the measure, we recruited and met with a national TEP, and CMS hosted a public comment. CMS solicited public comments on the measure, and we took all comments into consideration, addressing them individually. Therefore, performance results and data were provided to members of the TEP and then made public through public comment. TEP members and commenters included representatives of the measured entities (ASCs). The exact number of measured entities (ASCs) varies with each measurement period. In the Medicare FFS 2015 Dataset we used for measure development, there were 149,468 general surgery procedures performed at 3,251 ASCs. In the Medicare FFS CYs 2014-2015 Dataset we used for calculating ASC-level measures, there were 286,999 general surgery procedures performed at 1,642 ASCs (with at least 25 cases). (See section 1.7 of Measure Testing Form for a complete description of the number of measure entities).

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.

We provided data and results to the TEP and obtained TEP input on five occasions throughout the measure development process. We hosted two teleconference meetings with the TEP, solicited TEP input via email on the risk model, and provided measure updates to the TEP twice via email in response to public comments we received on the measure.

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.

Not applicable; the measure has not yet been implemented. Feedback during development as obtained through a TEP and public comment as described in 4a2.1.1.

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

Not applicable; the measure has not yet been implemented. See section 4a2.3 below for how TEP and public comment feedback was considered during measure development.

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

Not applicable; the measure has not yet been implemented.

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.

This measure was developed with input from national TEP consisting of patients, surgeons, methodologists, researchers, and providers. We also held a three-week public comment period soliciting stakeholder input on the measure methodology, and publicly posted a summary of the comments received as well as our responses (available in the Downloads section at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/PC-Updates-on-Previous-Comment-Periods.html.

CMS and The Center for Outcomes Research and Evaluation (CORE) investigated issues identified during measure development public comment. Specifically, CORE and CMS:

- Renamed the measure "Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers" to reflect the procedures included in the measure cohort.

- In response to feedback received during the measure development public comment period, reviewed all of the individual CPT® codes within CCS categories and removed 15 individual procedures (CPT® codes) from the measure that are outside the scope of general surgery practice, including two specifically suggested for removal by a commenter.

- Reviewed statistically selected variables for face validity for the final risk model, and retained one variable (opioid use) because experts advised it was an important risk predictor and expressed a strong preference for including it in the model even though it was not statistically selected.

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.

Not applicable. Measure is not yet in use.

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.

Not applicable. Measure is not yet in use.

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

Not applicable. Measure is not yet in use.

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)

2539 : Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy

2687 : Hospital Visits after Hospital Outpatient Surgery

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

Not yet submitted to NQF: Hospital Visits after ASC Orthopedic Procedures (CMS)

Submitting to NQF in this November 2017 round: Hospital Visits after ASC Urology Procedures (CMS)

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.

Not applicable. The measures' outcomes are harmonized.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable. There are no competing measures.

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: Gen_Surg_ASC_NQF_Appendix_v2.0_-1-.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services (CMS)

Co.2 Point of Contact: Vinitha, Meyyur, Vinitha.Meyyur@cms.hhs.gov, 410-786-8819-

Co.3 Measure Developer if different from Measure Steward: YNNH/Yale Center for Outcomes Research and Evaluation

Co.4 Point of Contact: Danielle, Purvis, Danielle.purvis@yale.edu, 203-200-5342-

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.

CORE convened a TEP comprised of clinicians, patients, and experts in quality improvement to provide input on key methodological decisions.

TEP Members

- Robin Blomberg, BA, MA – National Forum of End-Stage Renal Disease, Network 16 (Representative for Kidney Patient Advisory Council); Seattle, WA

- Kirk Campbell, MD – New York University Hospital for Joint Diseases (Clinical Assistant Professor of Orthopedic Surgery); New York, NY

- Gary Culbertson, MD, FACS – Iris Surgery Center (Surgeon; Medical Director); Sumter, SC

- Martha Deed, PhD – Consumers Union Safe Patient Project (Patient Safety Advocate); Austin, TX

- James Dupree, MD, MPH – University of Michigan (Urologist; Health Services Researcher); Ann Arbor, MI

- Nester Esnaola, MD, MPH, MBA – Fox Chase Cancer Center (Professor of Surgery; Associate Director for Cancer Health Disparities and Community Engagement); Philadelphia, PA

- John Gore, MD, MS – University of Washington (Associate Professor of Urology); Seattle, WA

- Lisa Ishii, MD, MHS – Johns Hopkins School of Medicine (Associate Professor); American Academy of Otolaryngology-Head and Neck Surgery (Coordinator for Research and Quality); Baltimore, MD; Alexandria, VA

- Atul Kamath, MD – Perelman School of Medicine, University of Pennsylvania (Assistant Professor and Clinical Educator Director of Orthopedic Surgery); Hospital of the University of Pennsylvania (Attending Surgeon); Philadelphia, PA

- Tricia Meyer, PharmD, MS, FASHP – Scott & White Medical Center (Regional Director of Pharmacy); Texas A&M University College of Medicine (Associate Professor of Anesthesiology); Temple, TX

- Linda Radach, BA – Consumers Union Safe Patient Project (Patient Safety Advocate); Austin, TX

- Amita Rastogi, MD, MHA, CHE, MS – Health Care Incentives Improvement Institute (Chief Medical Officer); Newtown, CT

- Donna Slosburg, RN, BSN, LHRM, CASC – ASC Quality Collaboration (Executive Director); St. Pete Beach, FL

- Julie Thacker, MD, FACS – Duke Health and Hospital System (Medical Director of Evidence-Based Perioperative Care); Duke School of Medicine Clinical Research Unit (Medical Director, Department of Surgery); Durham, NC

- Thomas Tsai, MD, MPH – Brigham and Women's Hospital (General Surgeon); Harvard School of Public Health (Research Associate); Boston, MA

The CORE measure development team met regularly and was comprised of experts in internal medicine, quality outcomes measurement, and measure development. CORE convened surgical and statistical consultants with expertise relevant to outpatient surgery and quality measurement to provide input on key methodological decisions.

CORE Measure Development Team

- Faseeha Altaf, MPH – Project Lead, CORE
- Haikun Bao, PhD – Analytic Lead, CORE
- Mayur Desai, PhD, MPH – Project Consultant, CORE
- Elizabeth Drye, MD, SM – Project Director, CORE
- Harlan Krumholz, MD, SM – Director, CORE
- Zhenqiu Lin, PhD – Analytics Director, CORE
- Megan LoDolce, MA – Project Manager, CORE
- Erica Norton, BS – Research Associate, CORE
- Danielle Purvis, MPH – Project Coordinator, CORE
- Craig Parzynski, MS – Analytic Consultant, CORE
- Jennifer Schwartz, PhD, MPH – Project Lead (Formerly at CORE)
- Rushi Shah, BS – Research Assistant, CORE
- Mona Sharifi, MD, MPH – Clinical Consultant, Instructor of Pediatrics, Yale University School of Medicine

Consultants

- Robert Becher, MD, MS – Surgical Consultant, Assistant Professor of Surgery at Yale University School of Medicine
- Sean O'Neill, ND, PhD – University of California, Los Angeles (Resident, General surgery); Los Angeles, CA
- Sharon-Lise Normand, PhD, MSc—Statistical Consultant, Professor of Biostatistics, Department of Health Care Policy, Harvard Medical School

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? Not applicable; not yet endorsed

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

Ad.6 Copyright statement: Not applicable.

Ad.7 Disclaimers: Not applicable.

Ad.8 Additional Information/Comments: None.