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 #: 3366

Corresponding Measures:

De.2. Measure Title: Hospital Visits after Urology Ambulatory Surgical Center Procedures

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 urology 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 urology procedures and improve follow-up care by capturing and illuminating, for providers and patients, post-procedure hospital visits that are not often 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 urology procedure performed at an ASC.

S.6. Denominator Statement: The target population for this measure is Medicare FFS patients age 65 years and older, who have undergone a urology procedure in ASCs.

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 urology procedure. 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

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.

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 of hospital visits after urology ambulatory surgical center procedures captures adverse patient outcomes associated with ASC care and an important area for quality improvement.
- The developer also provides a logic model demonstrating interventions that can be undertaken by ASC, including patient education, medication reconciliation, technical quality of surgery, and other ASC interventions to prevent unplanned hospital visits.

Question for the Committee:

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

Guidance from the Evidence Algorithm

Box 1: The measure assesses a healthcare outcome →Box 2: The developer has provided empirical data that there is a relationship between the measured outcome and at least one healthcare outcome → Pass

The highest possible rating is pass.

Preliminary rating for evidence: ☒ Pass ☐ No Pass

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer notes a performance range of 3.7% to 10.1%, with median performance of 5.8%.

Disparities

The developer examined potential disparities affecting patient who are dually eligible for Medicare and Medicaid and patients with low socioeconomic status as determined by the AHRQ SES Index. The developer found that observed hospital visit rates were higher for dual-eligible vs non-dual-eligible patients (7.5% vs. 5.9%), and higher for low SES patients (scores below 42.7 on the AHRQ SES index), vs high SES patients (scores above 42.7) (6.2% vs. 5.9%).

Questions for the Committee:

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

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

- Measure has evidence
- There are several studies quoted by the authors that indicate factors controllable by the surgical center that can reduce the opportunity for hospital visits after surgery (assessment of voiding function, patient counseling, alpha blocker use, surgery technical quality and prevention of infections.
- Yes, provided
- Developer connected specific processes to outcome
- Evidence supports significance of variations and co-morbidity, but measure designs does not seem optimal to improve outcomes or determine accountability for acute encounters in first 7 days post-procedure

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

- Range of 3.7 to 10.1% with median performance of 5.8%
- There appears to be a gap
- CMS conducted a two-year study of the risk-standardized hospital visit rate for urology surgery. Results ranged from 3.4% to 10.2%; 24%-ile-5.32; 50%-ile – 5.8%; 95%-ile-7.6%. There was moderate levels of disparity based on dual-eligible status (dual-eligible patients were at a higher risk for hospital visits. However, the median rates varied minimally for centers with the lowest proportion of patients with social risk factors, indicating that SES adjustment will have minimal impact.
- Yes, small differences in performance
- There was a range of performance. Performance declined with duals and low SES
- Yes

1c. Composite Performance Measure - Quality Construct (if applicable):

- NA
- Yes
- See above and below

Criteria 2: Scientific Acceptability of Measure Properties

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

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators:

- Larry Glance
- Karen Joynt Maddox
- Marybeth Farquhar
- Eugene Nuccio
- Christie Teigland
- Steve Horner

Review A and Review B

Evaluation of Reliability and Validity (and composite construction, if applicable):

- Summary of Methods Panel Review Process
    - In their preliminary analyses, subgroup members did not reach consensus on the validity of the measure One potential rationale was confusion over NQF's policy for accepting face validity testing to fulfill validity testing requirements. During the call, NQF staff clarified that because this is a new measure, a face validity assessment could fulfill the testing requirements for validity, assuming the results are adequate. Empirical validity testing is only required for maintenance measures.
    - Scientific Methods Panel members also raised concerns about the lack of adjustment for social risk factor. While the subgroup noted the lack of compelling analysis to support the exclusion of dual status in the risk adjustment model, NQF staff clarified that this concern, alone, cannot justify a measure's failure by the Scientifc Methods Panel. However, the Standing Committee is tasked with reviewing the risk adjustment model and determining if there are conceptual and empirical rationales to include social risk factors.
    - Ultimately, subgroup members found the measure to be reliable and valid. The All-Cause Admissions and Readmissions Standing Committee will evaluate this measure in the Fall 2018 cycle.

Standing Committee Action Item(s):

- The Standing Committee can discuss reliability and validity, or agree to accept the ratings of the Scientific Methods Panel. It is important to note that the appropriateness of inclusion or exclusion of social risk factors was not within scope for the Scientific Methods Panel ratings.

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?
    - Please note that the Scientific Methods Panel was not tasked with considering the appropriateness of including social risk factors in the risk adjustment model.

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

Scientific Acceptability

Measure Number: 3366

Measure Title: Hospital Visits after Urology Ambulatory Surgical Center Procedures

Type of measure:

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

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

Data Source:

☒ Claims ☐ Electronic Health Data ☐ Electronic Health Records ☐ Management Data

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

Data:

☐ Enrollment Data ☐ Other denominator files

Level of Analysis:

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

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

☐ Integrated Delivery System ☐ Other

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

   Methods Panel members did not note any concerns about the measure specifications.

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
   Reliability testing was conducted at the measure score level. Please note, NQF does not require data element reliability testing if data element validity has been demonstrated.
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Score-level reliability was demonstrated in two ways: a signal-to-noise ratio (SNR) analysis using the Adams method and a split-sample interclass correlation coefficient (ICC) (2,1).
   - A summary of the Scientific Methods Panel members review of the methods of reliability testing is presented below. This input is intended to serve as a resource to inform Standing Committee Deliberations. Please note, staff summarized this information as received from SMP members.
     - **PANEL MEMBER 1:** Methods were appropriate.
       - Data elements – compared percent agreement across development and validation samples
       - Score level: used (1) split-sample approach & (2) signal-to-noise ratio at facility level
     - **PANEL MEMBER 2**: ICC from split-sample testing
     - **PANEL MEMBER 3**: Data Element Reliability: No percent agreement or kappa score reported. Used audited fields from claims data and avoided us of fields that were thought to be coded inconsistently. Identified each variable through empiric analyses. Method used is appropriate.
       - Measure Score Reliability: Used intra-class correlation coefficient (ICC) using a split-sample (test-retest) method. Estimated facility-level reliability using Adams formula. Method used are appropriate.
     - **PANEL MEMBER 4**: The data sets seem old (e.g., Oct 2013 – Sept 2014). How often are the data updated for reporting to facilities?
       - Two types of reliability testing for measure score (Interclass Correlation & Split Half). Authors state that they did data element reliability testing, but no information on how the testing was done is included in submission.
     - **PANEL MEMBER 5:** Methods are appropriate.

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

   Submission document: Testing attachment, section 2a2.3

- The results of the split-sample ICC (2,1) = 0.45
- The results of the signal to noise ratio S(for facilities with >=30 cases) median reliability = 0.69
    - Please note that at least for the split-sample analysis, the developers did NOT limit their testing data to facilities with >=30 cases (i.e., testing aligned with specifications)
- Developer notes these results indicate that there is moderate reliability in the measure score using the Landis and Koch scale for interpretation. Some concern was raised by the Scientific Methods Panel members on the ICC for the overall measure score reliability.
- A summary of the Scientific Methods Panel members review of reliability testing results is presented below. This input is intended to serve as a resource to inform Standing Committee Deliberations. Please note, staff summarized this information as received from SMP members.
    - **PANEL MEMBER 1:**
        - Data elements – frequencies were in close agreement across development and validation data
        - Score
            - Split -sample measure score reliability was 0.45 which is acceptable
            - median facility reliability (measure of signal-to-noise ratio) was 0.69 which is also acceptable
    - **PANEL MEMBER 2:** ICC 0.45, median facility-level reliability 0.69.
    - **PANEL MEMBER 3:**
        - Data Element Reliability: Used audited data field from CMS Medicare Database. Sample sizes are sufficient for both development sample and validation sample.
        - Measure Score Reliability: split -sample measure score reliability was 0.45 indicating fair agreement. Facilities with at least 30 procedures yielded a median reliability score of 0.69 indicating good agreement. (I used Cicchetti (1994) for assessing guidelines for interpretation of ICC).
    - **PANEL MEMBER 4:** The ICC = 0.45; split half = 0.69. Adequate values.
        - Their results for their data element reliability were inadequate. Simply showing that two samples had similar mean values across variables does not constitute data element reliability.
    - **PANEL MEMBER 5:** Split-sample overall measure score reliability yielded an ICC score of 0.45 which the developers state "indicates moderate reliability. " Facility-level reliability testing indicated median reliability was 0.69 which the developers state "indicates substantial reliability. " I believe the ICC score of 0.45 for the overall measure score reliability indicates low reliability, not moderate. The facility ICC of 0.69 is better but still may be below the standard that has been debated by the Scientific Methods Panel (but not set in stone certainly). The developers argue that the split sample measure is conservative and expected to produce lower reliability scores. The argument that this measure is more similar to assessing personality disorder (one can imagine a lot of noise) than weight (which one could expect to be consistent). I disagree, whether a person had a hospital visit 7 days following colonoscopy, especially with the level of risk adjustment applied, would seem to be much more straightforward and consistent.

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

    Submission document: Testing attachment, section 2a2.2

    ☒**Yes**

    ☐**No**

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

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

Submission document: Testing attachment, section 2a2.2

☐**Yes**

☐**No**

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

Please note, NQF does not require data element reliability testing if data element validity has been demonstrated.

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

    Ultimately, the Scientific Methods Panel gave this measure an overall rating of moderate reliability. Individual member scores ranged from low to moderate. A summary of panel members rationales for their rating is provided under item 11 below.

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - **PANEL MEMBER 1:** Median facility reliability was greater than 0.60, which suggests substantial score reliability.
    - **PANEL MEMBER 2**: Testing indicates moderate reliability.
    - **PANEL MEMBER 3:** The results of the ICC differ depending on which scale is used. The developer used the standards established by Landis and Koch (1977) which provide a better rating, then other established standards.
    - **PANEL MEMBER 4:** The measure score reliability was OK; the data element reliability was non-existent.
    - **PANEL MEMBER 5:** Relatively low ICC scores, especially at the overall measure reliability level but also at facility level.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - Methods Panel members did not note any concerns with the measure exclusions.

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

    Submission document: Testing attachment, section 2b4.

    - Two Methods Panel members noted concern regard the measure's ability to identify meaningful differences in performance.
        - **PANEL MEMBER 2:** The range in performance is quite small – 0.5% between the 25th and 75th percentile. 19 outliers were identified among 1,204 ambulatory surgery centers. Median odds ratio is 1.27.
        - **PANEL MEMBER 5:** No empirical testing of validity at this time, only face validity. At least 2 of 14 TEP members disagreed with the validity statement. Facility measure scores ranged from 3.7% to 10.1%, with a median RSHVR of 5.8% (the 25th and 75th percentiles were 5.6% and 6.1%, respectively). While the range is large, the difference in scores between the 25th and 75th percentiles is not large and may not discriminate between facilities.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.

    - Methods Panel members did not note any concerns regarding comparability of results.

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

    Submission document: Testing attachment, section 2b6.

- Methods Panel Members did not note any concerns regarding missing data.

16. **Risk Adjustment**

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

- This measure uses a statistical risk-adjustment model with nine risk factors. Specifically, the measure uses a two-level hierarchical logistic regression model to estimate ASC-level risk-standardized hospital visit rates (RSHVRs).
- The nine risk factors included are:
  - 1. Age (years > 65)
  - 2. Work Relative Value Units (work RVUs)
  - 3. Benign prostatic hyperplasia with obstruction (ICD-9-CM diagnosis codes 60001, 60021, 60091; ICD-10-CM diagnosis codes N401, N403)
  - 4. Complications of specified implanted device or graft (Condition Category 176)
  - 5. Number of qualifying procedures: 1, 2, 3 or more
  - 6. Poisonings and allergic and inflammatory reactions (CC 175)
  - 7. Major symptoms, abnormalities (CC 178)
  - 8. Parkinson's and Huntington's diseases; seizure disorders and convulsions (CC 78, 79)
  - 9. Ischemic heart disease (CC 86, 87, 88, 89)

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

- The developer provided a conceptual rationable for including social risk factors and conducted empirical analyses to demonstrate their impact. Ultimately, the developer chose not to include these factors in the final risk adjustment model.
- Scientific Methods Panel members raised concerns about the decision not to include dual status in risk-adjustment approach (i.e., dual-eligible status had a statistically significant association (OR: 1.30, 95% CI: 1.13 -1.48, p = 0.0001) with the risk of a hospital visit)
- A Methods Panel member provider the following comment on the conceptual relationship:
  - **PANEL MEMBER 1:** Inclusion of SES variables had no significant effect on provider ranking.

16d.**Risk adjustment summary:**

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

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

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

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

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

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

- This measure uses a two-level hierarchical logistic regression model. The c-statistic was .61 in the development and validation samples. Some Methods Panel members raised concerns that this was a bit low. Please note, NQF does not maintain a set threshold for an adequate c-statistic.

- A summary of the Methods Panel members' assessments of the risk adjustment approach is provided below. These summaries are intended to inform Standing Committee discucssion.
    - **PANEL MEMBER 1:**
        - Hierarchical GLM
        - Used purposeful stepwise selection in bootstrap data sets to select risk factors
        - Quality is quantified using PE ratio
        - Model includes age, six comorbidity variables based con CMS Condition Categories, and work RVU as a measure of surgical complexity
        - Model performance
            - C statistic of 0.61 (in validation data) is on the low end, even for readmission models which typically have very low C statistics (note, a C stat of 0.5 means model is no better than the flip of a coin)
            - Calibration – as assessed using calibration curves (in validation data) is very good
    - **PANEL MEMBER 2**: C-statistics aren't great (in the .6 range), but acceptable.
        - Disagree with the developers' takeaway from the results of social risk factor testing. Dual eligibility had a significant relationship with the outcome, but the developers still do not include it because they don't think social risk factors should be included in risk adjustment. Their argument is loosely based on the fact that facility performance doesn't change much when that variable is added, but that is an inappropriate argument (the same could be made for any other risk factor in isolation – that's not the standard to which risk elements are held). Also I don't understand the rationale for only showing the top quartile of the SES factors in the graphs in Figure 1 – what about the bottom three quartiles?
    - **PANEL MEMBER 3:** Thorough and appropriate.
    - **PANEL MEMBER 4:** Discussion of socio-demographic variables was extensive. Work Relative Value Units was the only socio-demographic variable in model (age is a given). Operational definition of this RF is not clear. Correlation scattergrams shown.
        - The results are poor (c-statistics 0.610 to 0.615).
    - **PANEL MEMBER 5:** The measure uses a two-level hierarchical logistic regression model to estimate ASC-level risk-standardized hospital visit rates (RSHVRs). This approach accounts for the clustering of patients within ASCs between facility variation that may be due to quality and variation in sample size across ASCs. This approach controls for, as stated by the developers "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."
        - From 1,000 bootstrapped models, the developers selected 9 variables as the final risk-adjustment variables.

VALIDITY: TESTING

17. **Validity testing level: ☒ Measure score ☐ Data element ☒ Both**
18. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

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

    ☐ **N/A (score-level testing not conducted)**
19. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

- The developer provided results of an assessment of the measure's face validity.

- The developer assessed face validity in various ways; however, only the TEP assessment meets NQF's requirements for face validity.
- Please note, NQF accepts face validity testing for new measures. Empirical testing is only required at the time of maintenance review.
- A summary of Methods Panel members assessment of validity testing is provided below. This is intended to inform the Standing Committee's deliberations.
  - **PANEL MEMBER 1:** Face validity assessed using TEP
    - Empiric validity is assessed by assessing predictive validity of risk adjustment model
  - **PANEL MEMBER 2:** Face validity from TEP
  - **PANEL MEMBER 3:** TEP used for validity. Adequate method for assessing validity of measure score.
  - **PANEL MEMBER 4:** Returning measures need empirical validity testing. The face validity method is impressive, but not adequate for returning measures.
  - **PANEL MEMBER 5:** Sufficient

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

    Submission document: Testing attachment, section 2b2.3

    - The developer provided the following results from face validity testing. Of the 14 TEP respondents, 12 (86%) indicated that they somewhat, moderately, or strongly agreed and 2 respondents moderately disagreed with the statement about whether the results from the measure can be used to differentiate poor vs good quality. They also provided reason for disagreement from one of the two who disagreed.
    - A summary of Methods Panel members assessment of the results of validity testing is provided below. This is intended to inform the Standing Committee's deliberations.
      - **PANEL MEMBER 1:** Results from TEP indicate strong support of face validity
        - Empiric testing of risk adjustment model indicate acceptable model performance
      - **PANEL MEMBER 2:** Reasonable face validity per TEP
      - **PANEL MEMBER 3:** Results indicate a high level of agreement (86%) among TEP members as to the overall face validity of the measure.
      - **PANEL MEMBER 4:** Face validity results were impressive, but not adequate for returning measure. Empirical testing is required.
      - **PANEL MEMBER 5:** Sufficient

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

    Submission document: Testing attachment, section 2b1.

    Methods Panel members responses varied to this question.

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

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

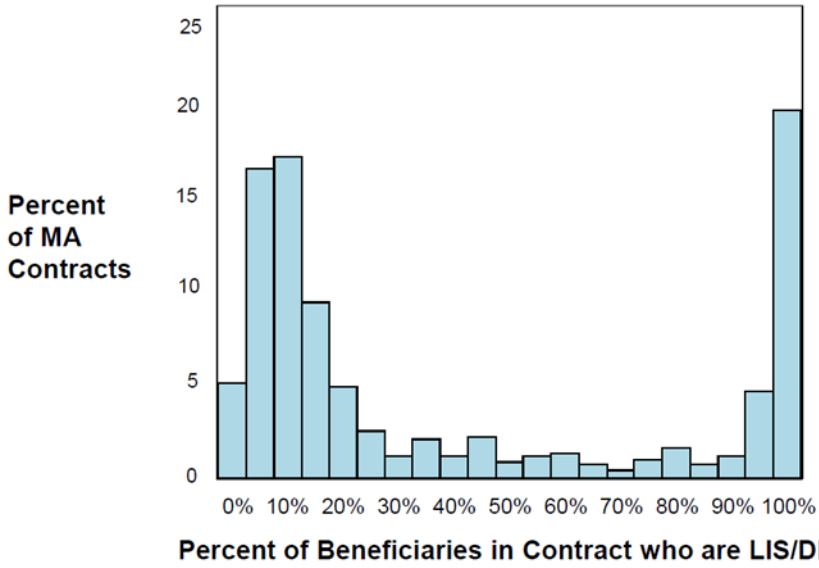
    Methods Panel members responses varied to this question. Please note that data element validity testing was not provided.

23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**
    - Ultimately, the Scientific Methods Panel gave this measure an overall rating of moderate for validity. Individual member ratings ranged from insufficient to high.
    - A summary of individual Methods Panel reviewers rataionale for their rating of validity is provided below. This information is intended to inform the Standing Committee's discussion.

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

- **PANEL MEMBER 1:** Although risk adj model performance is within acceptable range for readmission models, it would be difficult to assign more than a moderate rating given C stat of 0.61.
- **PANEL MEMBER 2:** I have concerns with the social risk factor adjustment and low c-statistic.
- **PANEL MEMBER 3:** Documented processes for determining face validity of the measure. Also used various methods and the literature to substantiate their claim of validity.
- **PANEL MEMBER 4:** Empirical testing is required.
- **PANEL MEMBER 5:** While three indicators of social risk were evaluated no SES factors tested were selected for inclusion: 1) Medicaid dual-eligibility (not a reliable proxy for low income) 2) race (African American vs. White only, a limitation), and 3) the AHRQ SES index (based on the American Community Survey which is a very imprecise and blunt measure of an individuals' SES). It is important to understand how the ACS block group level data are generated and what the data represent. The ACS is an annual survey. In 2015, the ACS sampled approximately 3.5 million housing units, which represents about 2.5% of households nationwide. The final sampling ratio is much lower; final interviews (which include occupied and vacant housing units) comprise about 1.6% of housing units nationwide. These data are aggregated to the block group level representing about 250,000 geographic areas. The ACS data thus generally represent a very small sample of SES information averaged across multiple disparate neighborhoods, resulting in a relatively imprecise assignment of characteristics to individuals.
  Observed hospital visit rates were higher for dual-eligible and low SES patients: 7.5% for dual-eligible patients compared to 5.9% for non-dual-eligible patients, and 6.2% for low SES patients compared to 5.9% for higher SES patients. Inclusion of each of these risk factors in the models (controlling for other risk-adjusters) showed a significant disparities association for dual-eligible patients only (dual-eligible: OR: 1.30, 95% CI: 1.13 -1.48, p = 0.0001) (not surprising given the impreciseness of ACS data).

  Distributions of the measure score for ASCs with a low % of patients with dual status (1st quartile) and high % of patients with dual status (4th quartile) were shown and indeed showed little variation in measure rates. However, 1.96% was considered "low percent duals" (1st q) and 7.54% was considered "high percent duals" (4th q). Based on data published by CMS, these distributions/cut-offs DO NOT represent the distribution of duals in the Medicare population. I would expect to something more similar to Medicare Advantage plan contracts which tend to either have a very high percentage (greater than 95%) of dual beneficiaries or a very low percentage (less than 20%) of dual beneficiaries (see below). Using quartiles is probably not appropriate to see differences between high dual and low dual facilities.



The developers conclude that dual-eligible status had a statistically significant association with the risk of a hospital visit, but conclude that this association may be a result of either disparate care received due to their socio-demographic status or increased risk of hospital visits not accounted for by risk adjustment. Since they have adjusted for other risk factors and are evaluating only within facility differences, I disagree with the conclusion. CMS has stated within facility differences are deemed "potentially appropriate for risk adjustment" while between facility differences "represent true differences in quality." Facilities providing worse care should be identified by lower measure rates if this measure is appropriately specified. Facilities serving a larger proportion than 7-8% of patients may be unfairly penalized by not adjusting for this social risk factor that was shown in this evaluation to be significant and represent a disparity.

ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - **PANEL MEMBER 5:** Acceptability of social risk factors used in testing (appropriateness of level of data used) and rationale for not including even when evidence shows a disparity.

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

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

- Moderate
- OK
- Ok
- None
- See above and below - claims data seems to ignore provider performing intervention and ASC where performed and use only acute care facility as basis of measure; seems imprecise method to understand co-morbidity and close performance gaps within communities of care

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

- No
- ICC for the overall measure seems low
- The measure follows standard risk-adjusted readmission ratios and rates relative to a national average. Although the methods have been shown to be reliable, the ICC of 0.44 indicates poor reliability.
- Split sample ICC--low to moderate, c stat moderate
- No
- As above - limited value in modeling reliability at facility level if different than ASC where interventions performed
- 2b1. Validity -Testing: Do you have any concerns with the testing results?
- Moderate
- Moderate
- Face validity ok per NQF
- no
- As above - appears to have validated measure of wrong perspective on GU interventions and outcomes
- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data): 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate

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

- Low c statistic for adjustment of social risk factors
- Social factor adjustment
- [pg 10 Testing Attachment] – "it is likely acceptable to see lower values of reliability given quality's latent and unpredictable nature." Quality measurement is the sole purpose for developing this readmission metric; if quality is too diverse of a topic to accept an ICC less than 0.5, it begs the question of whether the readmission metric should be a measure of quality. The c-statistic is reported as 0.61, representing a relatively weak model; however, the confidence intervals were not reported.
- Small differences between best and worse, only 19 outliers identified--will this actually measure any meaningful difference?
- I don't think so
- As above and below - unless I am misunderstanding bundle and measure design
- 2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure?2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?
- Has outcome
- It seems that social risk factor adjustment should be included.
- They note that duals had more hospital utilization but finish that they are constrained from including SES measures in risk adjustment.
- Yes
- Somewhat troubled by exclusions for those with less than 9-12 months of trailing Medicaid coverage; co-morbidity likely higher among newly covered if interventions delayed due to SES factors tied to Medicaid eligibility, even for those over 65

Criterion 3. Feasibility

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

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

- This measure uses claims data that has been shown to be operationalizble, however, the measure is not yet in use.
- There are no fees, licensing, or requirements to use the measure.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- Appears feasible
- Measure is feasible
- Data is readily available and appears to be feasible.
- Claims/enrollment based, should be feasible
- None
- Feasibility would not seem hampered by including all related Part/ASC claims for index intervention

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- The measure is new and not yet in use, but will be used in the Ambulatory Surgical Center Quality Reporting Program (ASCQR) for public reporting beginning with CY 2022.

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

- All ASCs with at least one eligible case were included in a dry run in August, 2018 (n=1,149). ASCs received their results for the October 1, 2015 through September 30, 2017 performance period. CMS provided education and resources for the ASCs before and during the dry run, responded to all questions, and presented information in writing and on a webinar to stakeholders.
- According to the developers, the measure has not yet been refined based on this feedback, but it is being reviewed and considered, and the measure will be refined before implementation.

Additional Feedback:

N/A

***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**

RATIONALE:

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

N/A – 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 measure has not been used yet but the developer states: "In designing the measure, we sought to minimize the potential of this measure to result in the denial of future care to high-risk individuals. We developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of hospital visits will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time."

Potential harms

N/A

Additional Feedback:

N/A

***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**

RATIONALE:

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

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

given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Has not been publically reported
- Ok
- NC
- Not being used, did get some feedback
- Might need to be careful in applying penalties if not risk adjusted for SES
- Don't view facility level as the most appropriate level of attribution. Claims data sources don't seem to include provider or facility detail for index ASC where procedures are being done. Unless TIN's for ASC's and acute care facilities are related, attribution for providers and groups seems obscure, hampering value of measure to improve outcomes or evaluate quality of ASC interventions and providers

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

- Appears to be but has not been vetted in real world setting
- Ok
- Metric not publicly reported yet.
- Looks like they are going to do a dry run to test.
- Useful information but need to understand SES factors
- Feel usability to improve outcomes hampered by issues above if only facility-level claims data used for downstream events

Criterion 5: Related and Competing Measures

Related or competing measures

- NQF did not identify competing measures.

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

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

- n/a
- Related to multiple ASC measures
- None known
- Would like to know if AUD intends to use measure for PQRS incentives or those relating to Medicare bundling pilots or programs

Public and Member Comments

NQF received no public or member comments on this measure as of January 25, 2019.

Developer Submission

Brief Measure Information

NQF #: 3366

Corresponding Measures:

De.2. Measure Title: Hospital Visits after Urology Ambulatory Surgical Center Procedures

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 urology 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 urology procedures and improve follow-up care by capturing and illuminating, for providers and patients, post-procedure hospital visits that are not often 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 urology procedure performed at an ASC.

S.6. Denominator Statement: The target population for this measure is Medicare FFS patients age 65 years and older, who have undergone a urology procedure in ASCs.

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 urology procedure. 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

Uro__ASC__NQF_Evidence_Attachment_FINAL.docx

1a.1 <u>For Maintenance of Endorsement:</u> 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*)**:** 3366

Measure Title: Hospital Visits After Urology Ambulatory Surgical Center Procedures

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

Date of Submission: 11/7/2018

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *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:

- Health outcome: 3 a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

☒ Health outcome: Hospital Visits (emergency department [ED] visits, observation stays, and unplanned inpatient admissions)

☐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.12 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) procedures often reflect procedure-related adverse events and quality issues. Common reasons for unplanned hospital visits following ASC surgeries include urinary retention, pain, bleeding, and infection. Strategies and interventions that have been shown to reduce unplanned hospital visits after outpatient surgery procedures include:

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

2) Prevention of urinary retention through:

-Appropriate assessment of voiding function prior to discharge [2];

-Patient counseling prior to procedures regarding risks and management options for postoperative urinary retention [2]; and

-Preoperative α-blocker use [3].

3) Appropriate patient education on preparation prior to procedures. [4]

4) Improving the technical quality of the surgery, including the choice of procedural technique and anesthesia. [5]

5) Prevention of surgical site infections through evidence-based guideline-concordant care. [6,7]

6) Prevention of adverse drug events through medication reconciliation. [8]

The measure will identify risk-adjusted variation in performance across ASCs 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. *Archives of Surgery.* 2007; 142(3):263-268.

2. Geller EJ. Prevention and management of postoperative urinary retention after urogynecologic

surgery. Int J Womens Health. 2014;6:829-838.3.

surgery.

3. Mohammadi-Fallah M, Hamedanchi S, Tayyebi-Azar A. Preventive effect of tamsulosin on

postoperative urinary retention. Korean J Urol. 2012;53(6):419-423. 4. Romero A, Joshi GP. Adult Patient

for Ambulatory Surgery: Are There Any Limits? *ASA Newsletter.* 2014;78(9):18-20.

5. 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.

6. Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR. Guideline for prevention of surgical site infection, 1999. *American Journal of Infection Control.* 1999;27(2):97-134.

7. Agency for Healthcare Research and Quality (AHRQ). 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. Available at: http://teacherweb.com/NY/StBarnabas/Law-PublicPolicy/JCINT-2005.pdf. Accessed June 7, 2017.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

A hospital visit after outpatient surgery is unexpected, and many of the reasons for hospital visits are preventable. Patients often present to the hospital for complications of surgical care, including infection, post-operative bleeding, urinary retention, nausea and vomiting, and pain. The outcome of unplanned hospital visits following outpatient surgery is a widely accepted measure of outpatient surgical care quality and reflects important features of healthcare structure, process, and service, including patient selection and management, technical aspects of the surgery, and delivery of guideline-concordant care.

Factors associated with patient selection, preparation, and post-discharge planning are important predictors of adverse events and unplanned hospital visits following outpatient surgery [1-3]. Demographic characteristics, such as older age, and a broad range of clinical comorbidities have been associated with post-procedure hospital visits [1-3]. Studies also point to the importance of post-discharge factors, such as ability to manage pain and availability of a responsible caregiver, in reducing poor outcomes [2].

The risk of unplanned hospital visits is also influenced by various technical aspects of the surgery, including anesthetic technique [2-3] and length of surgery [3]. In addition, outcome rates may be influenced by clinical pathways [2] and delivery of guideline-concordant care [4-6]. In particular, there are growing efforts to systematically address issues of surgical site infection [4, 5] and medication reconciliation [6]. For example, the Agency for Healthcare Research and Quality (AHRQ) developed a quality improvement collaborative for the ambulatory surgery environment to reduce healthcare-associated infections and surgical harms in ASCs through the use of a surgical safety checklist and improved safety culture through teamwork and communication. Partners involved in the collaborative concluded that efforts to increase the availability of

meaningful data would be beneficial to more accurately assess outcomes in the ASC setting, and would facilitate an ASC's ability to follow patients after discharge. [7]

Moreover, many ASC providers are often unaware of their patients' hospital visits after surgery because patients often present to an emergency department (ED), leading to understated adverse event rates and suggesting the need for better measurement to drive quality improvement. Therefore, both patients and providers will benefit from an outcome measure of hospital visits – a broad, patient-centered outcome that reflects the full range of reasons leading to hospital use among patients undergoing outpatient surgery.

In summary, interventions to improve the quality of care for patients undergoing outpatient surgical procedures – including appropriate patient selection, improving surgical techniques, implementing protocols to address common problems such as adequate control of postoperative pain, patient education about potential adverse effects of the surgery, and reconciling patient medications – may reduce unplanned hospital visits following outpatient surgery.

Citations

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

2. Romero A, Joshi GP. Adult patient for ambulatory surgery: are there any limits? *American Society of Anesthesiologists, Inc 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. *Canadian Journal of Anesthesia/Journal Canadien D'Anesthésie.* 2013; 60(7):675-683.

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

5. Agency for Healthcare Research and Quality (AHRQ). 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.

7. Agency for Healthcare Research and Quality (AHRQ). Toolkit to Improve Safety in Ambulatory Surgery Center. 2017. Available at: https://www.ahrq.gov/professionals/quality-patient-safety/hais/tools/ambulatory-surgery/index.html. Accessed May 17, 2017.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

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

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

1b. Performance Gap

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

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

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

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

This measure aims to reduce adverse patient outcomes associated with ASC urology procedures and improve follow-up care by capturing and illuminating, for providers and patients, post-procedure hospital visits that are not often 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.*

Please note that this is a new measure, not a maintenance measure. More details regarding the performance scores calculated during measure development are presented in Section 2 of the Testing Form. This measure is in the process of being implemented in the Ambulatory Surgical Center Quality Reporting (ASCQR) program (for Calendar Year 2022 payment determination and public reporting), and therefore CMS has collected data for confidential reporting prior to public reporting.

Summary of National Results for Urology ASC Measure Dry Run

CMS conducted a confidential reporting period (dry run) between August 1 through August 30, 2018. Results from the dry run are summarized in this section. The risk-standardized hospital visit rates (RSHVRs), among all ASCs with at least one eligible urology procedure during the performance period of October 1, 2015 to September 30, 2017 (n=1,149) ranged from 3.42% to 10.16% (mean of 5.88%). This facility-level variation in scores suggests opportunities for quality improvement. The distribution has a skewness coefficient of 1.00 and a kurtosis coefficient of 4.16 meaning that the data skews slightly right with a moderate tail. (Note that this form does not support our including a graphic representation of this data.)

Below we show the measure score (RSHVRs) among those ASCs with at least 35 eligible urology procedures (n=570). The facilities had a median case size of 139 eligible surgeries and a mean case size of 247.29 eligible surgeries.

Percentile, RSHVR

Min, 3.42

1st, 3.90

5th, 4.73

10th, 4.97

25th, 5.32

50th, 5.83

75th, 6.48

90th, 7.15

95th, 7.61

99th, 8.27

Max, 10.16

Mean, 5.95

Std deviation, 0.89

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 data in 1b.2.

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.*

Please note that this is a new measure, not a maintenance measure.

We provide a detailed disparities analysis in Section 2b3.4b of the testing form, and those results are summarized below.

We analyzed the following social risk factors: race, AHRQ SES Index, and dual-eligible status using the Medicare FFS FY 2015 dataset.

Overall, we found that observed hospital visit rates were higher for dual-eligible vs non-dual-eligible patients (7.5% vs. 5.9%), and higher for low SES patients (scores below 42.7 on the AHRQ SES index), vs high SES patients (scores above 42.7) (6.2% vs. 5.9%). Additionally, when each of these factors were added to the patient-level risk model, we found that patients with dual-eligible status only, were at a higher risk for the outcome (OR: 1.30, 95% CI: 1.13 -1.48, p = 0.0001).

When examining the impact at the ASC facility level, we found only slightly higher measure scores for ASCs with higher proportions of patients with social risk factors (ASCs in the 4th quartile for social risk; Table 2 in Section 2b3.4b compared to the ASCs with a lower proportion of patients with social risk factors (ASCs in the 1st quartile for social risk), but the distributions largely overlapped. The median RSHVR varied minimally across quartiles for all three variables (5.8% - 6.0%). When we examined the measure scores in facilities with the highest proportion of patients with social risk factors (the 4th quartile) in more detail, we found that there was no relationship between the measure score and the proportion of patients with the social risk factor (Figures 1, 2, and 3 in Section 2b3.4b). Finally, measure scores for facilities when including the risk factor variable in the model did not meaningfully differ (correlation coefficients between RSHVRs with and without adjustment for all three risk factors were near 1, and mean differences in RSHVRs were near zero).

These results suggest that adjusting for social risk factors will have almost no effect on the measure scores for facilities, even for those with a high proportion of patients with social risk factors. The measure, as specified, does not include social risk factors as variables in the model.

Social risk factors: CMS's work on stratifying outcome measures for social risk factors

Patients with social risk factors are at a higher risk for the outcome described for this measure. However, CORE currently does not recommend adjusting this measure for social risk factors because we do not know if the increased risk for the outcome is due to disparities in the quality of care that is delivered to patients with the social risk factor, disparities in the care that is accessible to or chosen by patients with the social risk factor, or the social risk factor itself independent of the care delivered or received. (Note, however, that for facilities with the highest proportion of patients with social risk factors, there is no correlation between the percent of patients with social risk factors and the measure score.) Importantly, CMS, on an independent, parallel track, is developing stratification methodologies to reveal patient level disparities in quality measures.

CMS has developed two stratification methodologies to measure quality for patients with social risk factors and identify hospitals with healthcare disparities. These methods respond to directives from a report from the Assistant Secretary for Planning and Evaluation (ASPE) evaluating the relationship between social risk and Medicare payment programs. This report recommends, among other things, to 1) develop statistical

techniques to report performance measures for patients with social risk factors, and 2) introduce health equity measures to illuminate disparities in health care quality [cite ASPE report]. Both methods are intended to supplement overall hospital quality measures, which are currently and will continue to be publicly reported.

Method 1: The hospital-specific disparity method

The hospital-specific disparity method assesses differences in health outcomes between dual and non-dual eligible patients within a hospital. The goal of this method is to assess the difference in outcomes for two patients who walk into the same hospital with the same condition and medical history, but have a different dual eligibility status. In other words, we want to know the likelihood of a difference in outcome for two patients that are the same except for their dual eligibility status. The modelling strategy builds on current risk-adjusted outcome measures by including a "disparity factor" to assess the degree of disparity between subgroups of patients. This approach accounts for differences in patients' severity of illness at each hospital and allows us to attribute differences in outcomes to dual eligibility status.

Method 2: The group-specific outcome rate method

The group-specific outcome rate method, or dual eligible outcome rate method, assesses relative quality for dual eligible patients across hospitals by giving each hospital a risk-standardized outcome rate for their dual eligible patients. It answers the question, "how does the risk-standardized outcome rate for dual eligible patients at a specific hospital compare to other hospitals?" It also allows us to compare each hospital's outcome rate for dual eligible patients to the national average outcome rate for dual eligible patients. This method risk adjusts for patients' illness severity to capture differences among hospitals rather than differences among patients so that hospitals can be compared fairly.

CMS implemented confidential reporting of both disparity methods using the pneumonia readmission measure in September 2018, meaning that hospitals have received their results in hospital specific reports (HSRs) and had the opportunity to provide feedback on the methods we developed and ask questions about their results. The disparity methods are being considered for future public reporting in the Hospital Inpatient Quality Reporting (IQR) Program.

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 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/Downloads/Version-10_Hospital-Visits_Urology-ASC-Procedures_Measure-Technical-Report_052017.pdf

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 Urology_ASC_Measure_NQF_Data_Dictionary_v1.0-636685738163686742.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.

No

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

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 urology 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 urology procedure performed at an ASC identified using the Centers for Medicare & Medicaid Services (CMS) Medicare administrative claims data. The codes used to identify ED visits and observation stays are in the attached Data Dictionary, sheet "S.5 Numerator-ED Obs Def."

Time Period for Data

Numerator time window: within 7 days of ASC procedure.

Denominator time window: urology ASC procedures performed during the measurement period

Identification of Planned Admissions

The measure outcome includes hospital visits within 7 days following the urology procedure, unless that inpatient admission is deemed a "planned" admission. We used CMS's Planned Readmission Algorithm v4.0 to identify 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 patients' index urology procedure, considering a few, specific, limited types of care 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 urology procedure, as "unplanned" and thus counts these inpatient admissions in the measure outcome.

Details of the planned admission algorithm and International Classification of Diseases, 9th Revision (ICD-9)/ International Classification of Diseases, 10th Revision (ICD-10) codes to identify planned admissions are in the attached Data Dictionary, sheets: (1) "S.5 Planned Adm Alg Overview," (2) "S.5 Planned Adm Alg Flowchart," and (3) "S.5 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 "S.5 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)*

The target population for this measure is Medicare FFS patients age 65 years and older, who have undergone a urology procedure in ASCs.

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

The target population is Medicare FFS patients aged 65 years and older who are undergoing outpatient urology procedures performed at ASCs. 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 the urology procedure to ensure that we have adequate data for identifying comorbidities for risk adjustment.

To identify eligible ASC urology procedures, we first identified a list of procedures from Medicare's 2015 ASC list of covered procedures, which includes procedures for which ASCs can be reimbursed under the ASC payment system. This list of surgeries is publicly available at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalOutpatientPPS/Hospital-Outpatient-Regulations-and-Notices-Items/CMS-1589-FC.html (refer to Addendum AA on the website).

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 publicly available, 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. We, therefore, further limited the list of covered ASC procedures to "major" and "minor" procedures defined using Medicare's Global Surgical Package [1]. Specifically, we identified "major" and "minor" surgeries using the global surgery indicator (GSI) values of 090 and 010, respectively, which correspond to the number of post-operative days included in Medicare's global surgery payment for the procedure. However, we also included cystoscopy with intervention, which has the GSI value of 000, since this is a common procedure, often performed for therapeutic intervention by surgical teams, and has an outcome rate similar to other procedures in the urology measure cohort.

Finally, to initially define the urology cohort, we used the Clinical Classifications Software (CCS) developed by the Agency for Healthcare Research and Quality (AHRQ). The CCS is a tool for clustering procedures into clinically meaningful categories using CPT® codes by operation site. We included all procedures defined by the CCS as "operations on the urinary system" and "operations on the male genital organs" and retained all of those typically performed by urologists. Examples of urology procedures include removal of prostate gland, cystoscopy, and fragmenting of kidney stones. The coding list for the body systems is available at: http://www.hcup-us.ahrq.gov/toolssoftware/ccs/AppendixDMultiPR.txt.

The codes used to define the procedures in the urology cohort are in the attached Data Dictionary, sheet "S.7 Codes Used to Define Cohort."

Citations

1. Department of Health and Human Services, Centers for Medicare and Medicaid Services. Global surgery fact sheet 2017. https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/GloballSurgery-ICN907166.pdf. Accessed June 7.

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 urology procedure. 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.)*

N/A.

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:

Rate/proportion

If other:

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

Better quality = 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 rates (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 an ASC's patients, multiplied by the national observed rate of unplanned hospital visits. 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 urology 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 (see Appendix C). 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. An ASC's P/E ratio is then multiplied by the overall national rate of unplanned hospital visits to calculate the ASC-level RSHVR. 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 C of the measure´s technical report for details.The measure's technical report can be found at https://www.qualitynet.org/dcs/ContentServer?cid=1228776662386&pagename=QnetPublic%2FPage%2FQnetTier3&%20c=Page

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. Available at:

http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/2015_Measure_Evaluation_Criteria.aspx. Accessed June 7, 2017.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

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.

N/A

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.)*

N/A

2. Validity – See attached Measure Testing Submission Form

Urology_ASC_NQF_Testing_Attachment__082418_FINAL2.2-636709715462406801.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: Hospital Visits after Urology Ambulatory Surgical Center Procedures
Date of Submission:

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 in order to identify appropriate surgical procedures for inclusion, comorbidities for risk adjustment, and the outcome of hospital visits. 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 Fiscal Year (FY) 2015 (October 1,2014 – September 30, 2015) Medicare claims data from the Health Account Joint Information (HAJI) database that included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims.

a. Datasets used to define the cohort:

-Outpatient urology procedures performed at ASCs were identified using the full set of Medicare beneficiaries' claims from the FY 2015 Carrier non-institutional claims, which included the ASC facility claims.
-Enrollment database and denominator files: These datasets contain Medicare Fee-For-Service (FFS) enrollment, demographic, and death information for Medicare beneficiaries used to determine inclusion criteria.

b. Datasets used to capture the outcome (hospital visits):

-The outcomes of emergency department (ED) visits and observation stays after urology ASC procedures were identified from FY 2015 hospital outpatient institutional claims, and inpatient hospital admissions from FY 2015 inpatient institutional claims.

c. Datasets used to identify comorbidities for risk adjustment:

-Inpatient and outpatient claims (institutional and non-institutional carrier) data from the year prior (FY 2014: October 1, 2013 – September 30, 2014) were used to identify comorbidities for risk adjustment for these patients.

2. To test facility-level variation in the measure score, we also used the FYs 2014-2015 Medicare claims data from the HAJI database that included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims.
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) 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 overall measure score reliability (split-sample) for a 2-year reporting period, we used 4 years of Medicare claims data from the HAJI database for FYs 2012-2015 (October 1, 2011 – September 30, 2015). We created two patient samples per facility that were equivalent in size to 2 years of data.
5. To calculate individual facility-level measure score reliability for a 2-year reporting period, we used the FY 2014-2015 Medicare claims data from the HAJI database that included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims.

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

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

The dates of the data vary by testing type. Our data spanned across FYs 2011 – 2015 (October 1, 2010 – September 30, 2015). More information is provided in Section 1.7.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> 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 <u>measured entities</u> 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 (ASCs) varied by testing type; see Section 1.7 for details.

1.6. How many and which <u>patients</u> 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 Medicare claims data from the HAJI database that included Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims to develop and test the measure. The measure cohort inclusion and exclusion criteria are specified in the Intent to Submit Form, Sections S.5 to S.9.

The datasets, number of measured entities, number of urology procedures, and demographic profile for the patients used in each type of testing are as follows:

<u>1. Medicare FFS FY 2015 Dataset</u>

-Dates: October 1, 2014 – September 30, 2015

-Number of facilities: 1,062 ASCs
-Number of urology procedures: 65,169
-Demographic characteristics: average age of 75.53 years; 31% female
-Dataset used for: defining the cohort, testing the exclusion criteria, disparities testing

2. Development Sample and Validation Sample

The Development and Validation Samples were derived by selecting two random samples from the Medicare FFS FY 2015 Dataset. The Development Sample included 70% of the urology ASC procedures in the Medicare FFS FY 2015 Dataset, and the Validation Sample included 30% of the urology ASC procedures in the Medicare FFS FY 2015 Dataset

Development Sample

-Dates: October 1, 2014 – September 30, 2015

-Number of facilities: 1,017 ASCs
-Number of urology procedures: 45,619
-Demographic characteristics: average age of 75.52 years; 31% female
-Dataset used for: evaluating the consistency of data elements, and testing the patient-level risk-adjustment model

Validation Sample

-Dates: October 1, 2014 – September 30, 2015

-Number of facilities: 905 ASCs
-Number of urology procedures: 19,550
-Demographic characteristics: average age of 75.53 years; 31% female
-Dataset used for: evaluating the consistency of data elements, validating the patient-level risk-adjustment model

3. Medicare FFS FYs 2014-2015 Dataset

-Dates: October 1, 2013 – September 30, 2015

-Number of facilities: 1,204 ASCs
-Number of urology procedures: 130,144
-Demographic characteristics: average age of 75.6 years; 30% female
-Dataset used for: performance measure score testing; facility-level reliability testing

4. Medicare FFS FYs 2011-2015 Dataset

To calculate measure score reliability for a 2-year reporting period, we used 4 years of Medicare claims data. Using this Medicare FFS FYs 2011-2015 Dataset, we created two patient samples per facility that were equivalent in size to 2 years of data.

-Dates: October 1, 2011 – September 30, 2015

-Number of facilities: 1,490 ASCs
-Number of urology procedures: 255,137
-Demographic characteristics: average age of 75.7 years; 29.7% female

-Dataset used for: measure score reliability

Note: The total data needed to calculate the measure spans 3 years and 7 days. The measure aggregates 2 years of procedures for measure score calculation – that is, it uses a 2-year measurement period. 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 FY 2011 data to gather risk factors for patients undergoing procedures in FY 2012). Finally, the measure uses 7 days post each year to gather the outcome of unplanned hospital visits.

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 sociodemographic status (SDS) variables (Medicaid dual-eligibility status and African-American race) and a composite measure (the AHRQ-validated Socioeconomic Status [SES] index score). In addition, we examined the facility-level proportions of dual-eligible patients, of African-American patients, and of low-SES patients based on the AHRQ SES Index. These analyses were performed with the Medicare FFS FY 2015 Dataset and data from the Census Bureau's American Community Survey.

We selected social risk factors and race variables 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.

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 IMPACT Act [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.

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)

Previous studies examining the validity of data on patients' race and ethnicity 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 using 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. Additionally, patients' dual eligibility for Medicare and Medicaid is an indicator whose data are readily available for use. 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 that patients live in. 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. *American Journal of Otolaryngology*. 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. *American Journal of Medical Quality*. 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*)

Measure Score Reliability

We performed two types of reliability testing.

Both reliability testing methods described in this section use a facility volume cutoff of 35 procedures. Our rationale for this is described below.

In general, CMS sets the volume cutoff for publicly reporting facility measures scores based on two considerations. CMS considers the empiric results of reliability testing conducted on the dataset used for public reporting. CMS also considers the volume cutoff for score reporting used for related measures (for example, Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy) and seeks to align where possible the cutoffs for similar measures that are concurrently reported. This measure (and the related measure, NQF 3470, Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures , also under NQF review in this cycle) are both currently in a confidential national "dry run." For the dry run, CMS empirically determined that measure scores for facilities with 35 or more procedures are reliable. Regardless of the score reporting volume cutoff, all facilities and their cases are used in calculating the measure scores. In the dry run, CMS will report scores for ASCs with fewer procedures than the volume cutoff of 35 procedures as having "too few cases" to support a reliable estimate. In summary, the measure specifications do not prejudge the ideal volume cutoff. The minimum sample size for public reporting is a policy choice that balances considerations such as the facility-level reliability testing results on the reporting data and consistency across measures for consumers.

First, we estimated the overall measure score reliability by calculating the intra-class correlation coefficient (ICC)using a split-sample (i.e. test-retest) method. This form of measure reliability testing evaluates, on a

whole, how reliable measure results are across all facilities. To calculate the ICC, we used the Medicare FFS FY 2012-2015 Dataset. For ASCs with two or more urology procedures, procedures were randomly split into the two samples (2 years of combined data for each sample). The ICC evaluates the agreement between the risk-standardized hospital visit rates (RSHVRs) calculated in the two randomly selected samples. The ICC was estimated as ICC (2, 1), described in Shrout and Fleiss [1], and assessed using conventional standards [2].

Second, we estimate the facility-level reliability. While split-sample reliability is the most relevant metric from the perspective of overall measure reliability, it is also meaningful to consider the separate notion of "unit" reliability, that is, the reliability with which individual units (here, ambulatory surgery centers) are measured. This is because the reliability of any one facility's measure score will vary depending on the number of procedures performed. Facilities with more procedural volume will tend to have more reliable scores, while facilities with less procedural volume will tend to have less reliable scores. Therefore, we also use the formula presented by Adams and colleagues [3] to calculate facility-level reliability as an additional, complementary metric.

References

1. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

2. Landis J, Koch G, The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

3. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

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*)

Measure Score Reliability

Split-sample measure score reliability yielded an ICC [2,1] score of 0.44.

Facility-level reliability testing indicated that for facilities with at least 30 procedures, the median reliability was 0.71 indicating substantial reliability. This includes 560 of 1,204 facilities.

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?*)

Measure Score Reliability Results

Split-Sample Reliability

The ICC [2,1] score of 0.44, calculated for 2 years of data, indicates moderate measure score reliability.

The ICC[2,1] is a conservative measure of split-sample reliability because it assumes that the multiple measurements are drawn from a larger sample of tests, and that the measured providers are drawn from a larger sample of providers. Given the conservative nature of the ICC[2,1] and the complex constructs of risk-adjusted outcome measures, a lower reliability score is expected.

Guidelines for the interpretation of the ICC[2,1] statistic are limited. Landis & Koch [1] created a convention to assess the reliability but stated "In order to maintain consistent nomenclature when describing the relative strength of agreement associated with kappa statistics, the following labels will be assigned to the corresponding ranges of kappa … Although these divisions are clearly arbitrary, they do provide useful "benchmarks".

In other words, 'acceptability' depends on context. For example, if we were measuring adolescent weight twice with the same scale, and assessing whether the weights were above a certain threshold, we would expect the two measurements to agree almost exactly (ICC[2,1] ~ 1); otherwise, we would discard the scale. At the other extreme, if we were measuring a latent personality trait such as a personality disorder, we would

expect a much lower level of agreement. In fact, Nestadt et al. assessed ICCs for several standard tools for assessing personality disorder and found test-retest reliabilities in the range of 0.06-0.27 [2]. Notably, Nestadt et al. conclude that these tools "may still be useful for identifying [personality disorder] constructs."

The current context is measuring provider quality, or, specifically, provider propensity to provide appropriate care as measured by subsequent outcomes. In this context, many factors can impact the quality we are trying to measure at any given point in a health care episode or throughout a measurement period. So, it is likely acceptable to see lower values of reliability given quality's latent and unpredictable nature. Although a reasonable amount of reliability should be sought in these types of measures, they should not be expected to achieve reliability values similar to measures of weight, using the example from above. Most risk-adjusted measures have achieved split-sample reliability values in the fair to moderate range and this could represent a reasonable level for these types of measures. Finally, it is important to note that risk-adjusted outcomes measures do account for their intrinsically lower reliability by using statistical methods accounting for uncertainty through the use of confidence intervals.

Facility-level Reliability

The median reliability score of 0.71, calculated with 2 years of data, is considered "substantial" [1].

The split-sample reliability score of 0.44, discussed in the previous section, represents the lower bound of estimate of the true urology measure reliability. Using the approach used by Adams et al [3], we obtained median reliability score of 0.71. This pattern was also observed by Yu, Mehrotra and Adams [4]. For example, they found mean reliability for a PCP visits utilization measure to be 0.94 using the approach used by Adams and colleagues [3], although the split-sample reliability score was 0.68.

Our interpretation of these results is based on the standards established by Landis and Koch (1977) [1]:

< 0 – Less than chance agreement;

0 – 0.2 Slight agreement;

0.21 – 0.39 Fair agreement;

0.4 – 0.59 Moderate agreement;

0.6 – 0.79 Substantial agreement;

0.8 – 0.99 Almost Perfect agreement; and

1 Perfect agreement

Takien together, these results indicate that there is sufficient reliability in the measure score.

Citations:

1. Landis J, Koch G. The measurement of observer agreement for categorical data, Biometrics 1977;33:159-174.

2. Nestadt G, et al. Concordance between personality disorder assessment methods, Psychol Med, 42:657-667.

3. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

4. Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing**

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

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

We demonstrated measure validity through the application of established measure development guidelines, and through assessment by external groups.

<u>Validity Indicated by Established Measure Development Guidelines:</u>
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 [1], 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" [2].

<u>Validity as Assessed by External Groups:</u>
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 21-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 during measure development in summer 2016 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 in October 2016. This NQF application includes the measure's final specifications, inclusive of the revisions after consideration of the public comments.

Finally, following measure development, the measure also underwent the federal rulemaking process (notice and comment rulemaking). CMS finalized the measure for use in the ASCQR program beginning with CY 2022 payment determination (82 FR 59470).

<u>Face Validity as Determined by the TEP</u>
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 statement via an online survey following the final TEP meeting: "The risk-standardized hospital visit rates obtained from the urology ASC measure as specified can be used to distinguish between better and worse quality facilities." The survey offered participants six response options ranging from "strongly disagree" to "strongly agree."

<u>List of TEP Members</u>

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) Thomas Tsai, MD, MPH – Brigham and Women's Hospital (General Surgeon); Harvard School of Public Health (Research Associate); Boston, MA
15) Katherine Wilson, RN, BA, MHA – AMSURG Corp (Vice President of Quality); Nashville, TN – She was not polled for face validity as she participated in early measure development.

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. National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report. Available at: http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed June 7, 2017.
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.

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

Face Validity as Determined by the TEP:

Validity was assessed by the TEP. The TEP provided input on the cohort, risk model, and outcome to strengthen the measure and supported the final measure with high agreement. A total of 14 TEP members completed the face validity survey. Of the 14 respondents, 12 respondents (86%) indicated that they somewhat, moderately, or strongly agreed and 2 respondents moderately disagreed with the following statement: "The risk-standardized hospital visits rates obtained from the urology ASC measure, as specified,

can be used to distinguish between better and worse quality facilities." (Note: One TEP member was not polled as she only participated in the early stages of measure development.)

As mentioned above, two TEP members "moderately disagreed" with the validity statement. Both members concluded the measure did not adequately differentiate quality of care in ASCs. One TEP member was specifically concerned about the utility of the measure in differentiating performance given the small number of facilities identified as outliers in measure testing. Specifically, we had identified outliers by estimating an interval estimate (similar to a confidence interval) around each facility score and counted those facilities that had a 95% interval estimate entirely above or entirely below the national crude rate using two years of data for 1,204 facilities and 130,144 procedures. This is based on CMS's conservative approach to assign scores to one of three performance categories for risk-adjusted hospital and outpatient outcome measures – 1) "no different than national average," 2) "better than the national average," or 3) "worse than the national average." – with 95% confidence.

However, among the respondents who "strongly agreed" with the validity statement even though there were relatively few outliers, one noted that there is room for improvement among average performing facilities. Another TEP member said that the measure demonstrated an appropriate rate of effectiveness/variation to identify performance across ASCs. Facility measure scores ranged from 3.7% to 10.1%, with a median RSHVR of 5.8% (the 25th and 75th percentiles were 5.6% and 6.1%, respectively).

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 a high level of agreement (86%) with the overall face validity of the measure. Measure validity is also ensured through the processes employed during development, including obtaining regular input from the TEP and modeling methods with demonstrated validity used in claims-based measures.

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 (urology procedures and cystoscopy with intervention 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 FY 2015 Dataset resulted in an initial cohort of 65,194 ASC urology procedures. We then applied the following exclusion criteria (see the Intent to Submit 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 urology ASC procedure.

This resulted in excluding 25 (0.04%) urology ASC procedures. Thus, the final Medicare FFS FY 2015 Dataset included 65,169 urology ASC procedures performed at 1,062 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 urology 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 (0.04%) of urology 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** 9 **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 risk-standardized hospital visit rates (RSHVRs). This approach accounts for the clustering of patients within ASCs and variation in sample size across ASCs.

The risk-adjustment model has 9 variables (age, six comorbidity variables, and two surgical variables, including a surgical complexity variable – Work RVU of the procedure). 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. With the exception of benign prostatic hyperplasia with obstruction, which we define using individual (ICD-9) diagnosis codes, we define comorbidity variables using CMS Condition Categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9 and ICD-10 diagnosis codes.

Model Variables:

1. Age (years > 65)
2. Work Relative Value Units (work RVUs)
3. Benign prostatic hyperplasia with obstruction (ICD-9-CM diagnosis codes 60001, 60021, 60091; ICD-10-CM diagnosis codes N401, N403)
4. Complications of specified implanted device or graft (CC 176)
5. Number of qualifying procedures: 1, 2, 3 or more
6. Poisonings and allergic and inflammatory reactions (CC 175)
7. Major symptoms, abnormalities (CC 178)
8. Parkinson's and Huntington's diseases; seizure disorders and convulsions (CC 78, 79)
9. Ischemic heart disease (CC 86, 87, 88, 89)

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

Not applicable. This measure is risk-adjusted.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of*

p<0.10; correlation of x or higher; patient factors should be present at the start of care) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

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 2b3.3a 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 rates. 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 orthopedic ASC measure); 2) a focused literature review; and 3) TEP and expert input.

To define the candidate risk factors, we defined the clinical risk factors in claims data using Version 22 of the CCs from CMS's Hierarchical Condition Categories (HCC) grouper, which classifies over 15,000 ICD-9 diagnosis codes into 200 clinically coherent and mutually exclusive groups of codes, or condition categories [3]. In some cases (for example, morbid obesity), individual ICD-9 codes were used to define the risk factor. The measure does not apply the hierarchical logic of the HCC. Based on prior validation work conducted for similar measures, we have confidence that model variables defined using the CCs are reasonable proxies for clinical conditions. Specifically, as discussed in the response to the next question, CMS has validated similar risk-adjustment models that use the CCs against models that use chart-abstracted data for risk adjustment. Note that we have specified the model in ICD-10 for future use; the process we used to specify the model is described in Section 2b1.2.

To address surgical procedural complexity, we used the work RVUs of the procedure, an approach employed by the American College of Surgeons National Surgical Quality Improvement Program [4].

We reviewed the candidate risk factors with TEP members and clinical consultants. None of the clinical experts suggested removing any of the candidate risk factors from the list. Several TEP members suggested that we consider additional risk adjustment for procedural complexity, beyond work RVU. One TEP member suggested we consider risk adjusting for benign prostatic hyperplasia, nocturia, urinary frequency, use of alpha blockers, and anesthetic type. We reviewed the suggested risk factors and added benign prostatic hyperplasia, nocturia, and urinary frequency. We were not able to include alpha blockers because we do not have data on patient-level medication use, and we did not include anesthetic type because we do not risk adjust for discretionary procedure differences (such as approach to anesthesia or surgical techniques).

Finally, to consolidate similar risk factors, we checked the bivariate direction and strength of association of the individual risk factors defined by CCs or ICD-9 codes and then combined risk factor diagnoses into clinically coherent comorbidity variables. For example, a "cancer" variable was created that combined several individual cancer diagnoses.

Variable Selection

To select the final set of variables to include in the risk-adjustment model, we performed a bootstrap selection method. Briefly, 1,000 samples were selected with replacement from the Development Sample dataset. For each of the 1,000 samples, a parsimonious logistic regression model was selected by iteratively removing non-significant candidate variables from the model using a stepwise purposeful selection approach described by Hosmer and Lemeshow [5]. Our goal was to minimize the number of variables in the model while preserving model performance (as measured by the c-statistic). All variables significant at $p<0.05$ were retained in the final model. This approach led to 1,000 models from which we then selected all variables that entered the model at least 70% of the time for our final model. This allowed us to select variables that reliably and consistently enter the model across the 1,000 bootstrap samples, and avoid spurious relationships that may occur due to low volume and event rate.

In the attached Data Dictionary:

- Sheet "2b3.3a Risk Model Specs" indicates the final risk variables selected, their odds ratios and 95% confidence intervals.
- Sheet "2b3.3a ICD-9 to CC" provides the cross-walk of Version 22 CCs and ICD-9 codes used to define risk variables in the measure.
- Sheet "2b3.3a ICD-10 to CC" provides the cross-walk of Version 22 CCs and ICD-10 codes used to define risk variables in the measure.

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 available in Medicare claims data 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 [6-11] and IMPACT Act–funded work by the National Academy of Science, Engineering and Medicine (NASEM) and the Department of Health and Human Services Assistant Secretary for Policy and Evaluation (ASPE) [12-14].

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, we performed a literature search 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 [5,6].

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 [7]. Alternatively, patients with SES risk factors, such as lower education, may require differentiated care – for example, 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 [8,9]. Similarly, African-American patients have been shown to have less access to high-quality hospitals compared to white patients [10]. 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.

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using General Equivalence Mapping (GEM) software. As part of the annual coding updates for the measure, we have continued to reevaluate ICD-10-based measure specifications. We reviewed the 2016 ICD-10 coding system in detail and enlisted the help of clinicians to select and evaluate which of the ICD-10 codes that mapped to the ICD-9 codes were appropriate for use in this measure. Upon updating the codes, we tested the performance of the measure's risk model, and impact on risk-standardized hospital visit rates at the ASC level in the most recent measurement years of data available. We then solicited input from clinical and measure experts to confirm the clinical appropriateness of the changes to the specifications given the updates to the ICD-10 codes.

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. *Statistical Science.* 2007; 22(2):206-226.
3. Pope, G., et al., Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review, 2000. 21(3):26.
4. 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. *Journal of the American College of Surgeons.* 2010; 211(6):715-723.

5. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York: Wiley; 2000.
6. Bhattacharyya N. Healthcare disparities in revisits for complications after adult tonsillectomy. *American Journal of Otolaryngology*. 2015 Mar-Apr; 36(2):249-253.
7. Menachemi N, Chukmaitov A, Brown LS, et al. Quality of care differs by patient characteristics: outcome disparities after ambulatory surgical procedures. *American Journal of Medical Quality*. 2007 Nov-Dec; 22(6):395-401.
8. Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. *New England Journal of Medicine*. 2014; 371:2298-2308.
9. 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 Affairs*. 2011; 30:1904-1911.
10. Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. *JAMA Surgery*. 2014; 149:475-481.
11. 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.
12. 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.
13. 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.
14. 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?

From 1,000 bootstrapped models, the following candidate variables were selected greater than 70% of the time, and thus we reselected as the final risk-adjustment variables:

1. Age (years > 65)
2. Work Relative Value Units (work RVUs)
3. Benign prostatic hyperplasia with obstruction (ICD-9-CM diagnosis codes 60001, 60021, 60091; ICD-10-CM diagnosis codes N401, N403)
4. Complications of specified implanted device or graft (CC 176)
5. Number of qualifying procedures: 1, 2, 3 or more
6. Poisonings and allergic and inflammatory reactions (CC 175)
7. Major symptoms, abnormalities (CC 178)
8. Parkinson's and Huntington's diseases; seizure disorders and convulsions (CC 78, 79)
9. Ischemic heart disease (CC 86, 87, 88, 89)

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) Medicaid dual-eligibility 2) race, and 3) the AHRQ SES index. For these analyses we used the Medicare FFS FY 2015 Dataset and data from the Census Bureau's American Community Survey. These data included 1,062 ASC facilities and 65,169 urology procedures. Our goal for these analyses were two-fold: 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 the models' 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 minus 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 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. To further ensure that association are not obscured by a relatively low proportion of patients with social risk factors in the majority of the facilities, we examined relationship of the RSHVR with the proportion of each social risk factor for facilities in the top quartile of patients with that social risk factor. Scatterplots and Pearson correlations are presented.

Results

Observed hospital visit rates were higher for duel-eligible and low SES patients (scores below 42.7 on the AHRQ SES index): 7.5% for dual-eligible patients compared to 5.9% for non-dual-eligible patients, and 6.2% for low SES patients compared to 5.9% for higher SES patients (scores above 42.7 on the AHRQ SES index). African-American patients had lower hospital visit rates when compared to non-African-American patients (5.6% vs. 6.0%, respectively). Furthermore, inclusion of each of these risk factors in our models (controlling for other risk-adjusters in our model) showed a significant disparities association for dual-eligible patients only (dual-eligible: OR: 1.30, 95% CI: 1.13 -1.48, p = 0.0001). No significant association was observed for race or AHRQ SES Index (race: OR: 0.96, 95% CI: 0.83-1.12, p=0.64; AHRQ SES Index: OR: 1.02, 95% CI: 0.92-1.12, p=0.75).

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 (0.999 for dual-eligible, African-American, and low SES patients) and mean differences in RSHVRs were near zero (0.0006, 0.0002, and -0.0002 for dual-eligible, African-American, and low SES patients, respectively). This indicates that including these social risk factors in ASC-level measure scores will result in limited differences in ASC measure results after accounting for other factors (demographic, comorbidities, and surgical procedure complexity) included in the risk model.

Distributions of the measure score for ASCs with a low % of patients with social risk factors (1st quartile) and high % of patients with social risk factors (4th quartile) by each social risk factor are shown in Table 2 below. The results showed slightly higher (worse) measure scores for the 4th quartile ASCs (those with higher proportions of patients with the social risk factors) compared to the 1st quartile ASCs, but the distributions largely overlapped. The median RSHVR varied minimally across quartiles for all three variables (5.8% - 6.0%).

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

	Medicaid dual eligible		African-American race		Low SES	
	1st Quartile (≤1.96%)	4th Quartile (≥7.54%)	1st Quartile (0.00%)	4th Quartile (≥6.34%)	1st Quartile (≤4.93%)	4th Quartile (≥18.30%)
Number of ASCs	106	109	125	108	108	109
Number of patients	11,842	11,705	9,490	17,011	11,286	14,348
Maximum RSHVR	10.8%	8.8%	8.4%	8.8%	10.8%	9.3%
90th	7.3%	7.4%	7.3%	7.4%	6.9%	7.5%
75th	6.4%	6.6%	6.6%	6.6%	6.3%	6.6%
Median	5.9%	6.0%	6.0%	6.0%	5.8%	6.0%
25th	5.3%	5.5%	5.5%	5.5%	5.2%	5.6%
10th	4.8%	5.0%	5.2%	5.0%	4.6%	5.3%
Minimum RSHVR	4.0%	4.5%	4.1%	4.2%	4.0%	4.9%

Finally, for the quartile of facilities with the highest proportion of patients with social risk factors, we examined the relationship between the proportion of patients with each risk factor (x-axis) and the ASC risk-standardized hospital visit rates (RSHVRs) (y-axis) for the measure (Figures 1, 2, and 3 below). The results show that there is no correlation between the proportion of patients with social risk factors and the measure result (all p-values > .05).

Figure 1: Relationship between dual-eligible status and ASC risk-standardized hospital visit rates (RSHVRs) (facilities in the highest quartile for the proportion of dual-eligible patients)

Pearson correlation coefficient: -0.04

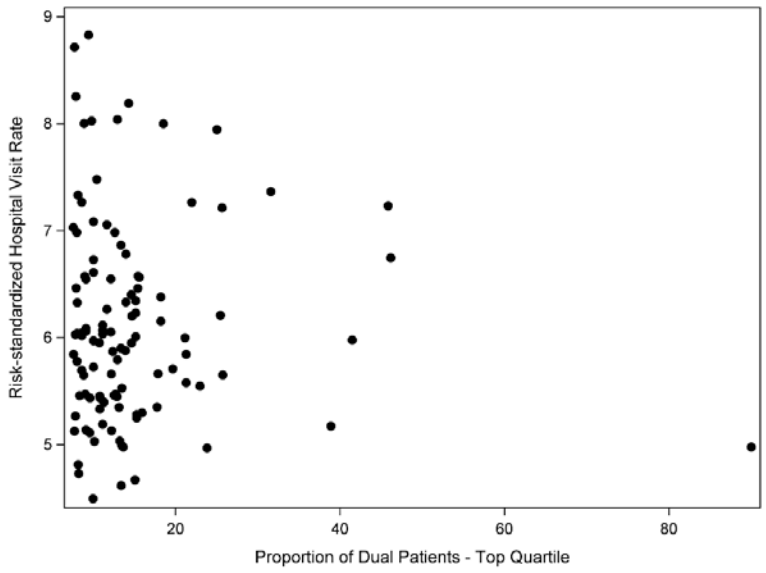


Figure 2: Relationship between race and ASC risk-standardized hospital visit rates (RSHVRs) (facilities in the highest quartile for the proportion of African American patients)

Pearson correlation coefficient = 0.09

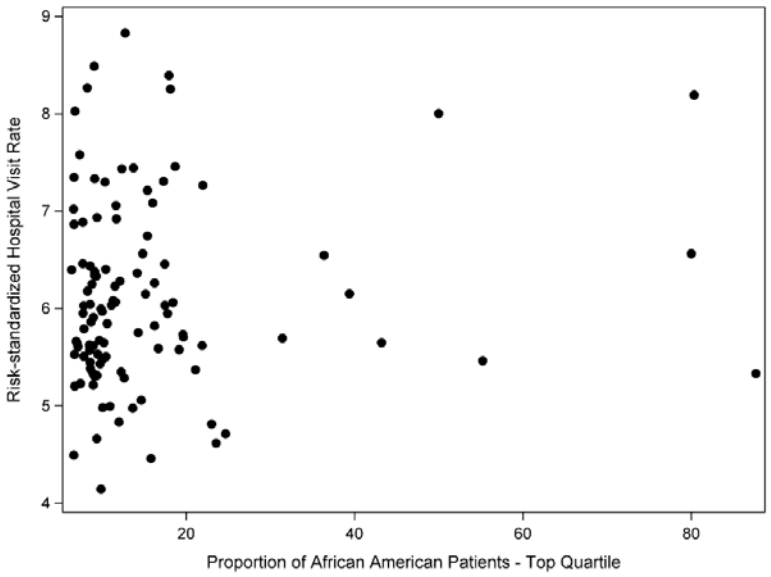
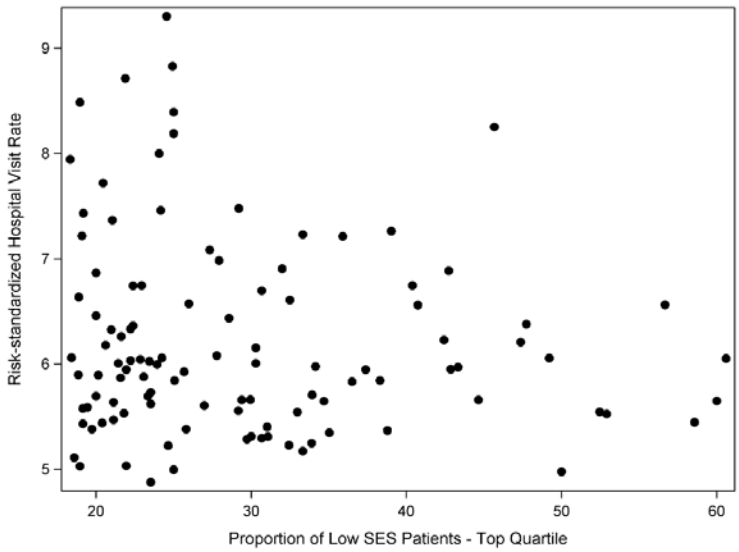


Figure 3: Relationship between socioeconomic status and ASC risk-standardized hospital visit rates (RSHVRs) (facilities in the highest quartile for the proportion of low SES patients).

Pearson correlation coefficient = -0.12



In summary, we conclude that dual-eligible status had a statistically significant association with the risk of a hospital visit, and a small shift in the RSHVR distribution (.1% - .2% at the 10th and 90th percentiles). This association, however, may be a result of either disparate care received due to their socio-demographic status or increased risk of hospital visits not accounted for by risk adjustment, but this cannot be discerned. Furthermore, we observed no substantial impact of dual-eligibility or other patient-level social factors on the ASC-level measure scores, and in facilities in the top quartile for the proportion of patients with each of the social risk factors we do not see a relationship between the measure score and the proportion of patients with the social risk. Based on the above, we do not adjust for these social risk factors.

This is consistent with CMS's decision to not risk adjust or stratify the measure by social risk factors (82 FR 59468) and concern that ASCs should not be held to different standards for patients with social risk factors (82 FR 59446). CMS remains committed to considering options for accounting for social risk factors within individual measures and in the ASCQR program as a whole (82 FR 59447).

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model <u>or</u> 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.

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 rate deciles to assess calibration, and the range of observed hospital visit rates between the lowest and highest predicted 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 discrimination (predictive ability) 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 datasets but fails to provide valid predictions in new patients. Estimated calibration values of γ0 far from 0 and estimated values of γ1 far from 1 provide evidence of over-fitting.

If stratified, skip to 2b3.9

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

Development Sample results:

c-statistic=0.610
Predictive ability (lowest decile %, highest decile %): 3.2% to 11.2%

Validation Sample results:

c-statistic=0.615
Predictive ability (lowest decile %, highest decile %): 3.1% to 11.4%

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

FY 2015 Development Sample results:

Calibration: (0, 1)

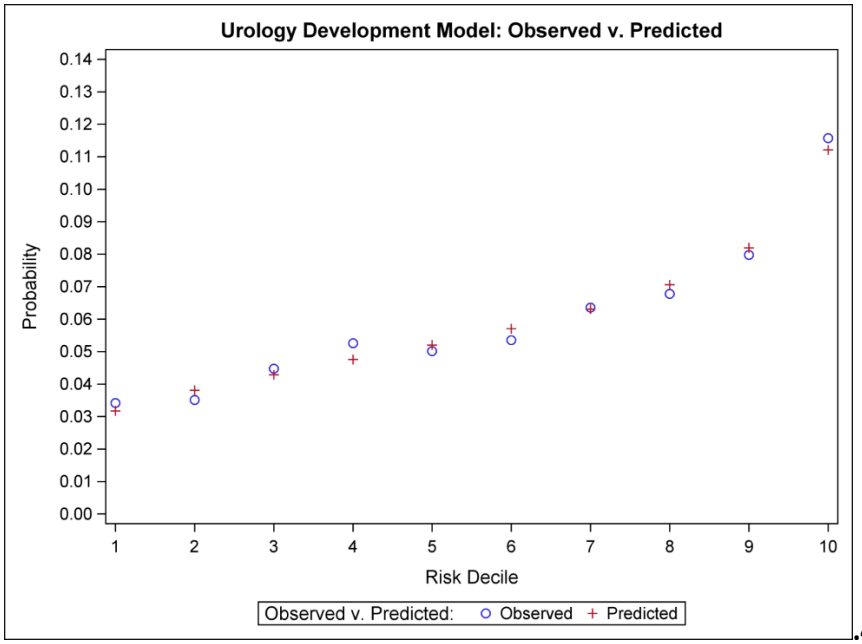
FY 2015 Validation Sample results:

Calibration: (-0.05, 0.98)

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

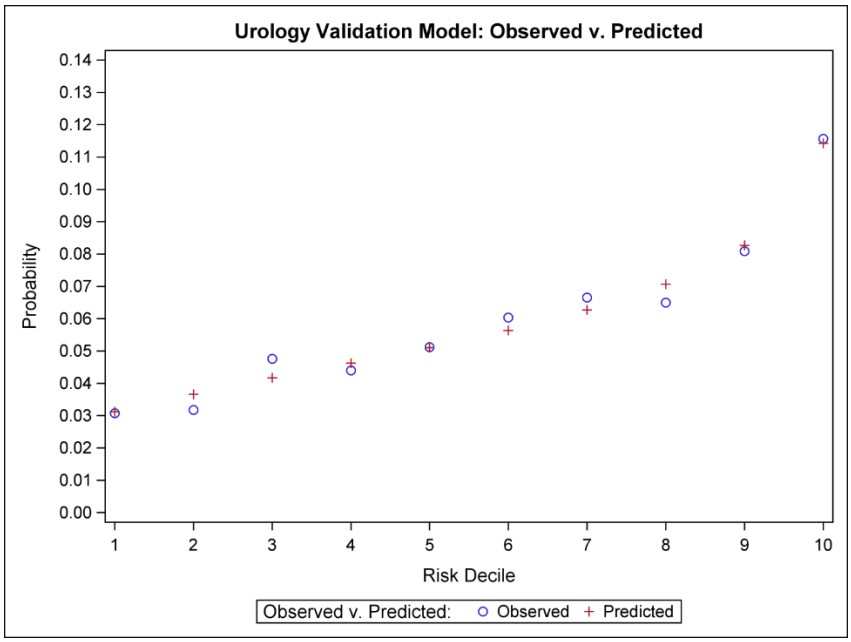
Below are plots of observed vs. predicted values for the hospital visit outcomes across deciles of patient risk in the Development Sample (Figure 1) and Validation Sample (Figure 2). The plots, which showed that the predicted risk closely approximated the observed risk in most deciles, suggest reasonable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

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



.4b

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



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-statistic in the Development Sample was 0.610; the c-statistic in the Validation Sample was slightly higher (0.615). Although, the c-statistic represents good discrimination, it is important to note that we evaluated over 75 potential risk-adjusters and only 9 were significantly associated with the outcome. While the c-statistic may seem low compared to other risk-adjusted measures, this can be explained, in part, by the outcome we are evaluating. If the outcome is more strongly related to quality of care rather than patient characteristics, we would expect lower c-statistics because patient factors are less predictive of the outcome. From this measure, urinary retention (our most frequent reason for returning to the hospital) is a good

example. Our Technical Expert Panel indicated that urinary retention is particularly hard to predict and is related to operative and post-operative care. Therefore, in this instance, our c-statistic seems adequate for the outcome under study.

The risk decile plots, which showed that the predicted risk closely approximated the observed risk in most deciles, suggest good calibration. The mean predicted unplanned hospital visit rate in the Development Sample ranged from 3.2% in the lowest decile of predicted urology procedure hospital visits to 11.2% in the highest predicted risk decile, a range of 8.0%; 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*)

Below in Table 2, we include information on the consistency of data elements used in risk adjustment.

Table 2 shows the frequencies across the two split samples for all variables included in the final model. According to the results presented below, frequencies of the risk variables were similar in the Development and Validation Samples, indicating good variable consistency.

Table 2: Risk Variable Frequencies, Development and Validation Samples (Medicare 20% FFS Cohort)

Variable (definition)	Development Sample, 10/01/2014 – 09/30/2015		Validation Sample, 10/01/2014 – 09/30/2015	
	#	%	#	%
Number of procedures	45,619	-	19,550	-
Age: mean (standard deviation [SD])	75.52	6.7%	75.53	6.7%
Work Relative Value Units: mean (SD)	6.28	3.7%	6.26	3.6%
Benign prostatic hyperplasia with obstruction	14,441	31.7%	6,110	31.3%
Complications of specified implanted device or graft (This variable includes codes for complications of implanted devices and grafts, including cystostomies, vascular and genitourinary devices, and urethral catheters.)	2,309	5.1%	1,021	5.2%
Number of qualifying procedures: 1	41,176	90.3%	17,587	90.0%
Number of qualifying procedures: 2	4,094	9.0%	1,807	9.2%
Number of qualifying procedures: 3 or more	349	0.8%	156	0.8%
Poisonings and inflammatory allergic reactions (This variable includes codes for adverse drug effects and allergies).	2,434	5.3%	1,013	5.2%
Major symptoms, abnormalities (This variable includes diagnoses of fever, sleep disorders, altered consciousness, and abdominal pain.)	32,921	72.2%	13,989	71.6%
Parkinson's and Huntington's diseases; seizure disorders and convulsions	1,697	3.7%	722	3.7%
Ischemic heart disease	15,295	33.5%	6,558	33.5%

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, multiplied by the national observed rate of unplanned hospital visits. 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 urology 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 greater than one indicates that the ASC's patients and 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.

We characterize the degree of variability by:

1) Providing the median odds ratio (MOR) [1]. The median odds ratio represents the median increase in odds of a hospital visit if a procedure on a single patient was performed at a higher risk ASC compared to a lower risk ASC. It is calculated by taking all possible combinations of ASCs, always comparing the higher risk ASC to the lower risk ASC. The MOR is interpreted as a traditional odds ratio would be.
2) Reporting the distribution of the RSHVR.
3) Because the measure score is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate to determine if an ASC is performing better than, worse than, or no different from expected. An ASC is considered better than expected if its entire risk-standardized rate interval estimate falls below the national mean, and considered worse if the entire confidence interval falls above the national mean. It is considered no different if the confidence interval overlaps the national mean.

Reference:

1. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, Råstam L, Larsen K. (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health, 60(4):290-7.

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 **median odds ratio** was 1.27.

The risk-standardized measure scores estimated using Medicare FFS data (FYs 2014-2015) had a median value of 5.8%. The values ranged from 3.7% to 10.1%. The **percentiles of the distribution** were as follows:

Min	1st	5th	10th	25TH	50TH	75TH	90TH	95TH	99TH	Max
3.7%	4.6%	5.0%	5.2%	5.6%	5.8%	6.1%	6.6%	7.0%	8.2%	10.1%

Using a bootstrapped 95% interval estimate, we found 19 significant outliers among 1,204 ASCs. Of the 1,204 ASCs, 4 were categorized as better than expected, 15 as worse than expected, and 1,185 as no different than expected.

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

The median odds ratio suggests a meaningful increase in the risk of a hospital visit if a procedure was performed at a higher risk ASC compared to a lower risk ASC. A value of 1.27 indicates that a patient has a 27% increase in the odds of a hospital visit if the same procedure was performed at higher risk ASC compared to a lower risk ASC indicating the impact of quality on the outcome rate is substantial.

The median RSHVR is 5.8% which indicates that patients are expected to have an ED visit, observation stay, or admission to the hospital after an ASC urology procedure on average 5.8% of the time. During our Technical Expert Panel, many participants indicated that this rate was too high given that the expectation for ASC-based procedures is that patients selected for the procedures will not need follow-up acute care, and that ultimately the goal should be near 0. Further, the 10th and 90th percentiles (5.2 and 6.6) represent substantial deviations from this median: a facility performing at the 10th percentile is performing 10% better than an average performer, while a facility performing at the 90th percentile is performing nearly 14% worse than an average performer. Furthermore, the best performing ASCs (3.7%) are performing 36% better than an average performer, while the worst performing ASCs (10.1%) are performing 75% worse than an average performer. This variation shows a clear quality gap, as some facilities can achieve substantially lower rates than the average performer, while other facilities are performing worse than an average performer. It is important to note that here the average performer refers to an ASC with the same case and procedure mix performing at the average.

We identified few outliers, which is expected given the measure's low outcome rate and conservative 95% CIs. The measure's low outcome rate (combined with lower volumes) will reduce the precision of estimates leading to wider confidence intervals. This, however, does not diminish the importance of the measure; we observed many avoidable complications as part of the outcome and substantial variance in both observed and risk-adjusted rates among ASCs. Identifying those facilities that are outliers with a very high degree of confidence using the 95% CI can be informative to consumers and ASCs; CMS is currently discussing expanding the performance period to three years which could increase the number of outliers that can be detected in the measure.

Overall, our results suggest that there is substantial need to both reduce the expected rate and the variation in rates across ASCs, and that this improvement goal is achievable.

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.***

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

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**.

ALL data elements are in defined fields in a combination of 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).

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.

Measure development and testing show that the measure cohort can be defined and outcomes can be reported using routinely collected Medicare claims data. This measure is not yet in operational use.

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)
Public Reporting 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 it only recently completed development and will be submitted to the National Quality Forum (NQF) for initial endorsement. The measure is in the process of being implemented in the Ambulatory Surgical Center Quality Reporting (ASCQR) program, for public reporting in January 2022.

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.*)

CMS is in the process of implementing this measure for public reporting in calendar year 2022 in the Ambulatory Surgical Center Quality Reporting (ASCQR) Program. CMS recently completed a confidential reporting period for ASCs.

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.

All ASCs with at least one eligible case were included in the 2018 dry run, n=1,149.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

The Centers for Medicare & Medicaid Services (CMS) recently conducted a confidential reporting (dry run) for Hospital Visits after Urology ASC Procedures (urology ASC measure [ASC-18]). These measures will be included in the Ambulatory Surgical Center Quality Reporting (ASCQR) Program for Calendar Year (CY) 2022 payment determination and public reporting.

The objectives of the dry run were to educate ASCs and other stakeholders about the measure, allow facilities to review their measure results and data confidentially prior to public reporting, answer questions from facilities and other stakeholders, test the production and reporting process, and identify potential changes to measure specifications.

The dry run occurred August 1 through August 30, 2018. At the beginning of the dry run, ASCs received their measure results in a Facility-Specific Report (FSR) for the October 1, 2015 through September 30, 2017 performance period. CMS took comments on the measure during a national provider call and answered questions through a question and answer (Q&A) email inbox. CMS worked with facilities to help them

understand the measure methodology and their own data. Throughout the dry run, CMS received recommendations on measure refinements.

The steps in the dry run process are outlined below. For the dry run, CMS:

1. Announced the dry run to ASCs, Quality Improvement Organizations/Quality Innovation Networks (QIO/QINs), and other stakeholders via email blasts on CMS listservs and presentations during existing monthly ASC and QIO/QIN webinars. The announcements included information on the dry run timeline and process, a measure overview, and the process for accessing a Facility-Specific Report (FSR). CMS and the CORE team encouraged facilities to participate and provided contact information to ask questions and provide feedback.

2. Prepared and posted resource materials on the QualityNet website prior to the start of the dry run. Specifically, CMS posted:

Measure Technical Reports: provide the background and rationale for the development of the orthopedic and urology ASC measures, describes the approach to risk model development and testing, and provides detailed measure specifications.

Measure Updates Reports: provide a description of the 2018 measure refinements and coding updates made for the dry run of the measures.

Measure Code Sets: provide the codes used to define the cohort, risk adjustment and outcome for each measure.

Mock Facility-Specific Report: MS Excel® file containing real national data and mock state results, facility-specific results, and patient-level data. This report serves as an example of the FSR each participating facility will receive for the 2018 dry run.

Facility-Specific Report User Guide: provides facilities with interpretation of the measures and measure results, an overview of the measure methodologies, instructions for interpreting the FSR, and links to the resources on QualityNet.

Five Frequently Asked Questions (FAQs): list of FAQs and responses; includes both general questions related to dry run processes and measure specifications questions.

Measure-Specific Fact Sheets: a short document that provides a general overview for each measure, a brief description of the development and purpose of the measure, instructions regarding how to participate in the dry run, links to QualityNet resources, and the Q&A email addresses.

Dry Run Timeline: graphic that shows important dates for the dry run.

Condition Category Crosswalks: provides the assignment of International Classification of Diseases, 9th and 10th edition (ICD-9 and ICD-10) codes to condition categories used to adjust for patient risk factors.

3. Provided facilities with their results, including:

Confidential FSRs that contained national, state, and facility performance results, patient-level data, and case mix information for a facility's patients compared to other facilities in the same state and in the United States, for the October 1, 2015 through September 30, 2017 performance period.

FSR User Guide that provided facilities with interpretation of the measures and measure results, an overview of the measure methodologies, instructions for interpreting the FSR, and links to additional resources. CMS provided these files confidentially through the QualityNet Secure Portal the week of August 1, 2018. QIO/QINs received a summary of results for all eligible facilities in their respective states.

4. Responded to all stakeholder Q&A inquiries before and throughout the dry run: The dry run announcements distributed prior to and during the dry run informed stakeholders of the Q&A period and provided instructions on how to submit comments and questions.

CMS directed facilities and stakeholders to send their comments and questions to an email inbox (ascmeasures@yale.edu).

CMS responded to each email received.

5. Conducted a National Provider Call to present the measure methodology, dry run process, plans for measure implementation, and answer stakeholder questions. CMS hosted the call through a webinar presentation on August 21, 2018. CMS informed stakeholders of the National Provider Call through email notifications and webinar announcements. CMS also posted information about the call on the QualityNet website.

CMS posted a recording and transcript for the call along with the agenda and slides, on the QualityNet website.

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.

Feedback on both the ASC orthopedic measure (NQF #3470), also being reviewed in this cycle by the Readmissions Standing Committee) and this measure (urology) was obtained through distribution of facility-specific reports (FSRs), a national provider call, and Email Q&A during the confidential reporting period

FSR Reports

CMS created and distributed an FSR report for all 2,699 ASCs that were open and had at least one qualifying orthopedic or urology case. CMS successfully uploaded the reports to 1,862 ASCs that had an active QualityNet Secure portal account holder with the designated role of Security Administrator to receive FSRs. Of these, 511 (27.4%) had at least one QualityNet user successfully download their report.

National Provider Call

CMS held a National Provider Call on August 21, 2018 from 2:00 to 4:00 PM ET with 299 participants. The topics of the questions and comments were similar to those received through the Q&A inbox. CMS has made the National Provider Call materials, audio recording, and transcript publicly available on the QualityNet website.

Email Q&A Period

CMS received 117 inquiries via the ASC measure email inbox (ascmeasures@yale.edu) before and during the dry run. Facilities inquired about how to participate in the dry run, how to access the FSR, interpretation of their results, and the measure's methodology. CMS worked with these facilities to answer their questions in more detail, as described in the next section. CMS responded to every inquiry received.

CMS received most of the questions from ASCs (99.1%); and received one question from a QIO.

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

Summary of National Provider Call

The majority of the questions asked during the August 21st National Provider Call were regarding the measure's methodology, such as the definition of the ASC procedures included in the measures' cohort, the definition of an unplanned hospital visit, risk adjustment methodologies, and interpreting the measures' results. Other questions pertained to the implementation of the measures in the ASCQR program, how facilities can access their FSRs and other dry run materials, and how ASCs can participate in the dry run.

In summary, there were a total of 17 questions: 12 questions regarding measure methodology, four regarding the dry run itself, and one regarding implementation.

Summary of Questions and Comments Received During the Dry Run Period

CMS received and responded to a variety of questions via the measure inbox from July 12 to August 30, 2018. The most common types of questions were requests to reupload FSRs (46.5%), followed by questions about the dry run process (18.1%), and requests for publicly available materials (7.1%).

Specific Feedback

One facility identified a case in their results where the patient was seen in the emergency department prior to the ASC procedure, which was counted as a hospital visit outcome. This ASC recommended updating the measure algorithm to differentiate between emergency department visits that occur before the ASC procedure versus after the ASC procedure when the hospital visit, and the ASC procedure occur on the same day.

Response provided to facility: The measures consider hospital visits occurring on the day of the procedure (day 0) an outcome since the vast majority of these admissions occur directly following the procedure; however, on rare occasion these hospital visits may occur before the ASC procedure. We appreciate facilities' feedback and will consider these cases in the refinement of the measure. Any refinements of the measure will be publicly posted along with other potential updates of the measure specifications.

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

CMS received most of the questions from ASCs (99.1%); CMS received one question from a QIO.

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.

We did not yet refine the measure based on this feedback, as the measure has not undergone any updates following the recent (September 2018) dry run.

Together with CMS, we will assess all feedback raised during the dry run and will consider incorporating changes in future measure updates, prior to public reporting. We will meet with clinical experts to determine how best to refine the measures based on the dry run. Throughout this process CMS and its contracted developers will continue to ensure that, as the measures' refinements are made, the measures continue to be accurate and valid for measuring outcomes and quality of care.

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.

Since this measure is not currently in use, there are no trend results available to assess improvement.

We expect there to be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with orthopedic procedures performed at ASCs and follow-up care by capturing, and making more visible to providers and patients, unplanned hospital visits following orthopedic procedures performed at ASCs.

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; this measure is in the process of being implemented. To date, the feedback from facilities does not suggest any unintended impacts on patients. In designing the measure, we sought to minimize the potential of this measure to result in the denial of future care to high-risk individuals. We developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of hospital visits will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time.

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

Not applicable; this measure is in the process of being implemented.

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

3357 : Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers

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

Submitted to NQF in same Fall 2018 cycle: Hospital Visits after ASC Orthopedic Procedures

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:** Urology_ASC_NQF_Appendix.pdf

Contact Information

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

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

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

Co.4 Point of Contact: Dr. Doris, Peter, doris.peter@yale.edu

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

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

- Katherine Wilson, RN, BA, MHA – AMSURG Corp (Vice President of Quality); Nashville, TN

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 urology procedures and quality measurement to provide input on key methodological decisions.

CORE Measure Development Team

- Faseeha Altaf, MPH – Project Coordinator, CORE

- Haikun Bao, PhD – Analytic Co-Lead, CORE

- Mayur Desai, PhD, MPH – Project Lead – CMS Orthopedic ASC measure, 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 Assistant, CORE

- Craig Parzynski, MS – Analytic Co-Lead, CORE

- Jennifer Schwartz, PhD, MPH – Project Lead – CMS Urology ASC measure, CORE

Consultants

-Robert Becher, MD, MS – Surgical Consultant, CORE

-Simon Kim, MD, MPH – Surgical Consultant, Urologic Oncologist, Assistant Professor, Urology; Case Western Reserve University School of Medicine; Cleveland, OH

-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?

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: Not applicable.