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: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2687

Corresponding Measures:

De.2. Measure Title: Hospital Visits after Hospital Outpatient Surgery

Co.1.1. Measure Steward: The 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 procedure performed at a hospital outpatient department (HOPD) 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: The goal of this measure is to reduce adverse patient outcomes associated with preparation for same-day surgery, the surgery itself, and follow-up care, by capturing and making more visible to providers and patients unplanned hospital visits following outpatient surgery. The measure score provides an assessment of quality that is publicly reported, and also informs quality improvement.

S.4. Numerator Statement: The outcome is all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after

the surgery or 2) an unplanned hospital visit (emergency department [ED] visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the surgical procedure.

S.6. Denominator Statement: Outpatient same-day surgeries performed at HOPDs for Medicare FFS patients aged 65 years and older.

S.8. Denominator Exclusions: 1. Surgeries for patients without continuous enrollment in Medicare FFS Parts A and B in the 7 days after the surgery.

2. Surgeries for patients who have an ED visit on the same day but billed on a separate claim, unless the ED visit has a diagnosis indicative of a complication of care.

3. Surgeries that are billed on the same hospital claim as an emergency department (ED) visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.

4. Surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit.

5. Surgeries that are billed on the same outpatient claim as an observation stay.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Sep 03, 2015 **Most Recent Endorsement Date:** Dec 08, 2015

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

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

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

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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

Summary of prior review in 2015

- The developer provides several strategies and interventions that may reduce unplanned hospital visits after same-day surgery, including appropriate patient selection, patient education, and nausea and pain management.
- The rationale submitted by the developer references literature indicating that 40-60% of admissions after HOPD surgery are due to adverse effects of the surgery, anesthesia, or due to other suspected medical problems such as chest pain reports, and that as many as 40% are

preventable. However, this rationale does not specifically support an association between at least one of the various interventions posited and fewer post-HOPD surgery admissions.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- The developer noted that almost all of the evidence has been updated since the last submission in 2015.
- The goal of this measure is to reduce adverse patient outcomes associated with preparation for same-day surgery, the surgery itself, and follow-up care, by capturing unplanned hospital visits following outpatient surgery and making these visits more visible to providers and patients.
- The developer provides a list of strategies and interventions to improve same-day outpatient surgical procedural quality and reduce unplanned hospital visits following outpatient surgery:
    - Appropriate patient selection
    - Appropriate patient education
    - Improving technical quality, including procedural technique and anesthesia
    - Appropriate management of post-operative nausea, vomiting, and pain
    - Educating patients about potential adverse events, whom to contact with questions, and when and where to seek follow-up care
- The developer provides a list of studies supporting actions providers can take to improve care, specifically:
    - Use of multi-modal approaches for treatment of post-operative pain
    - Routine multi-modal nausea and vomiting prophylaxis
    - Identifying and managing patient-level risk factors, such as prevention of hyperglycemia for patients with diabetes

Question for the Committee:

- Is there at least one thing that the provider can do to achieve a change in the measure results?
- The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?

Guidance from the Evidence Algorithm

The measure assesses performance on a health outcome of all-cause, unplanned hospital visits (box 1) → The relationship between all-cause, unplanned hospital visits and adverse patience outcomes is demonstrated through empirical data (box 2) → Pass

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Scores on this measure are a ratio of predicted hospital visits to expected hospital visits. Lower scores represent higher quality.
- The developer provided a distribution of performance scores (risk-standardized hospital visit ratios (RSHVRs)) for facilities for the performance period of January 1, 2018 – December 31, 2018. The data included 3874 facilities and reported:

	Mean	Std Dev	Min	25th	50th	75th	Max
RSHVRs	1.01	0.15	0.54	0.93	0.99	1.07	2.39

- The developer also reported the performance scores by deciles.

Disparities

- The developer explored disparities by Medicare-Medicaid dual-eligibility status and Agency for Healthcare Research and Quality Socioeconomic Status (AHRQ SES) Index. The measure scores are presented for facilities for the performance period of January 1, 2018 – December 31, 2018.
- The developer reported quartile data for both variables.

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

- No new evidence that I am aware of in addition to what what cited by the developer as new in the report. However, given that ED visits are now part of the measure, is there any data to shed light on the % of ED revisits during the postop period that are related to the procedure?
- There is direct evidence that certain actions will result in the desired outcome. Citations: (1)Hall DE, Arya S, Schmid KK, et al. Association of a Frailty Screening Initiative With Postoperative Survival at 30, 180, and 365 Days. JAMA Surg. 2017;152(3):233-240. doi:10.1001/jamasurg.2016.4219 (2) Long DR, Lihn AL, Friedrich S, et al. Association between intraoperative opioid administration and 30-day readmission: a pre-specified analysis of registry data from a healthcare network in New England. Br J Anaesth. 2018;120(5):1090-

1102. doi:10.1016/j.bja.2017.12.044 (3) Hospital readmission after ambulatory laparoscopic cholecystectomy: incidence and predictors Citation DataJournal of Surgical Research, ISSN: 0022-4804, Vol: 219, Page: 108-115 Publication Year2017

- Seems like appropriate evidence to support that there are interventions the provider can do to impact the measure results.
- Evidence is sufficient. Developer provided substantial additional evidence.
- evidence seems lacking on the ability of this measure to be tied to variation due to preventable adverse events. From personal experience, it seems like a majority of HOPDs converting patients to inpatient stay are often not necessarily due to poor pain management or adverse outcomes but due to issues of billing and prior authorization. Providers may bill as an outpatient procedure for easier prior authorization and then convert patients to inpatient stays for further 'observation' in order to increase revenue as a DRG. In that case, this is more of a measure of billing quality and convenience of booking patients for surgery, than it is of actual perioperative care quality.
- If 40-60% of indications for re-admission are surgery related, then almost half are not
- Open-Ended Response
- Pass - The developer provides a list of studies supporting actions providers can take to improve care.
- Evidence measures support outcome

1b. Performance Gap

- It was unclear to me from the data provided how many statistical outliers were identified with the measure, and what was the range of RSHVR's (and width of CI's) for those different from the rest of the group. This would provide more insight into the importance of the measure both for discrimination and actionability for those that are outliers. The range of max & min are quite wide, but not sure if there is a sample size issue with some hospitals. Furthermore, one wonders what % of these visits are related to surgery? This is cited to be 40-60% in the document which is for inpatient revisits—but may be different when incorporating ED visits as well. As for disparities, it is unclear to me from the quartile-based data provided whether there were disparities based on dual-eligible or AHRQ low SES status.
- Yes, a gap was demonstrated in Figure one such that it does deserve to remain a national performance measure. The social determinants of health were not found to be significantly different therefore they decided not to include them in the final measure.
- There is some performance gap and disparities that warrant a measure.
- Yes performance gap was provided. I would like to see stratification by type of facility - specifically are there issues with safety net hospitals versus others. How do social determinants of health relate to the outcome for this measure?
- n/a
- moderate gap
- Open-Ended Response
- The magnitude of the gap in care does warrant a national performance measure. RSHVRs range from 0.54 to 2.39. Stratified facility-level analysis (1b.4) by AHRQ SES does not reveal

significant disparities. However, when the measure is used to classify sites as better or worse than expects, only about 7% meet these criteria.

- Acknowledges Moderate Gap, agree

1c. Composite Performance Measure

- **Yes, including ED, Obs and inpatient revisits as a composite measure (as opposed to inpatient readmissions only) provides a more comprehensive assessment of revisits that "matter" from both a cost and acuity standpoint. All three "categories" are likely to be influenced by the same (potentially preventable) factors. Weighting is not done here but perhaps should be considered to adjust for both fiscal and severity considerations -- for example, should a hospital with a 10% revisit rate where 90% are to the ED be considered as having the same performance status as another with a 10% overall rate where 90% are readmitted to the inpatient setting?**
- **This is an outcome measure.**
- **N/A**
- **n/a**
- **n/a**
- **concerns surrounding validity still exist**
- **Open-Ended Response**
- **NA**
- **Meets**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Scientific Methods Panel

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by a subgroup of the Scientific Methods Panel. The full Panel accepted the subgroup's rating and did not pull the measure for discussion. A summary of the measure and the Panel evaluation is provided below.

Reliability

- Ratings for reliability: H-5; M-4; L-0; I-0 → Measure passes
- Reliability testing conducted at the measure score level using signal-to-noise analysis (Adams' method) with IQRs; minimum 30 procedures: 0.839 (median); 0.759 (all facilities)
- Reviewers noted few concerns with the clarity of the specifications (e.g., risk adjustment, qualifying events, exclusions), and generally agreed the testing approach was acceptable.

Validity

- Ratings for validity: H-1; M-7; L-1; I-0 → Measure passes
- Validity testing conducted at the measure score level. The measure was compared with hospital-wide readmission rate (HWR) and results indicated a weak positive correlation as expected by developers (0.033, p = 0.07).
- Developer also presented face validity results, however, because this is a maintenance measure, empirical validity testing should be the basis for evaluation.
- Risk model discrimination and calibration: c statistic = 0.684; developer reports good discrimination and predictive ability based on risk decile plot.
- Reviewers expressed concern, but generally accepted the validity testing results as a weak, but acceptable, demonstration of validity.

Questions for the Committee regarding reliability:

- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

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

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

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

2a1. Reliability-Specifications

- None- this part of there methods appear fairly tight.
- The reliability was well described. I have no concerns about the likelihood that this measure can be consistently implemented.
- Agree that some specification clarity concerns but overall no concerns about implimentation.
- Acceptable reliability
- how are urgent care visits assessed within the logic?
- concern about linking readmission to surgery reliability
- Open-Ended Response
- To quote a SMP member "For facilities with more than 30 procedures, the reliability is quite high. Although the application suggests that CMS will use the 30 procedure cutoff when this measure is used for public reporting, there is no assurance that CMS will not lower the cutoff to, say, 10, when the reliability of this measure would likely be questionable." The specification raising some concern about consistent implementation.
- Meets
- No
- None
- No concerns about reliability.
- No
- n/a
- yes: how specific can claims be to link readmission to the index procedure
- Open-Ended Response
- Basically sound but only applicable to entities with >30 cases
- No concerns

2b. Validity -Testing

- No
- I would like to know how they take into consideration if a patient has a complication but goes to another hospital for treatment. Would this complication be counted and attributed to the correct hospital?
- No concerns about validity.
- No
- It seems that the cases selected for inpatient surgery generally are a different type and severity than those completed on HOPD settings and I'm not sure I fully understand the use of HWR for validity testing. Additionally specific procedures in HOPD may have different probability for hospital revisits. This may not be reflective for what a patient is looking for if they are undergoing a sports medicine procedure vs a urological procedure.
- The AUC is fair; doesn't have value >0.7; not sure if 50% of cases are non-operative related readmission how important this measure is
- Open-Ended Response
- Overall, the empirical validity testing was not very impressive. I agree with the stated concerns and suggestions of the SMP. More could have been done for empirical validity testing.
- Minimal concerns
- It was unclear to me from the data provided how many statistical outliers were identified with the measure, and what was the range of RSHVR's (and width of CI's) for those different from the rest of the group. This would provide more insight into the importance of the measure both for discrimination and actionability for those that are outliers. The range of max & min are quite wide, but not sure if there is a sample size issue with some hospitals. Furthermore, one wonders what % of these visits are related to surgery? This is cited to be 40-60% in the document which is for inpatient revisits—but may be different when incorporating ED visits as well. Missing data does nat appear to be an issue here.
- The data sources are reliable so they are not a threat to the validity. There is little opportunity for missing data. except for an incomplete medical record not providing medical diagnoses. Presumably, this occurrence would happen in both groups.
- No apparent threats to validity.
- No concerns
- I have concerns that if only 6% of facilitiates are statistically significantly worse or better than average, that this will allow minimal opportunity for hospitals to drive improvement and the benefit this has for patients seeking care. Will they assume ASCs are better to go to than HOPDs or that they should only pursue Inpatient care, despite increases in costs? What are some of the unforseen consequences. Concerned that if this measure failed to pass face validity, the benefit it will have towards consumers.
- no
- Open-Ended Response
- No concerns

- High capture rate for this data
- This is greatly limited by use of admin data - hence the mediocre c-stat. My concern is less risk-adjustment and more about the % (and hospital level variability) of events truly related to the procedure. This may be one of the reasons the developer chose to use a 7 day f/u period.
- The exclusions are well-justified. My biggest concern was the exclusion of emergency cases. Perhaps in the future, it would be interesting to see how well the healthcare system can compensate for patients who are not medically optimized. In a way, this would be analyzing how well centers can handle stress (i.e. the best trauma center).
- Some clarity issues with exclusions but no threats to validity.
- Seems robust and constructed with available data from claims
- n/a
- there is no discrimination about the type of outpatient surgery and risk adjustment with the readmission event
- Open-Ended Response
- Exclusions are complex but seem sensible. The risk model development and validation process was sound. . The tests of model adequacy were good. The analysis and rationale for not including social determinants was reasonable.
- Potential exclusion based on facilities selected

2c. Composite Performance Measure

- See previous response.
- not applicable
- N/A
- n/a
- n/a
- yes
- Open-Ended Response
- NA
- Meets

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.

- ALL data elements are in defined fields in electronic claims and the data are routinely collected as part of the billing process.
- The data are coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

- No fees, licensing, or other requirements reported to use any aspect of the measure as specified.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

- No- the data for the measure is already being collected as routine by hospitals.
- I have no concerns about the data collection collection strategy. The information they are obtaining are all present in the medical and billing records.
- No concerns with feasibility.
- Data elements are all available from hospital billing sources
- n/a
- very feasible- claims data; all hospitals report this information
- Open-Ended Response
- No concerns
- This data is generally through electronic sources

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

Accountability program details

Hospital outpatient quality reporting program (HOQR) is a national pay-for-quality-data-reporting program mandated by the Tax Relief and Health Care Act of 2006. This act requires hospitals to submit data on measures on the quality of care furnished by hospitals in outpatient settings. The HOQR program provides hospitals with a financial incentive to report their quality of care measure data and CMS with data to help Medicare beneficiaries make more informed decisions about their health care. The level of measurement is the facility; the setting is the Hospital Outpatient Department.

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

1. CMS has created and distributed measure scores through "facility-specific reports" (FSRs) for all HOPDs that were open. These reports have been uploaded to HOPDs that have QualityNet account holders. CMS conducted a confidential national reporting of measure results (dry run).
2. CMS took public comments on the measure during development and during rulemaking, and answered questions through a question and answer (Q&A) email inbox before and during the dry run, as well as a Q & A session as part of the National Provider Call.
3. In response to the measure-specific feedback received, the developer made several changes to the measure. This included updates to exclusion for surgeries that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care; exclusion for surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit; and exclusion for surgeries billed on the same claim as an ED visit, where the measure continues to exclude surgeries billed on the same hospital outpatient claim as an ED visit unless the primary diagnosis on the facility claim is indicative of a complication of care.

Additional Feedback:

***Questions for the Committee**:*

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

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

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

No improvement in performance was seen between the 2017 dry run results and the 2018 performance data. The developer notes that the measure specifications changed between these two performance periods and that the measure was first publicly reported in January 2020. It expects improvement to increase now that the results are being reported.

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

Unexpected findings (positive or negative) during implementation

- The developer states no unintended consequences or findings have been identified.

Potential harms

- The developer does not report any potential harms.

Additional Feedback:

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

- Yes, according to the document
- Hospital compare is a great resource. There was ample Q and A sessions. They allowed for a dry run. There was ample opportunity for feedback to occur and it appears they responded to it. They made the appropriate responses to the feedback.
- Used in a publicly reported accountability program. Public comments and feedback used.
- Trial run was performed and sent to hospitals - first performance data published January 2020
- n/a
- public reporting already occurs
- Open-Ended Response
- No concerns
- Meets

4b. Usability

- If the measure was limited to inpatient readmissions, there may be a higher threshold for readmitting patients if they presented to the ED. However, this is addressed by including observation and ED visits. It is theoretically possible that a hospital could pressure surgeons to counsel patients to wait until they can be seen in clinic for postop concerns and not go to the ED - this could potentially delay management of serious postop events.
- I am curious if race/ethnicity could be used as a means of risk-stratifying. It is an obvious social determinant of health. It was mentioned in their rationale but was never furthered investigated. The other unintended consequence that may be of concern is the fact that poorly performing hospitals may be taking care of patients from a lower SES with more co-morbidities. Even though in their multi-variant analysis they did not find this had a major impact, as the number of facilities contribute data will this become more of an issue.
- Clear opportunity to improve care. No potential harms or unintended consequences.
- This measure encourages optimizing condition of patients prior to same-day surgery as well as their discharge planning and instructions. I see little harm, if any, from the measure
- Will they assume ASCs are better to go to than HOPDs or that they should only pursue Inpatient care, despite increases in costs? What are some of the unforseen consequences. Concerned that if this measure failed to pass face validity, the benefit it will have towards consumers.
- already in use
- Open-Ended Response
- Although the logic of the measure is sound, no evidence was provided that overall or individual quality has changed since the measure was implemented. Public reporting started in Jan of 2020, so perhaps improvements will be realized. It would be interesting to examine sites that improve to learn how they did it.
- Meets

Criterion 5: Related and Competing Measures

Related or competing measures

0697 : Risk Adjusted Case Mix Adjusted Elderly Surgery Outcomes Measure

1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

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

3366 : Hospital Visits after Urology Ambulatory Surgical Center Procedures

3470 : Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures

3490 : Admission and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

Harmonization

The developer states that the measure specifications are harmonized with related measures. The HOPD Surgery measure is a claims-based measure, therefore any differences in measure specifications create no burden to facilities as the measures are calculated from data produced during the billing process. The related measures target the same quality domains as the HOPD Surgery measure. The patient cohort is also similar in that the related measures target Medicare Fee-For-Service (FFS) patients aged 65 years and older. For those measures that focus on the same facility (HOPD) as the target, the procedure/clinical cohorts however, differ. The ASC-related measures have a different target (ASCs instead of HOPDs), and are divided into separate measures for general surgery, orthopedic surgery, and urologic surgery. The HWR measure and HOPD Surgery include overlapping but distinct surgeries (inpatient vs. outpatient) and overlapping but distinct patient outcomes (hospital visits within 7 days vs. readmissions within 30 days). They address a similar patient cohort (Medicare FFS patients 65 years of age and older). NQF 0687 overlaps with the HOPD Surgery measure in terms of target (patients over 65) and has an overlapping outcome. However, NQF 0687 includes all surgeries (in- and out-patient) and is not limited to outpatient surgeries. In addition, the outcomes that are part of NQF 0687 include complications that may result in an ED visit, observation stay, or inpatient admission (such as sepsis, surgical site infection, wound disruption, and urinary tract infection). It also includes mortality as an outcome, which is not included in the HOPD Surgery measure.

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

- Based on the measure document, other measures are similar but do not incorporate either the breadth of procedures or scope of revisit categories (ED, observation or inpatient readmission)
- It appears that this measure is harmonized with other similar measures.
- Multiple related measures. Seems to harmonize, different in that specific to outpatient surgery.
- My only question is why there are multiple measures for ASCs separating urology, ortho and general surgery procedures whereas there is only a single measure for Hospital Outpatient Surgery.
- n/a
- none
- Open-Ended Response
- Summarized in the measure worksheet
- Unsure

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June 19, 2020

- Comment by: **Federation of American Hospitals**

The Federation of American Hospitals (FAH) appreciates the opportunity to comment on this measure prior to the Standing Committee's evaluation. The FAH remains concerned with the risk adjustment approach to determine whether inclusion of social risk factors. The FAH believes that the risk adjustment approach should not consider the identification and testing of social risk factors as supplementary to clinical risk factors. This approach was identified as a concern by the NQF Disparities Standing Committee and developers must begin to include these factors within the testing of the model rather than the approach of "adding on" factors after the model is developed. This type of analysis would assist facilities and others in understanding how their inclusion could impact the model and provide additional information for groups examining this issue such as the NQF and Office of the Assistant Secretary for Planning and Evaluation. As a result, the FAH believes that this measure lacks sufficient information on the potential impact these social risk variables have on the risk adjustment model.

- **No NQF Members have submitted support/non-support choices as of this date.**

Combined Methods Panel Scientific Acceptability Evaluation

Measure Number: 2687

Measure Title: Hospital Visits After Hospital Outpatient Surgery

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**

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

Panel Member #2: s.3.2 (p5) several improvements to algorithm that improve accuracy and face validity as judged by experts: coding; same-day codes; certain procedures. Not clear to me how or whether urgent care visits within 7 days are counted.

***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.**

Panel Member #1: I found the measure specifications very clear and comprehensive, including the description of specification updates to inclusion and exclusion criteria.

Panel Member #3: No concerns.

Panel Member #5: I am unclear about some aspects of this. The provided materials are extensive and include a description of a National Provider Call QnA. Despite this, further explanation would be helpful. For example

→Does every overnight care episode with AM discharge count in the numerator?

→How does the algorithm determination of planned/unplanned fit in actual circumstances?

→Does the existing CMS list prevent a surgeon from listing every procedure as inpatient (for admission) and then switching to same day discharge after the fact to avoid registering a numerator/denominator event?

→The documents do not specify why folks are admitted. "Pain" is a very common reason for seeking care after surgery (including POD0, especially for obese patient with chronic pain). Is there additional documentation about unintended consequences of such a measure, particularly in the midst an opioid epidemic?

Panel Member #7: Adequate, although complex

Panel Member #8: The applicant indicates that a statistical risk model was used with 21 factors**.** However, from reviewing the application, I am unsure which factors were included in the final model, so this is difficult to evaluate. It would have been nice to see a table that had all factors included. This does make me concerned that the instructions are not clear enough for facilities to follow. Thus, I chose "no" above.

Panel Member #9: No concerns

RELIABILITY: TESTING

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

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

☐ **Yes** ☒ **No**

Panel Member #4: NA – score level testing conducted

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: The STN analyses was appropriately used based on the Adams tutorial for score level reliability.

   Panel Member #2: Signal-to-noise (Adams formula). Between/between+within

   Panel Member #3: Conducted signal-to-noise analysis

   Panel Member #4: Use of signal to noise testing is acceptable. "…signal-to-noise method…" [p8]

   Panel Member #7: Adequate

   Panel Member #8: A signal to noise ratio was calculated. This is an appropriate method.

   Panel Member #9: The hierarchical logistic regression model is appropriately used, but only for facilities with at least 30 procedures.

   Panel Member #10: Used SNR. Median SNR was 0.84. Did not provide information on reliability testing for low volume facilities. This is desirable but is not required.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: Reliability of measure score level was supported for all facilities (median=0.76) and for facilities that had 30 cases or more (median=0.839) which included 99% of all procedures.

   Panel Member #2: Rel = 0.76 for all facilities; 0.84 for those with >30 procedures

   Panel Member #3: Median signal-to-noise reliability score is high for all facilities (median reliability of 0.759) and all facilities with at least 30 procedures (median reliability of 0.839).

   Panel Member #4: Signal to noise test results are generally good. Only concern is results for all facilities regardless of denominator size the lower end of the IQR is borderline poor. This suggest to set a minimum threshold to approximately 30 cases. Results based on facilities with a minimum threshold of 30 cases was high: median signal to noise of 0.839.

   "…median facility-level reliability (signal-to-noise reliability) for all facilities (N=3974) was 0.759 (IQR 0.372-0.892); the median facility-level reliability for facilities with more than 30 procedures (n=2979) was 0.839 (IQR 0.696-0.915)." [p9]

   Panel Member #8: From the application: "The median facility-level reliability (signal-to-noise reliability) for all facilities (N=3974) was 0.759 (IQR 0.372-0.892); the median facility-level reliability for facilities with more than 30 procedures (n=2979) was 0.839 (IQR 0.696-0.915). The 2979 facilities represent 1,161,312 procedures or 99% of the total 1,172,087 procedures."

   Panel Member #9: For facilities with more than 30 procedures, the reliability is quite high. Although the application suggests that CMS will use the 30 procedure cutoff when this measure is used for public reporting, there is no assurance that CMS will not lower the cutoff to, say, 10, when the reliability of this measure would likely be questionable.

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)

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

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

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

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

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Panel Member #1: Methods and interpretation are appropriate, and reliability of measure scores were high, therefor the 'high' rating.

    Panel Member #3: Used appropriate method for testing score-level reliability; the median signal-to-noise reliability scores were sufficiently high.

    Panel Member #4: Excerpt from Q7 response (above): Results based on facilities with a minimum threshold of 30 cases was high: median signal to noise of 0.839.

    Panel Member #5: Signal to Noise, including (0.76 overall, 0.84 when lower volume sites excluded). Moderate rather than High because I am uncertain whether reliability estimate is artefactually high due to inadequate risk adjustment.

    Panel Member #7: Data for lower volume facilities vs. higher volume facilities (with expected differences in reliability) are not provided

    Panel Member #8: No concerns

    Panel Member #9: The reliability testing, based on hospitals with at least 30 procedures, are not sufficient to assess the reliability of the measure for hospitals with fewer procedures, but statistics tells us that measures for small hospitals will have less reliability. This problem can be solved by only using the measure for hospitals with at least 30 procedures.

    Panel Member #10: Used SNR. Median SNR was 0.84. Did not provide information on reliability testing for low volume facilities. This is desirable but is not required.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1: No concerns

Panel Member #2: The only exclusion worth worrying about is the one regarding surgery and observation day on the same claim, because it is not possible to determine sequence or attribution (about 5% of cohort). It would be useful to see what would be the impact on measure score if these are included with assumption that they are all in the numerator.

Panel Member #3: No concerns. Exclusion criteria exclude about 7% of surgeries, with the largest chunk being surgeries billed on the same claim as the observation stay (difficult to understand order of events)

Panel Member #4: No concerns. The MIF does a good job of providing a rationale for each exclusion. The rationale in each instance is logical.

Panel Member #7: None

Panel Member #8: Exclusions include those who are not: continuously enrolled for 7 days after the procedure; surgeries on the same day as an ED visit/or on the same claim (etc.) unless there is a diagnosis indicative of a complication of care; and surgeries billed on the same claim as an observation stay. These exclusions were determined via clinical review and are important because temporality cannot be determined from claims. I have no concerns with these exclusions.

Panel Member #9: No concerns

Panel Member #10: none

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

Submission document: Testing attachment, section 2b4.

Panel Member #1: No concerns.

Panel Member #2: Odds ratio of predicted (observed?) to expected hospital visits. Bootstrapping to determine 95% CI. Mean OR=1.28, showing room for improvement. Only 4% better than expected and 4% worse than expected. May not be much room for improvement, unless you go with full range of ORs across all facilities, which is reasonably wide (p41)

Panel Member #3: No concerns. See a distribution of results for facilities. 6% of facilities are statistically better or worse than average.

Panel Member #4: No concerns.

Panel Member #7: There are relatively small differences between facilities in this measure in the interquartile range.

Panel Member #8: From the application "The range of performance on the HOPD Surgery measure (RSHVR min-max of 0.54-2.39) demonstrates that there is a significant quality gap. Specifically, the best-performing HOPD (RSHVR of 0.54) is performing 46% better than average, whereas the worst-performing HOPD (RSHVR of 2.39) is performing 139% worse than the average. Furthermore, our outlier analysis identified about 300 or about 8 percent of HOPDs as outliers (3.77% significantly better and 3.98% significantly worse than expected). Note that the that average performer refers to an HOPD with the same case and service-line mix, performing at the average." I have no concerns with this conclusion.

Panel Member #9: No concerns

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.

Panel Member #1: NA

Panel Member #3: **Not applicable**

Panel Member #4: No concerns / Not applicable

Panel Member #7: N/A
Panel Member #8: Not applicable
Panel Member #9: No concerns

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: The missing data section notes NA. I am assuming there is no or limited missing data. A clarification note would be appreciated.

Panel Member #3:

Panel Member #4: No concerns / Not applicable

Panel Member #7: None

Panel Member #8: The authors indicate that missing data are not an issue. However, I am not sure I believe this. I feel that missing data (and incorrect data) are always an issue in claims data and that this should have been assessed.

Panel Member #9: No concerns

16. **Risk Adjustment**

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

Panel Member #10: Fit a hierarchical generalized linear model (HGLM) to calculate hospital PE ratio and used bootstrapping to calculate 95% CI. Risk factors consist of 25 patient-level variables (age, comorbidities, and indicators of surgical complexity [RVU, body system operated on])

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

Panel Member #1: I was very impressed by the extent of testing done and considerations regarding the possibility to include risk-adjustment by social risk-factors. The final decision not to adjust for social-risk factors is strongly supported.
Panel Member #2: Developer makes the case that the social risk factors analyzed have no significant or even detectable effect on the performance score.

Panel Member #8: From the application "The social factors included: 1. Medicare-Medicaid dual-eligibility status; and 2. Agency for Healthcare Research and Quality Socioeconomic Status (AHRQ SES) Index"

Panel Member #10: Did not include social risk in risk adj model because they demonstrated:

- High correlation (r=0.99) between risk-standardized outcomes with and without inclusion of SES (Medicaid eligibility) as a risk factor
- Distribution of risk-standardized scores were similar in hospitals with high and low proportion of low SES patients

16d.**Risk adjustment summary:**

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

Panel Member #2: 21 factors included

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

Panel Member #4: NA – risk factors present at onset of care

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

Panel Member #4: c-stat: 0.68

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

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

Panel Member #1: This measure has a strong risk-adjustment model, well developed, tested and validated.

Panel Member #3: R/A model uses 21 factors; decided to leave out social risk factors after looking at the impact on the model; good model discrimination and calibration.

Panel Member #4: Development of a risk adjustment methodology is appropriate for the given measure. The consideration of SES was deliberate and the conclusion is logical for the non-inclusion of SES. The c-stat is a bit lower than one would like to see, but acceptable.

Panel Member #5: Description of with and without social factors, C stat unchanged (still poor with c of less than 0.7, albeit in a potentially discounted set of procedures). The risk adjustment is usual ICD10 comorbidities and does not reflect some variables that may be important. Addressing these is outside the scope of this evaluation.

Panel Member #7: Despite evidence for inclusion of the AHRQ SES Index, developers did not use this variable in final risk adjustment model.

Panel Member #8: The applicant indicates that a statistical risk model was used with 21 factors**.** However, from reviewing the application, I am unsure which factors were included in the final model, so this is difficult to evaluate. It would have been nice to see a table that had all factors included. However, the methods used to examine the impact on social risk factors on the measure are appropriate (distribution of social risk factors across measured entities, patient level observed hospital visit rates for patients with social risk factors, strength and significance of the social risk factors in the model [9OR] comparing the c-statistic for risk adjustment models with and without social risk factors, comparison of measure scores between facilities with the highest and lowest proportion of patients with social risk factors, and the assessment of the relationship between the

RSHVR and percent of patients with social risk factors in facilities in the highest quartile for proportion of patients with social risk factors).

Panel Member #9: The clinical risk adjustment model is well developed and tested. A careful analysis shows why the benefits of adjusting for the two available social risk failures do not outweigh the concerns.

Panel Member #10: Fit a hierarchical generalized linear model (HGLM) to calculate hospital PE ratio and used bootstrapping to calculate 95% CI. Risk factors consist of 25 patient-level variables (age, comorbidities, and indicators of surgical complexity [RVU, body system operated on])

Model discrimination in validation data set is acceptable C stat (0.684). Model calibration (-0.05, 0.96) is excellent. Calibration plot is excellent. Predictive validity is high.

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

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

    ☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member #1: Methods are appropriate. However, given the challenge to identify an external marker that assesses a similar construct, methods could have included additional analyses as suggested below.

    I appreciate the challenge in finding an external measure of a similar domain to serve as a comparison measure for empirical validity testing, which is well explained. A similar post-surgery measure for hospital-wide readmission in inpatient services was selected as the best candidate for comparison. Due to many differences between these two measures, the hypothesis was for a weak positive correlation.

    Since this was the hypothesis, a strong support for the score level validity would not be expected.

    Other forms of validity testing as known groups validity could have been used to better assess the measure score validity. For example, hospitals with a case mix of patients with higher rates of characteristic expected to be associated with better outcomes would be expected to have higher scores compared to hospitals with higher rates of case-mix of patients with characteristic expected to be associated with worse outcomes. Such evidence would provide additional support for the measure score validity.

Additionally, a 1 item survey about face validity of the measure was conducted among 13 participating team members.

Panel Member #2: Compared to hospital-wide readmission rates (HWR) with chi-square tests of differences in HWR classification by HOPD Surgery measure. Face validity via TEP review and input

Panel Member #3: For empirical validity testing, hypothesized a weak, positive relationship with the surgical domain of the HWR measure. Conducted face validity with a multi-stakeholder TEP, asking if this measure would identify higher quality facilities from lower quality facilities.

Panel Member #4: Disagree with the premise that the measure being evaluated would be necessarily correlated with the HWR measure. Rationale is twofold:

[1] These are differing units of analysis.

[2] The measure steward states "It is possible the same surgeons and surgical teams are performing surgeries covered by both measures". However, no analysis / evidence of this is noted as to the degree to which this occurs. Additionally, it is not only the surgeon that influences the outcomes, but numerous other factors as well (e.g. the team, systems in place in each setting).

Empirical validity:

"We examined whether better performance on the HOPD Surgery measure was correlated with better performance on measures that are related… we identified readmission measures, specifically the …HWR measure, as a potential candidate for comparison." [p10]

Face validity:

"During measure development, we asked our TEP, made up of 15 members including patient representatives, expert clinicians, methodologist, researchers, and providers, to formally assess the measure's face validity…." [p12]

Panel Member #8: Empirical validity of the measure score was done by correlating the measure score to comparator measures (convergent validity). Face validity was assessed by the technical expert panel. These are appropriate measures.

Panel Member #9: Methods for assessing face validity were appropriate. The absence of a measure with the proposed measure should be correlated limits the relevancy of the result cited.

Panel Member #7: Adequate

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: As the developers expected, the external measure used for comparison had a weak positive and non-significant association with measure 2687 (r=0.033, p=0.07). An additional analyses of distribution of positive or negative outliers from measure 2687 by quartiles of the external measure provided some evidence for an expected association. Overall, I find these results to only weakly support the measure's validity.

The systematic assessment of face validity provided additional evidence for validity.

Overall, these analyses support the empirical validity on 0716 score level but only to a weak level. Since face validity alone does not meet NQF requirements for re-endorsement at measure maintenance, the validity assessment relies mostly on the empirical validity testing, which to my

view, as mentioned above, provide weak evidence that is at best moderate to support the score level validity of measure 2687.

Panel Member #2: Marginally acceptable

Panel Member #3: • The correlation coefficient indicated a very weak correlation (0.033); in looking at outliers, there is a trend that better performance on the HOPD Surgery measure is consistent with better performance on the HWR measure.

• The face validity results showed good agreement that the measure can identify high quality facilities from low quality facilities.

Panel Member #4: Empirical validity:

As noted on my response to Q21: Disagree with the premise that the measure being evaluated would be necessarily correlated with the HWR measure. Rationale is twofold:

[1] These are differing units of analysis.

[2] The measure steward states "It is possible the same surgeons and surgical teams are performing surgeries covered by both measures". However, no analysis / evidence of this is noted as to the degree to which this occurs. Additionally, it is not only the surgeon that influences the outcomes, but numerous other factors as well (e.g. the team, systems in place in each setting).

In addition to the issues I noted in my response to Q21, the "correlation" of 0.03 is not really telling us anything about a relationship between the results of these two measures.

Face validity:

Given this is a measure in maintenance, face validity is insufficient to maintain endorsement. While the results from one question cited in the TEP survey is positive, we do not have any of the other survey question results. Perhaps other questions were more (or as) germane to identifying the perceptions of the TEP. It would have been more helpful if the measure steward would have provided us with other survey results. The absence of information here in concerning.

Empirical validity:

"…The correlation coefficient indicates a very weak positive correlation (0.033, p=0.07) as expected.

… There are more "better than expected" HOPD Surgery outliers in the first (better performing) quartile of HWR performance, and more "worse" HOPD Surgery outliers in the fourth (worst performing) quartile of HWR performance. A chi square test indicated that this relationship was significantly different than what would be expected by chance alone (p=0.0331)." [p15]

Face validity:

"The results of the TEP rating of agreement with the validity statement: … Mean rating=5.2

All TEP members who responded to the survey indicated they agreed with the statement that "The risk-standardized hospital visit ratios obtained from the outpatient surgery measure as specified can be used to distinguish between better and worse quality facilities." 12 of the 13 indicated they moderately or strongly agreed. Two TEP members did not respond to the TEP survey.

Frequency of Ratings of Agreement

Rating # (%) of Responses

1 (Strongly disagree) 0 (0)

2 (Moderately disagree) 0 (0)

3 (Somewhat disagree) 0 (0)

4 (Somewhat agree) 1 (7.7)

5 (Moderately agree) 8 (61.5)

6 (Strongly agree) 4 (30.8)" [p16]

Panel Member #7: Associations between HOPD and HWR measure a week.

Panel Member #8: The measure developer states, "Taken together, these results support the validity of the HOPD Surgery measure." And I agree that evidence supports this.

Panel Member #9: The evidence for validity is primarily based on face validity testing.

Panel Member #10: Examined the correlation of this measure with related measures.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☒ **No**

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

Panel Member #4: As noted on my response to Q21: Disagree with the premise that the measure being evaluated would be necessarily correlated with the HWR measure. Rationale is twofold:

[1] These are differing units of analysis.

[2] The measure steward states "It is possible the same surgeons and surgical teams are performing surgeries covered by both measures". However, no analysis / evidence of this is noted as to the degree to which this occurs. Additionally, it is not only the surgeon that influences the outcomes, but numerous other factors as well (e.g. the team, systems in place in each setting).

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

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

☒ **Yes**

☐ **No**

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

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

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

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

☒ **Low** (NOTE: Should rate LOW if you believe that there <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

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

Panel Member #1: As mentioned above, empirical validity of score level was only partially supported. Face validity was strong, but not sufficient in lieu of NQF requirements at maintenance. However, since I do not believe that there is a good reason to assume this measure in not valid, the challenge seems more as being able to provide the evidence to support its validity. Since this is an active and important measure, and the level of scoring for its validity is somewhat subjective, I struggled between rating validity as being moderate or insufficient.

I rated validity as moderate to allow the measure to move forward. However, insufficient evidence for its score level validity is my message within these comments, urging developers to provide additional evidence for validity, possibly considering the analyses suggested above.

My recommendation to NQF is to emphasize this point before the next maintenance cycle.

Other than the above, I found no additional threats to validity.

Panel Member #2: This measure may be reaching sundown. Not a great number of negating outliers and the empirical validity marginally acceptable

Panel Member #3: The correlation coefficient was not much different than 0.00, making a very weak case. The outlier analysis supports a limited trend of facilities with better performance on the HOPD Surgery measure have better performance on the HWR measure. The previously conducted face validity results support this measure is useful for QI purposes.

Panel Member #4: As noted on my response to Q23:

Empirical validity:

As noted on my response to Q21: Disagree with the premise that the measure being evaluated would be necessarily correlated with the HWR measure. Rationale is twofold:

[1] These are differing units of analysis.

[2] The measure steward states "It is possible the same surgeons and surgical teams are performing surgeries covered by both measures". However, no analysis / evidence of this is noted as to the degree to which this occurs. Additionally, it is not only the surgeon that influences the outcomes, but numerous other factors as well (e.g. the team, systems in place in each setting).

In addition to the issues I noted in my response to Q21, the "correlation" of 0.03 is not really telling us anything about a relationship between the results of these two measures.

Face validity:

Given this is a measure in maintenance, face validity is insufficient to maintain endorsement. While the results from one question cited in the TEP survey is positive, we do not have any of the other survey question results. Perhaps other questions were more (or as) germane to identifying the perceptions of the TEP. It would have been more helpful if the measure steward would have provided us with other survey results. The absence of information here in concerning.

Panel Member #5: Expert panel plus prediction model – fine. I am curious about ethnography here, however, the QnA description would indicate this was not terribly revealing.

Panel Member #7: There are weak associations of the hospitalization after outpatient surgery and validation variables.

Panel Member #9: The evidence for validity is primarily based on face validity testing.

Panel Member #10: Fit a hierarchical generalized linear model (HGLM) to calculate hospital PE ratio and used bootstrapping to calculate 95% CI. Risk factors consist of 25 patient-level variables (age, comorbidities, and indicators of surgical complexity [RVU, body system operated on]) Model discrimination in validation data set is acceptable C stat (0.684). Model calibration (-0.05, 0.96) is excellent. Calibration plot is excellent. Predictive validity is high. This is the most important method to assess overall validity – as opposed to examining the correlation of a particular measure with other related measures

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Panel Member #1: I'd like to discuss with the SMP the level of support to the measure score validity given the evidence provided and my comments above.

Developer Submission

Additional evaluations and submission materials attachments…

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

xxxxxxxxxx.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** NQF 2687

Measure Title: Hospital Visits After Hospital Outpatient Surgery

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 4/15/2020

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

Outcome

☒ Outcome: The outcome is all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the surgery or 2) an unplanned hospital visit (emergency department [ED] visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the surgical procedure.

☐ Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

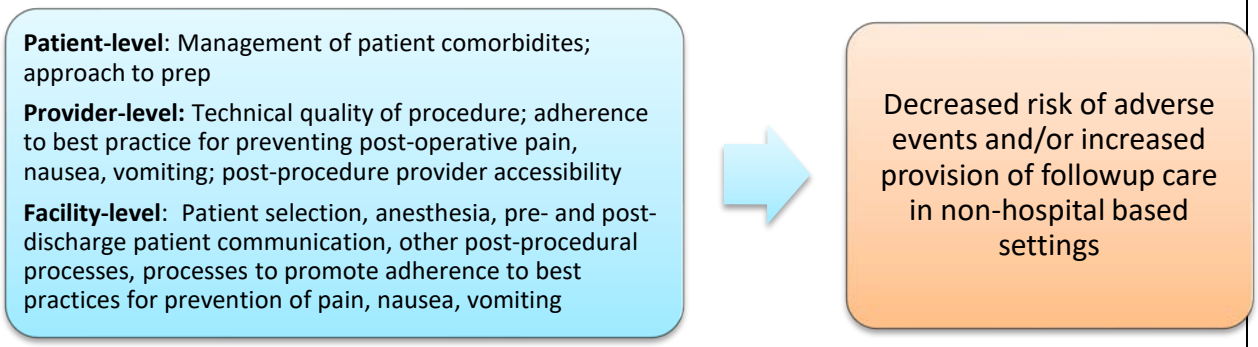
☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The outcome is all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the surgery or 2) an unplanned hospital visit (ED visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the surgical procedure.

Rationale: Unplanned hospital visits following same-day surgeries often reflect surgery-related adverse events and quality issues. Several strategies and interventions may reduce unplanned hospital visits after same-day surgery. They include: 1) appropriate patient selection for same-day surgery; 2) appropriate patient education on preparation prior to same-day surgery; 3) improving the technical quality of the outpatient surgery, including the choice of procedural technique and anesthesia; 4) appropriate implementation of interventions to manage common causes of hospital visits such as protocols to manage nausea and vomiting and postoperative pain; and 5) educating patients about potential adverse events post same-day surgery, symptoms to monitor, whom to contact with questions, and where and when to seek follow-up care.



1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Not applicable.

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

least one healthcare structure, process, intervention, or service.

[Note to NQF: Almost all of the evidence in this section has been updated, therefore we have not marked it all in red text. The new citations that were used since the measure was endorsed are marked in red font in the citations list.]

The outcome of unplanned hospital visits following outpatient same-day surgery is a widely accepted measure of outpatient surgical care quality. This measure provides the opportunity to improve quality of care and to lower rates of adverse events leading to hospital visits after outpatient surgery.

Estimates of hospital visit rates within the first 30 days following surgery vary from less than one percent to 28% depending on the type of surgery, the outcome measured (inpatient admissions alone or with ED visits, and observation stays), outcome timeframe (e.g., 7, 14, or 30 days), and patient characteristics (e.g. age, sex) (Christian, 2019; Mull, 2019, De Oliveira, 2015; Liu et al., 2018; Rosero et al., 2017, DeFroda, 2017, Gengler et al., 2017, Liu et al., 2018-2). For example, a 2018 retrospective study of patients undergoing outpatient shoulder arthroscopy found an inpatient admission rate within 7 days of 0.22% (Liu et al, 2018a). In contrast, a 2018 study of veterans age 65 or older found a 28% rate of hospital admissions (in-patient, emergency department, and observation stays) within 7 days for patients who had urological surgery, and a 6% rate of hospital admissions for patients who had orthopedic surgery (Mull et al., 2018).

Common causes of return visits following outpatient surgery include surgical errors, post-operative pain, infection, nausea, and vomiting (Rosero et al., 2017, Gildaseo et al., 2015, Liu et al., 2018a, Liu et al., 2018b). In one 2017 study of patients undergoing outpatient laparoscopic cholecystectomy, 60% of hospital return visits were due to these preventable events (Rosero et al., 2017). Other less common, but more serious, reasons for return hospital visits include bleeding, respiratory complications, deep vein thrombosis, cardiac complications, and urinary complications (Rosero et al, 2017; Gildasio, et. Al., 2015; DeOliveria, 2015; Liu et al., 2018a; Liu et al., 2018b; Rosero et al., 2017). Patient characteristics, such as age, sex, and comorbidities such as diabetes, can increase the risk of an admission (De Oliveria et al., 2015; DeFroda et al., 2017; Gengler et al., 2017; Christian et al., 2019). In addition, clinical procedural factors can increase the risk, such as the type of anesthesia used, and longer operation time (Defroda et al., 2017; Liu et al., 2018a; Gengler et al., 2017; Mingus et al., 1997; Christian et al., 2019).

Interventions to improve same-day outpatient surgical procedural quality can reduce unplanned hospital visits following outpatient surgery. Potential quality improvement actions include appropriate patient selection, improving surgical techniques, implementing protocols to address common problems such as adequate control of nausea and vomiting and postoperative pain, patient education about potential adverse effects of the surgery, reconciling patient medications, and organizing appropriate follow-up care with providers such as primary care physicians. For example, guidelines recommend multi-modal approaches for treatment of post-operative pain (Chou et al., 2016; Joshi et al., Mariano, et al, 2020) as well as routine multi-modal nausea and vomiting prophylaxis for all patients (Gan et al., 2014). Facilities can also provide support for identifying and managing patient-level risk factors; for example, identifying patients with diabetes can ensure optimal care during the perioperative period regarding prevention of hyperglycemia (Thompson et al., 2016).

A hospital visit following same-day surgery is an unexpected and potentially preventable outcome for patients scheduled for same-day surgeries that have a low anticipated risk. Providers (HOPDs and surgeons) are often unaware of their patients' hospital visits after surgery because patients often present to the ED or to different hospitals, leading to understated adverse event rates and suggesting the need for better measurement to drive quality improvement (Mezei G, 1999). Therefore, both patients and providers benefit from outcome measures of hospital visits – a broad, patient-centered outcome that reflects the full range of reasons leading to hospitalization among patients undergoing same-day surgery.

The HOPD Surgery measures is part of the Hospital Outpatient Quality Reporting (HOQR) Program, a pay-for-reporting program. HOPDs first saw their facility-specific measure scores in 2017, during a "dry run" that precedes public reporting. The measure was first publicly reported in January 2020, on Hospital Compare. Currently, there are no other publicly available quality reports of HOPDs that perform same-day surgery. Thus, this measure addresses an important quality measurement area and enhances the information available to patients choosing among HOPDs that provide same-day outpatient surgery. Furthermore, providing outcome rates to HOPDs makes visible to clinicians and hospitals meaningful quality differences and incentivizes improvement.

Citations

Christian RA, Gibbs DB, Nicolay RW, Selley RS, Saltzman MD. Risk factors for admission after shoulder arthroscopy. J Shoulder Elbow Surg. 2019 May;28(5):882-887.

Chou et al., Guidelines on the Management of Postoperative Pain. The Journal of Pain, Vol 17, No 2 (February), 2016: pp 131-157.

DeFroda SF, Bokshan SL, Owens BD. Risk Factors for Hospital Admission Following Arthroscopic Bankart Repair. Orthopedics. 2017 Sep 1;40(5):e855-e861.

De Oliveira GS Jr, Holl JL, Lindquist LA, Hackett NJ, Kim JY, McCarthy RJ. Older Adults and Unanticipated Hospital Admission within 30 Days of Ambulatory Surgery: An Analysis of 53,667 Ambulatory Surgical Procedures. J Am Geriatr Soc. 2015 Aug;63(8):1679-85.

Falck-Ytter Y, Francis CW, Johanson NA, Curley C, Dahl OE, Schulman S, Ortel TL, Pauker SG, Colwell CW Jr. Prevention of VTE in orthopedic surgery patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. American College of Chest Physicians Chest. 2012;141(2 Suppl):e278S.

Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. Anesth Analg. 2014 Jan;118(1):85-113.

Gengler I, Carpentier L, Pasquesoone X, Chevalier D, Mortuaire G. Predictors of unanticipated admission within 30 days of outpatient sinonasal surgery. Rhinology. 2017 Sep 1;55(3):274-280.

Joshi GP, Schug SA, Kehlet H. Procedure-specific pain management and outcome strategies. Best Pract Res Clin Anaesthesiol. 2014 Jun;28(2):191-201.

Liu J, Flynn DN, Liu WM, Fleisher LA, Elkassabany NM. Hospital-Based Acute Care Within 7 Days of Discharge After Outpatient Arthroscopic Shoulder Surgery. Anesth Analg. 2018a Feb;126(2):600-605.

Liu J, Kim DH, Maalouf DB, Beathe JC, Allen AA, Memtsoudis SG. Thirty-Day Acute Health Care Resource Utilization Following Outpatient Anterior Cruciate Ligament Surgery. Reg Anesth Pain Med. 2018b Nov;43(8):849-853.

Marino, E. Management of acute perioperative pain. UpToDate, January 2020. Accessed on March 6, 2020. https://www.uptodate.com/contents/management-of-acute-perioperative-pain

Mezei G, Chung F. Return hospital visits and hospital readmissions after ambulatory surgery. Annals of surgery. Nov 1999;230(5):721-727.

Mingus ML, Bodian CA, Bradford CN, Eisenkraft JB. Prolonged surgery increases the likelihood of admission of scheduled ambulatory surgery patients. *Journal of clinical anesthesia.* Sep 1997;9(6):446-450.

Rosero EB, Joshi GP. Hospital readmission after ambulatory laparoscopic cholecystectomy: incidence and predictors. J Surg Res. 2017 Nov;219:108-115.

Sanabria D, Rodriguez J, Pecci P, Ardila E, Pareja R. Same-Day Discharge in Minimally Invasive Surgery Performed by Gynecologic Oncologists: A Review of Patient Selection. J Minim Invasive Gynecol. 2019.

Thompson BM, Stearns JD, Apsey HA, Schlinkert RT, Cook CB. Perioperative Management of Patients with Diabetes and Hyperglycemia Undergoing Elective Surgery. Curr Diab Rep. 2016 Jan;16(1):2.

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

Not applicable. This is an outcome measure.

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

Not applicable. This is an outcome measure.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Not applicable. This is an outcome measure.**
Quote the guideline or recommendation verbatim about the process, structure or	Not applicable. This is an outcome measure.

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	Not applicable. This is an outcome measure.
Provide all other grades and definitions from the evidence grading system	Not applicable. This is an outcome measure.
Grade assigned to the **recommendation** with definition of the grade	Not applicable. This is an outcome measure.
Provide all other grades and definitions from the recommendation grading system	Not applicable. This is an outcome measure.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Not applicable. This is an outcome measure.
Estimates of benefit and consistency across studies	Not applicable. This is an outcome measure.
What harms were identified?	Not applicable. This is an outcome measure.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Not applicable. This is an outcome measure.

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.

Not applicable. This is an outcome measure.

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

Not applicable. This is an outcome measure.

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

Not applicable. This is an outcome measu

1b. Performance Gap

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

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

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

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

The goal of this measure is to reduce adverse patient outcomes associated with preparation for same-day surgery, the surgery itself, and follow-up care, by capturing and making more visible to providers and patients unplanned hospital visits following outpatient surgery. The measure score provides an assessment of quality that is publicly reported, and also informs 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.*

The distribution of performance scores (risk-standardized hospital visit ratios or RSHVRs) for facilities for the performance period of January 1, 2018-December 31, 2018 is shown below.

N=3874 facilities

Mean RSHVR (standard deviation) = 1.01 (0.15)

Minimum=0.54

25th percentile= 0.93

50th percentile=0.99

75th percentile=1.07

Maximum=2.39

Performance for the January 1, 2018-December 2018 performance period, by deciles, is shown below.

Decile//# Facilities//Minimum RSHVR//Maximum RSHVR

1//397//0.54//0.84

2//397//0.84//0.90

3//398//0.90//0.95

4//397/0.95//0.97

5//398//0.97//0.99

6//397//0.99//1.02

7//398//1.02//1.05

8//397//1.05//1.10

9//398//1.10//1.19

10//397//1.19//2.39

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.

See performance information provided above 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.*

A detailed analysis related to social risk factors distribution and relationship to the outcome and measure scores is presented in the testing attachment, section 2b3.4b.

Below we present measure scores (risk-standardized hospital visit ratios or RSHVRs) for facilities for the performance period of January 1, 2018-December 31, 2018 for the following disparities variables:

Dual eligible (DE): Measure scores stratified by facilities' proportion of patients with the DE variable

Range of DE variable values within quartiles:

Q1:<=3.06%

Q2:>3.06% and <=5.45%

Q3:>5.45% and <=9.15%

Q4:>9.15%

Characteristic//Q1//Q2//Q3//Q4

Number of HOPDs//745//743//746//745

Number of patients//324240//362810//286577//187685

Maximum RSHVR//2.39//1.77//1.96//2.02

90th//1.20//1.21//1.24//1.22

75th//1.09//1.10//1.10//1.12

Median//0.98//0.99//0.99//1.01

25th//0.88//0.88//0.90//0.92

10th//0.80//0.80//0.82//0.86

Minimum RSHVR//0.62//0.59//0.58//0.54

AHRQ Low SES: Measure scores stratified by facilities' proportion of patients with low AHRQ SES.

Characteristic//Q1//Q2//Q3//Q4

Number of HOPDs//744//744//746//745

Number of patients//309758//359295//307331//184928

Maximum RSHVR//2.04//2.28//2.39//1.86

90th//1.20//1.23//1.25//1.21

75th//1.10//1.10//1.11//1.10

Median//0.99//0.98//0.99//1.00

25th//0.90//0.89//0.89//0.91

10th//0.81//0.80//0.82//0.85

Minimum RSHVR//0.54//0.59//0.59//0.58

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

See performance information provided above in 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):*

Surgery, Surgery : General Surgery, Surgery : Perioperative and Anesthesia

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

Care Coordination, Safety, Safety : Complications

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

Elderly, Populations at Risk

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.)*

Methodology landing page (includes most recently published specifications as well as previous versions of methodology): https://www.qualitynet.org/outpatient/measures/surgery/methodology

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:** HOPD_Surgery_Measure_Data_Dictionary_v2019a.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.

Yes

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.

2019 Measure Updates

A public report that contains these updates and additional details is available at this URL: https://www.qualitynet.org/outpatient/measures/surgery/methodology

•Update to exclusion for surgeries that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.

Rationale: In these situations, it is not possible to use claims data to determine whether the surgery was the cause of, subsequent to, or during the ED visit. However, if the ED visit is coded with a diagnosis for a complication, the assumption is that it occurred after the surgery.

Prior to this update, surgeries billed on the same claim as an ED visit were excluded from the measure, unless the claim had a diagnosis indicating a complication of care occurred. This update further refines this exclusion to exclude surgeries that occur on the same day and on the same claim as the index surgery, unless there is a diagnosis of complication of care indicated on the claim. Additionally, we expand the exclusion criteria to exclude surgeries that are billed on the same hospital outpatient claim, but occur after the ED visit, regardless of whether complications of care are billed or not.

•Update to exclusion for surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit.

Rationale: In these situations, we assume that the surgery was subsequent to the ED visit and may not represent a routine surgery. Timing of the ED visits is determined using revenue center dates from the outpatient claim.

•Update the surgery measure's planned readmission algorithm by adopting changes made when going from v4.0_2019 to v4.0_2020 of the planned readmission algorithm.

Rationale: These changes improve the accuracy of the algorithm by updating the algorithm for coding changes and decreasing the number of hospital visits that the algorithm mistakenly designated as unplanned or planned.

For this update, we studied the 2019 versions of the AHRQ CCS map for diagnoses and procedures, respectively, to determine how the newly implemented ICD-10 codes in the 2018 code set were categorized, and to examine any code shifts that may have occurred from the previous version of the AHRQ CCS map to the most recent AHRQ CCS map. Review of these versions of the AHRQ CCS map was extensive, and included:

•Examination of seven AHRQ CCS diagnosis categories and 13 AHRQ CCS procedure categories to determine how the newly implemented ICD-10 codes should be incorporated into the planned readmission algorithm specifications; and,

•Examination of one AHRQ CCS diagnosis category and eight AHRQ CCS procedure categories codes that shifted to investigate where code shifts may affect the specialty cohort definitions and planned readmission algorithm.

We then solicited input from clinical and measure experts to confirm the clinical appropriateness of the AHRQ CCS categorization of the newly implemented ICD-10 codes and any changes warranted due to the code shifts that occurred. The experts also reviewed the newly implemented ICD-10 codes in the FY 2019 version of the ICD-10-CM/PCS codes to determine which, if any, should be added to the singular ICD-10 code lists that are also used in the algorithm (conditions that are not captured by AHRQ CCS categories). The intent was to maintain the clinical integrity of the algorithm.

These processes led to the following changes to the algorithm:

•Potentially planned procedures:

o The addition of four AHRQ CCS procedure categories (Procedure CCS 96, 118, 162, 163), which consisted of procedures that clinicians deemed potentially planned. Examples of these categories are “Other OR lower GI therapeutic procedures” (CCS 96) and “Other OR therapeutic procedures on joints” (CCS 162). We previously included subsets of ICD-10-PCS codes within CCS 96, 118, and 163 on the potentially planned procedures list.

o The addition of selected ICD-10-PCS codes within CCS group 112 (“Other OR therapeutic procedures of urinary tract”).

o The removal of CCS 95 (“Other non-OR lower GI therapeutic procedures”) and 174 (“Other non-OR therapeutic procedures on skin subcutaneous tissue fascia and breast”) as a whole; we previously included a subset of codes on the potentially planned procedures list.

The complete set of codes reflected in the v4.0_2020 planned readmission algorithm adopted as the PAA for the surgery measure are available in the data dictionary tabs “PAA PA1 Always Plnnd Px,” “PAA PA2 Always Plnnd Dx,” “PAA PA3 Pot Plnnd Px,” and “PAA PA4 Acute Dx.”

2018 Measure Updates

Please see this report for details:

https://www.qualitynet.org/files/5d0d367e764be766b01006a7?filename=HOPD_Surg_MsrUpdtRpt_2018.pdf

•Update to the approach for identifying concurrent high-risk procedure to (1) not require specific GSI values for procedures on CMS’s Hospital Outpatient Prospective Payment System “inpatient only”

procedures list, (2) exclude cases with a high-risk procedure identified on the outpatient or inpatient facility claim that are matched to a professional services claim having an eligible same-day surgery

Rationale: This improves the measure's ability to exclude all cases with concurrent high-risk surgery.

•Expansion of definition of complication of care for same-day, separate-claim ED visit exclusion. Additional CCS complication codes added following dry run:

*CCS 2616: Adverse effects of medical care

*CCS 2617: Adverse effects of medical drugs

Rationale: This improves accuracy of capturing the outcome by including same-day ED visits that are indicative of a complication of care.

•Update to the exclusion for surgeries billed on the same claim as an ED visit, where the measure continues to exclude surgeries billed on the same hospital outpatient claim as an ED visit unless the primary diagnosis on the facility claim is indicative of a complication of care

Rationale: This improves accuracy of capturing the outcome by including same-day, same-claim ED visits indicative of a complication of care.

•Update the surgery measure's planned admission algorithm by adopting changes made when going from v 4.0_2017 to v4.0_201910 of the planned readmission algorithm.

Rationale: These changes improve the accuracy of the algorithm by updating the algorithm for coding changes and decreasing the number of hospital visits that the algorithm mistakenly designated as unplanned or planned.

No changes were made in 2017.

2016 Measure Updates

Please see this report for details:

https://www.qualitynet.org/files/5d0d3a7e764be766b0104644?filename=2016HOPDSurgeryTechReport.pdf

1. Update to the exclusion criterion for Medicare FFS Enrollment

Rationale: The measure excludes surgeries for patients who are not continuously enrolled in Medicare FFS Parts A and B for at least 7 days after the surgery (rather than at least 30 days, as specified in the original measure). The measure will continue to exclude patients with fewer than 7 days post-surgery enrollment to ensure all patients have full data available for outcome assessment. This minor adjustment shortens the requirement for continuous enrollment in order to exclude index procedures only when necessary.

2. Addition of an exclusion criterion to exclude surgeries that are billed on the same day but on a separate claim as an ED visit, unless the ED visit has a diagnosis indicative of a complication of care

Rationale: It is unclear whether a same-day ED visit occurred before or after an eligible same-day surgery. However, the measure excludes surgeries with same-day, separate- claim ED visits unless the diagnosis for the ED visit is indicative of a post-surgery complication. The measure classifies these diagnoses using AHQR CCS groups. The measure considers ED visits with the following diagnoses as outcomes:

- AHRQ CCS 237 – Complication of device; implant or graft

- AHRQ CCS 238 – Complications of surgical procedures or medical care
- AHRQ CCS 257 – Other aftercare
- ICD-9-CM code 338.18 – Acute pain

In these scenarios, the procedure is counted in the index cohort and the ED visit is counted as an outcome.

3. Update to how the measure handles multiple qualifying procedures within 7 days. Rationale: The timeframe for outcome assessment was 7 days after each procedure that occurred within a 7-day period. With the updated specifications, the outcome is attributed to the surgery nearest to (and preceding) the hospital visit.

4. Adoption of the changes from the updated planned readmission algorithm Version 4.0_2017 from Version 3.0, which are based on findings from a validation study and the review of those findings by clinical experts, to the surgery measure's planned admission algorithm;

Rationale: These changes improve the accuracy of the algorithm by decreasing the number of hospital visits that the algorithm mistakenly designated as unplanned or planned.

5. Specification of the risk variables and complication-of-care variables mapped to the Hierarchical Condition Categories (HCC) Version 22 to accommodate ICD-10 codes. Rationale: This update accommodates the use of ICD-10 codes for risk variable definitions using version 22 of CMS's HCCs.

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 is all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the surgery or 2) an unplanned hospital visit (emergency department [ED] visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the surgical procedure.

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 all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the surgery or 2) an unplanned hospital visit (ED visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the surgical procedure. If more than one unplanned hospital visit occurs in the 7 days following the surgical procedure, only the first hospital visit within the outcome timeframe is counted in the outcome. If there are two surgical procedures within a 7-day period, we adjust the follow up period of the first procedure to be the time between the first procedure and the second procedure. The second procedure's follow up period remains 7 days post-procedure. Thus, hospital visit outcomes are assigned to the first procedure if they occur during the time

between procedures, while outcomes in the 7 days following the second procedure are assigned to the second procedure.

Planned Admission Algorithm

For inpatient admissions occurring after Day 1 following surgery, we only include unplanned admissions in the measure outcome. We consider admissions occurring on the day of the surgery (Day 0) and Day 1 post-surgery "unplanned" as the vast majority of these admissions are inpatient admissions directly following surgery. "Planned" admissions are those planned by providers for anticipated medical treatment or procedures that must be provided in the inpatient setting. We do not count these in the outcome because variation in planned admissions does not reflect quality differences.

To identify admissions as planned or unplanned we use an algorithm we previously developed for CMS's hospital readmission measures, CMS Planned Readmission Algorithm (PRA) Version 4.0. In brief, the algorithm uses the procedure codes and principal discharge diagnosis code on each hospital claim to identify admissions that are typically planned and may occur after a surgery. A few specific, limited types of care are always considered planned (for example, major organ transplant, rehabilitation, or maintenance chemotherapy). Otherwise, a planned admission is defined as a non-acute admission for a scheduled procedure (for example, total hip replacement or cholecystectomy). Post-discharge admissions for an acute illness or for complications of care are never considered planned.

Also, the measure never considers ED visits or observation stays as planned. The most recently published methodology report provides a detailed description of the planned admission algorithm adapted for the surgery measure: https://www.qualitynet.org/files/5d0d367e764be766b01006a7?filename=HOPD_Surg_MsrUpdtRpt_2018.pdf. The codes that define ED visits and observation stays are in the attached data dictionary, sheet "HOPD_Surgy__ED_Obs_Stay_Def"

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

Outpatient same-day surgeries performed at HOPDs for Medicare FFS patients aged 65 years and older.

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).

The surgery measure was developed to improve the quality of care delivered to patients undergoing hospital outpatient surgeries. In brief, the surgery measure includes all hospital outpatient departments (HOPDs) that performed qualifying surgeries during the performance period.

Further information on the measure development process is available in the Hospital Visits After Hospital Outpatient Surgeries: Measure Technical Report (2014) and 2016 Technical Report Addendum: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Hospital-Visits-after-Hospital-Outpatient-Surgery-Measure.pdf

Inclusion Criteria

1. Surgeries and procedures that are substantial and are typically performed as same-day surgeries

Rationale: The target cohort is low-to-moderate-risk surgeries that can be safely performed as same-day surgeries and do not typically require an overnight stay or an inpatient admission. In addition, they do not occur in conjunction with a same-day emergency department (ED) visit or observation stay. We define same-day surgeries using the CMS's list of covered ambulatory surgery center (ASC) procedures. The list is comprised of procedures for which the patients are expected to return home the same day as their procedure. We further restrict Medicare's list of covered ASC procedures using the Global Surgical Package (GSP) indicator and include two types of procedures from this list:

- Substantive surgeries performed at HOPDs (except eye surgeries)

Rationale: Ambulatory procedures include a heterogeneous mix of non-surgical procedures, minor surgeries, and more substantive surgeries. We want to include substantive surgeries but not very low-risk (minor) surgeries or non-surgical procedures, which typically have a high volume and a very low outcome rate. We define substantive procedures using the Medicare Physician Fee Schedule (MPFS) global surgery indicator (GSI) code 090.

- Cystoscopy procedures with intervention

Rationale: All endoscopy procedures are considered non-surgical procedures based on Medicare coding (GSI code 000). However, we include cystoscopy with intervention because it is a common procedure, often performed for therapeutic intervention by surgical teams, and the outcome rate and causes of hospital visits post-procedure are similar to those for surgeries in the measure cohort.

Please refer to the data dictionary "HOPD_Surg_Cohort" to review the list of qualifying same-day surgeries, including cystoscopy procedures with intervention. The data dictionary "HOPD_Surg_Eye_Exclusions" provides the list of eye surgeries that are excluded from the measure cohort.

2. Surgeries on patients aged 65 or over

Rationale: Medicare beneficiaries under age 65, typically, are a highly diverse group with a higher burden of disability, and it is therefore difficult to adequately risk adjust for the under 65 population.

3. When multiple procedures occur concurrently, only surgeries that are not performed concurrently with a high-risk procedure are included.

Rationale: Occasionally, more than one surgery may be performed and some of these surgeries may be higher-risk procedures. When multiple procedures occur, we only include surgeries that are not performed concurrently with high-risk procedures. Please refer to the data dictionary "HOPD_Surg_High_Risk_Exclusions" tab to review the list of high-risk procedures. High-risk procedures are identified using the Hospital Outpatient PPS Addendum B. A procedure is considered high-risk if it is flagged as "Inpatient Only" (not paid under OPPS) or "Outpatient Only" (paid under OPPS, but not on the list of ASC-approved procedures). Removal of these procedures aids with alignment of the measure's restriction to only include ASC-covered procedures.

4. Surgeries for patients with continuous enrollment in Medicare Fee-for-Service (FFS) Parts A and B in the 12 months prior to the surgery.

Rationale: Patients with full enrollment have all claims available for identifying comorbidities for risk adjustment.

5. Surgeries for patients with continuous enrollment in Medicare Fee-for-Service (FFS) Parts A and B in the 12 months prior to the surgery.

Rationale: Patients with full enrollment have all claims available for identifying comorbidities for risk adjustment.

Citations

1. Centers for Medicare & Medicaid Services (CMS). Three Day Payment Window. 2013; http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Three_Day_Payment_Window.html

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

1. Surgeries for patients without continuous enrollment in Medicare FFS Parts A and B in the 7 days after the surgery.

2. Surgeries for patients who have an ED visit on the same day but billed on a separate claim, unless the ED visit has a diagnosis indicative of a complication of care.

3. Surgeries that are billed on the same hospital claim as an emergency department (ED) visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.

4. Surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit.

5. Surgeries that are billed on the same outpatient claim as an observation stay.

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.)*

Exclusion Criteria

1. Surgeries for patients without continuous enrollment in Medicare FFS Parts A and B in the 7 days after the surgery.

Rationale: We exclude these patients to ensure all patients have full data available for outcome assessment.

2. Surgeries for patients who have an ED visit on the same day but billed on a separate claim, unless the ED visit has a diagnosis indicative of a complication of care.

Rationale: It is unclear whether a same-day ED visit occurred before or after an eligible same-day surgery. However, the measure will not exclude surgeries with same-day, separate-claim ED visits if the diagnoses are indicative of a complication of care because we want to continue to capture these outcomes. The ICD-9-CM and ICD-10-CM codes that define complications of care are in the attached Data Dictionary, sheet " HOPD_Surg_ED_Excl_CoC".

3. Surgeries that are billed on the same hospital claim as an emergency department (ED) visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.

Rationale: In these situations, it is not possible to use claims data to determine whether the surgery was the cause of, subsequent to, or during the ED visit. However, if the ED visit is coded with a diagnosis for a complication, the assumption is that it occurred after the surgery. The ICD-9-CM and ICD-10-CM codes that define complications of care are in the attached Data Dictionary, sheet " HOPD_Surg_ED_Excl_CoC".

4. Surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit.

Rationale: In these situations, we assume that the surgery was subsequent to the ED visit and may not represent a routine surgery. Timing of the ED visits is determined using revenue center dates from the outpatient claim.

5. Surgeries that are billed on the same outpatient claim as an observation stay.

Rationale: We do not include these cases in the calculation because the sequence of events is not clear.

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

Not applicable. This is not a stratified measure.

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

Statistical risk model

If other:

S.12. Type of score:

Ratio

If other:

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

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

1. Identify surgeries meeting the inclusion criteria described above in S.7.

2. Exclude procedures meeting any of the exclusion criteria described above in S.8/S.9.

3. Identify a binary flag for an unplanned hospital visit within 7 days of index procedures as described above in S.5.

4. Use patients' historical and index procedure claims data to create risk-adjustment variables.

5. Fit a hierarchical generalized linear model (HGLM) and calculate the ratio of the number of "predicted" hospital visits to the number of "expected" hospital visits for each facility, given its case/procedure mix using the results. This is the risk-standardized hospital visit ratio (RSHVR). The HGLM is adjusted for age, clinical risk factors, and procedure RVU and body system that vary across patient populations, are unrelated to quality, and influence the outcome. Details about the risk-adjustment model can be found in the measure methodology report: Hospital Visits after Hospital Outpatient Surgery Measure Technical Report at https://www.qualitynet.org/files/5d0d3a7e764be766b0104644?filename=2016HOPDSurgeryTechReport.pdf

6. Use statistical bootstrapping to construct a 95% confidence interval estimate for each facility's RSHVR. For more information about the measure methodology, please see the Hospital Visits after Hospital Outpatient Surgery Measure Technical Report posted on the web page provided in data field S.1.

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

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

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

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable. This measure is not based on survey or patient-reported data.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims, Enrollment Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Medicare administrative claims and enrollment data

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

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

Facility

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

Outpatient Services

If other:

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

Not applicable

2. Validity – See attached Measure Testing Submission Form

HOPD_Surgery_Testing_Attachment_FINAL_010520.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.

Yes

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.

Yes

2.3 For maintenance of endorsement

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

Yes - Updated information is included

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

Measure Number (*if previously endorsed*)**:** NQF 2687
Measure Title: Hospital Visits after Hospital Outpatient Surgery
Date of Submission: 1/6/2020

Type of Measure:

1.

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

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: Medicare enrollment data	☒ other: Medicare enrollment data, Master Beneficiary Summary File (MBSF) Database, Census Data/American Community Survey

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*).

For the original development of the Hospital Visits after Hospital Outpatient Surgery (HOPD Surgery) measure we used 2009-2011 Medicare data to develop a Medicare fee-for-service (FFS) cohort consisting of a 20% sample of same-day surgery claims from hospital outpatient departments (HOPDs) as outlined below. The measure cohort included patients with outpatient same-day surgery in 2010, and we used inpatient and outpatient data from 2009 to derive comorbidities for risk adjustment for these patients.

a. Datasets used to define the cohort:

-Carrier (Part B Physician) claims Standard Analytical File (SAF): This SAF contains a 20% sample of all base and line item claims billed by physicians performing surgeries at HOPDs.

-Medicare 100% Hospital Outpatient SAF: This dataset contains 100% of all HOPD facility claims for surgeries performed at HOPDs. This dataset links physician claims for surgeries performed at HOPDs to the corresponding HOPD facility claim in order to obtain a facility identifier for HOPDs.

-Enrollment database and denominator files: This dataset contains Medicare FFS enrollment, demographic, and death information for Medicare beneficiaries.

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

-The Centers for Medicare & Medicaid Services (CMS) Medicare Provider Analysis and Review File (MedPAR) Part A Inpatient institutional claims (100% of all claims): This dataset is used to identify inpatient hospital claims.

-Medicare 100% Hospital Outpatient SAF: This dataset is used to identify emergency department (ED) and observation stay visits.

c. Datasets used to identify comorbidities for risk adjustment:

-Inpatient and outpatient claims (institutional and non-institutional carrier) data from the year prior to the outpatient surgery (2009) were used to identify comorbidities for risk adjustment for these patients.

For updated measure testing provided in this submission we used paid, final action Medicare claims from January 1, 2018 to December 31, 2018 to identify procedures performed in the outpatient setting at Hospital Outpatient Departments (HOPDs), and subsequent hospital visits. In addition, we used CMS enrollment and demographic data from the Health Account Joint Information (HAJI) database to determine inclusion and exclusion criteria. Patient history is assessed using claims data collected in the 12 months prior to the outpatient surgery.

For all derived cohorts:

a. Datasets used to define the cohort:

-All cohort, outpatient surgeries performed at HOPDs were identified using the full set of Medicare beneficiaries' claims from the Carrier non-institutional claims, which included physician bills for hospital outpatient services. HOPD claims were linked to the outpatient institutional surgical claims or inpatient institutional surgical claim when CMS's 3-day window payment period applied.

-Enrollment database and denominator files: These datasets contain Medicare Fee-For-Service (FFS) enrollment, demographic, and death information for Medicare beneficiaries, which is used to determine inclusion/exclusion criteria.

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

-The outcomes of emergency department (ED) visits and observation stays after outpatient surgery were identified from hospital outpatient institutional claims, and inpatient hospital admissions (at acute care and critical access hospitals) from 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 to the outpatient surgery were used to identify comorbidities for risk adjustment for these patients.

To assess social risk factors, we used census as well as claims data (dual eligible status obtained through the Master Beneficiary Summary File (MBSF) Database; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score obtained through census data). The dataset used varies by testing type; see section 1.7 for details.

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

We used data from January 1, 2009 through December 31, 2011 for initial measure development and data from January 1, 2017 to December 31, 2018 for the updated testing results presented in this submission; these dates include one year of inpatient and outpatient claims to identify comorbidities for risk adjustment.

Please see section 1.7 for additional details.

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 measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The number of measured entities (HOPDs) varies by testing type; see section 1.7 for details.

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

The number of patients varies 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.

Dataset	Description of Dataset	Use and Section in the Testing Attachment
Dataset #1: Initial Development Dataset **Dataset #1a: Development dataset**	Administrative claims dataset including Part B Physician claims (20% sample) linked to HOPD facility claims to identify HOPD facilities; Medicare FFS enrollment database and denominator files. CMS MedPAR Part A institutional claims (100%) and Medicare Hospital Outpatient SAF (100%) were used to identify the outcome. Patient history is assessed using inpatient and outpatient claims data collected in the 12 months prior to the outpatient surgery. Outpatient surgeries are identified using Medicare's list of covered ASC procedures.	• Section 2b1 Validity testing (face validity) • Section 2b3.3a Identification and selection of risk-adjustment variables

Dataset #1b: Validation dataset	Dates of data for the outcome: January 1, 2010-December 31, 2010. Number of procedures: 212,104 Number of facilities: 4234 For measure development and testing, we randomly split the 2010 data into Development (Dataset #1a) and Validation Samples (Dataset #1b) (each sample including approximately 50% of outpatient surgeries contained in the 2010 data). For patients in these samples, we used data from 2009 to derive comorbidities for risk adjustment.	• Section 2b3.7 Risk model calibration statistics
Dataset #2: Endorsement Maintenance Testing Dataset	Final action 2018 Medicare claims (100%) from the Health Account Joint Information (HAJI) database were used identify procedures performed in the outpatient setting at Hospital Outpatient Departments (HOPDs), and subsequent hospital visits. In addition, we used CMS enrollment and demographic data from the HAJI database to determine inclusion and exclusion criteria. Patient history is assessed using inpatient and outpatient claims data collected in the 12 months prior to the outpatient surgery. Outpatient surgeries are identified using Medicare's list of covered ASC procedures. Dates of data for the outcome: January 1, 2018-December 31, 2018. Number of procedures: 1,172,087 Number of facilities: 3974 Number of facilities with >= 30 procedures: 2979 Mean age (SD): 74.659 (6.729) % Female: 49	• Section 2a2 Reliability • Section 2b1 Data Element & Measure Score Validity • Section 2b2 Testing of Measure Exclusion • Section 2b3.4b Selection of Social Risk Factors • Section 2b4 Meaningful Differences • Section 2b3.6 Predictive ability Section 2b3.6 Statistical model discrimination statistics

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.

We developed and used the conceptual framework described in section 2b3.3a below to identify potential social risk factors. Limited social risk factor data are available at this time, however, on Medicare beneficiaries [1]. We analyzed two well-studied social risk factors that could best be operationalized in data:

1. Medicare-Medicaid dual-eligibility status

Dual-eligibility for Medicare and Medicaid is available at the patient level in the Medicare Master Beneficiary Summary File. The eligibility threshold for over 65-year-old Medicare patients considers both income and assets. For the dual-eligible (DE) indicator, there is a body of literature demonstrating differential health care and health outcomes among beneficiaries, indicating that, while not ideal, the DE indicator allow us to examine some of the pathways of interest [1].

2. Agency for Healthcare Research and Quality Socioeconomic Status (AHRQ SES) Index

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 [2]. It is a widely used index that summarizes area-level measures of employment, income, education, and housing from the American Community Survey (ACS). Each of the index components is available at the census block level, which we then used to link to patient's residence using 9-digit ZIP code. The AHRQ SES index score summarizes 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.

The AHRQ SES Index's 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. In this submission, we present analyses using the census block group level, the most granular level possible using ACS data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the census block group level. 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 (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes.

We identify patients at risk due to social factors if they are in the bottom 25th percent of the ARHQ SES distribution.

Citations

1. 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 December 8, 2019.

2. Bonito A, Bann C, Eicheldinger C, Carpenter L. 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 provide facility-level measure score reliability using the signal-to-noise method, using the formula presented by Adams and colleagues [1,2]. Specifically, for each facility we calculate the reliability as:

Reliability=(σ_(facility-to-facility)^2)/(σ_(facility-to-facility)^2+ (σ_(facility error variance)^2)/n)

Where facility-to-facility variance is estimated from the hierarchical logistic regression model, n is equal to each facility's observed case size, and the facility error variance is estimated using the variance of the logistic distribution (pi^2/3). The facility-level reliability testing is limited to facilities with at least 30 admissions for public reporting.

Signal-to-noise reliability scores can range from 0 to 1. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real difference in performance.

We calculated the measure score reliability for all facilities, and for facilities with a volume cutoff of 30 procedures, using Dataset #2. Our rationale for this is described below.

Relationship of reliability testing to minimum volume per facility

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. CMS has empirically determined that measure scores (risk-standardized hospital visit ratios or RSHVRs) for HOPDs with 30 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 and in public reporting CMS

typically reports scores for facilities with fewer procedures than the volume cutoff 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.

Citations

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

2. 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

The median facility-level reliability (signal-to-noise reliability) for all facilities (N=3974) was **0.759** (IQR 0.372-0.892); the median facility-level reliability for facilities with more than 30 procedures (n=2979) was **0.839** (IQR 0.696-0.915). The 2979 facilities represent 1,161,312 procedures or 99% of the total 1,172,087 procedures.

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?*)

The median signal-to-noise reliability score is sufficiently high for both all facilities, and facilities with at least 30 procedures (the public reporting cutoff).

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

Citation:

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

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Empirical Validity Testing of the Measure Score

We examined whether better performance on the HOPD Surgery measure was correlated with better performance on measures that are related, meaning that at least to some extent the comparator measures assess the same domain of quality (complications requiring acute care after same-day surgery).

Hospital Outpatient Quality Reporting Measures

To identify related measures, we reviewed all of the measures that are currently publicly reported (for CY2020 Payment Determination) in the Hospital Outpatient Quality Reporting program (HOQR) and the Inpatient Quality Reporting Program (IQR). Note that, because Hospital Outpatient Departments are not a distinct entity but rather a diverse group of care settings (such as the ED, outpatient clinics, and outpatient surgery settings), many of the HOQR measures are not relevant comparators because they are restricted to particular settings (such as the ED or clinic) that do not overlap with the HOPD Surgery measure.

Of the 14 measures in the HOQR program that are not planned for retirement, none of the measures assessed the same quality domain. One measure, OP-32: Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy, assessed the same outcome. However, colonoscopy is a narrow and relatively low-risk procedure performed in a different setting (not the surgical suite); we therefore would not expect the measure scores from the colonoscopy measure to correlate with measure socres from the HOPD Surgery measure.

Hospital Inpatient Quality Reporting Measures

Of the Hospital Inpatient Quality Reporting measures, we identified readmission measures, specifically the Hospital-wide Readmission (HWR) measure, as a potential candidate for comparison. The HWR calculates rates of 30-day unplanned hospital readmissions for five different specialty cohorts: medicine, neurology, cardiovascular, cardiorespiratory, and surgery/gynecology), each with a fully developed and statistically tested risk model. (Methodology report is available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-

Instruments/HospitalQualityInits/Downloads/Hospital-Wide-All-Cause-Readmission.zip). The HOPD Surgery measure cohort and outcome overlap with the surgery/gynecology cohort (hereafter "surgery cohort") of HWR.

We hypothesized that the HOPD Surgery measure score would show a weak, positive relationship with the measure score for the surgery cohort of the HWR measure given that the measures assess overlapping but distinct surgeries (outpatient vs. inpatient) and overlapping but distinct patient outcomes (hospital visits within 7 days vs. readmissions within 30 days):

We expect some correlation because:

•It is possible that the same surgeons and surgical teams are performing surgeries covered by both measures, and in some hospitals those procedures may be co-located.

•Both measures count admissions to the hospital post-surgery in the outcome, although the HOPD measure also counts ED visits, which make up the majority of the return visits, as well as observation stays.

• The same organizational culture and processes may be in place to prevent visits to the hospital following surgery across both inpatient and outpatient procedures, such as timely recognition of post-operative complications and ensuring effective discharge plans [1].

However, we do not expect moderate or strong correlations because:

•The outcomes differ; not only does the HOPD Surgery measure include ED visits and observation stays in addition to admissions, but the period of observation for the outcome differs (7 days for the HOPD Surgery measure vs. 30 days for the HWR surgery cohort).

•The cohorts (procedures and patients) are distinct; inpatient procedures are generally more complex procedures done on higher-risk patients.

Instead, we hypothesize that the relationship, while positive, would be weak, because:

•Certain procedures, such as inguinal hernia repair, are more likely to be done on an outpatient vs. inpatient basis, whereas more complex procedures, such those within the CCS "Vascular stents and OR procedures, other than head or neck" are predominantly inpatient [2]. Further, the HWR surgery cohort includes more acutely ill patients.

•The two measure scores are the result of separate statistical models that assume a distribution of latent quality that is normally distributed. These estimates are shrunk toward an overall mean that depends on a hospital's own performance as well as the other hospitals in the measure. Each measure's score will ultimately have their own uncertainty associated with the estimate which will ultimately reduce correlation among the measure scores.

For this analysis we used the measure scores from the HOPD Surgery measure calculated from Dataset #2 (January 1, 2018-December 31, 2018) and evaluated their association with measure scores from the same facilities using the HWR surgery cohort measure score (July 1, 2017-June 30, 2018). Specifically. we examined the relationship of performance on the HOPD Surgery measure score against performance within quartiles for the HWR surgery cohort measure score (see Figure 1, section 2b1.3). We also calculated the Pearson correlation coefficient between the two measure scores, to characterize the strength and direction of the relationship. Finally, we examined the association of outlier status of the HOPD Surgery measure score with the quartiles of the HWR surgery cohort score (see Figure 2, section 2b1.3). Specifically, we identified outliers by estimating an interval estimate (similar to a confidence interval) around each hospital's measure score and identified those facilities that had a 95% interval

estimate entirely above or entirely below 1.0, as described in section 2b4.1. We then performed a chi square test to determine if the outlier relationship (between HOPD Surgery measure score outliers and quartiles of HWR performance) was significantly different than what would be expected by chance alone.

Face Validity as Determined by the TEP

During measure development, we asked our TEP, made up of 15 members including patient representatives, expert clinicians, methodologist, researchers, and providers, to formally assess the measure's face validity. We provided the TEP background on the NQF measure evaluation criteria and presented the measure specifications and testing and performance results for their evaluation.

List of TEP Members

1) David Chang, PhD, MPH, MBA—Massachusetts General Hospital (Associate Professor of Surgery, Department of Surgery; Director of Healthcare Research and Policy Development, Codman Center for Clinical Effectiveness in Surgery); Boston, MA
2) Gary Culbertson, MD—Iris Surgery Center (Plastic Surgeon; Medical Director); Sumter, SC
3) Martha Deed, PhD—Member of the public; North Tonawanda, NY
4) Richard Dutton, MD, MBA—Anesthesia Quality Institute (Executive Director); Park Ridge, IL
5) Nestor Esnaola, MD, MPH, MBA—Temple University School of Medicine (Professor of Surgery; Chief, Surgical Oncology); Philadelphia, PA
6) Charles Goldfarb, MD—Washington University School of Medicine (Associate Professor of Orthopaedic Surgery); St Louis, MO
7) Lisa Ishii, MD, MHS—Johns Hopkins School of Medicine (Associate Professor, Department of Otolaryngology-Head & Neck Surgery); Baltimore, MD
8) Sandra Koch, MD—Carson Medical Group (OB/GYN surgery); Carson City, NV
9) Tricia Meyer, PharmD, MS—Scott & White Memorial Hospital (Associate Vice-President, Department of Pharmacy); Texas A&M University College of Medicine (Associate Professor, Department of Anesthesiology); Texas A&M Rangel College of Pharmacy (Adjunct Associate Professor, Department of Anesthesiology); Temple, TX
10) Linda Radach, BA— Member of the public; Lake Forest Park, WA
11) Danny Robinette, MD—Surgery Center of Fairbanks (General Surgeon; Medical Director); Fairbanks, AK
12) Suketu Sanghvi, MD—The Permanente Medical Group, Kaiser Permanente (Ophthalmologist; Associate Executive Director); Oakland, CA
13) Christopher Tessier, MD—Manchester Urology Associates (Urologist); Manchester, NH
14) Thomas Tsai, MD, MPH—Brigham and Women's Hospital (General Surgery Resident; Administrative Chief Resident for Research); Harvard School of Public Health (Postdoctoral Fellow, Department of Health Policy and Management); Boston, MA
15) Katherine Wilson, RN, MHA—AmSurg Corp (Vice President, Quality); Nashville, TN

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The risk-standardized hospital visit ratios obtained from the outpatient surgery measure as specified can be used to distinguish between better and worse quality facilities."

TEP members indicated their agreement with the face validity of the measure on a six-point scale:

1=Strongly disagree
2=Moderately disagree
3=Somewhat disagree
4=Somewhat agree
5=Moderately agree
6=Strongly agree

Use of Established Measure Development Guidelines:
We developed this measure in consultation with national guidelines for publicly reported outcome measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcome measurement set forth in NQF guidance for outcome measures, CMS MMS guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" [3,4].

Citations

1. Brooke BS, De Martino RR, Girotti M, Dimick JB, Goodney PP. Developing strategies for predicting and preventing readmissions in vascular surgery. J Vasc Surg. 2012;56(2):556–562.

2. Steiner CA, Karaca Z, Moore BJ, Imshaug MC, Pickens G. Surgeries in Hospital-Based Ambulatory Surgery and Hospital Inpatient Settings, 2014: Statistical Brief #223. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006-2017 May. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb223-Ambulatory-Inpatient-Surgeries-2014.jsp. Accessed November 6, 2019.

3. 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.

4. 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 January 6, 2019.

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

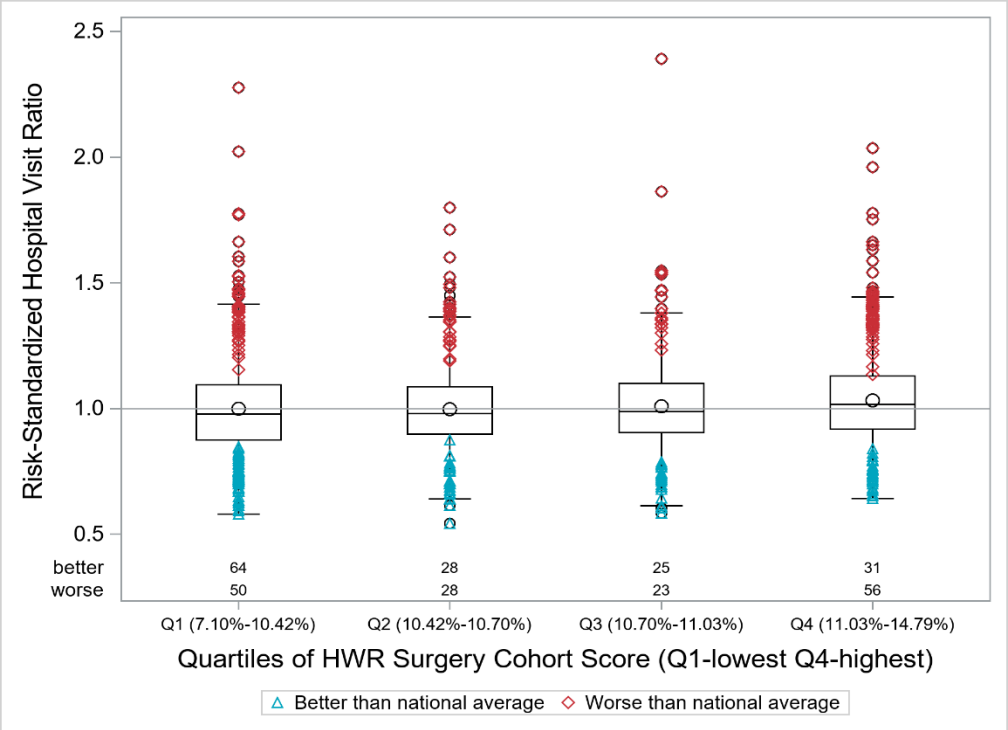
Empiric Validity Testing

To examine the external validity of the HOPD Surgery measure, we divided hospitals into quartiles based on their scores on the comparator measure, the surgery cohort of the HWR measure (range of scores 7.10%-14.79%). We then displayed the distribution of those hospitals' HOPD Surgery measure scores (RSHVRs) within each of the HWR quartiles in a box plot or "whisker" plot (Figure 1). (Note: The horizontal line within a box represents the median HOPD RSHVR of all the hospitals in the quartile, the open circle represents the mean, the horizontal boundaries of a box represent the 1st and 3rd quartiles).

We also compared outliers on the HOPD Surgery measure within quartiles of HWR performance (Figure 1 and Figure 2). In Figure 1 we display hospitals that are statistical outliers on the HOPD Surgery measure with a blue triangle (if they are better than expected) or a red diamond (if they are worse than expected) (outliers are identified as described in section 2b4.1, below). In Figure 2 we show the total number of "better than expected" and "worse than expected" facilities within each quartile of performance on the HWR measure (surgery cohort).

All analyses included facilities with at least 30 procedures.

Figure 1: Relationship between HOPD Surgery measure score and HWR (surgery cohort) measure score



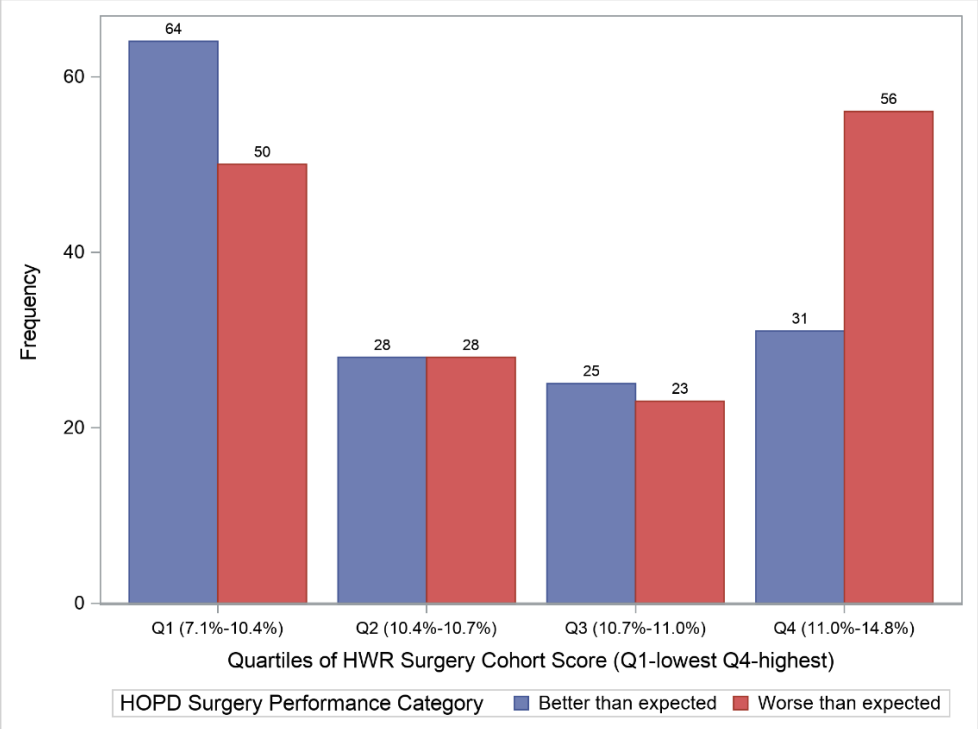
The results show a trend toward better performance on the HOPD Surgery measure with better performance on the comparator measure (HWR, surgery cohort). As shown in Figure 1, better performance on the HOPD Surgery measure shows a small positive trend with better performance across quartiles of performance on the HWR measure. The correlation coefficient indicates a very weak positive correlation (0.033, p=0.07) as expected.

The outlier (better, and worse, than expected) comparison is consistent with the trend toward better performance on the HOPD Surgery measure with better performance HWR measure (Figure 2). There are more "better than expected" HOPD Surgery outliers in the first (better performing) quartile of HWR performance, and more "worse" HOPD Surgery outliers in the fourth (worst performing) quartile of HWR performance. A chi square test indicated that this relationship was significantly different than what would be expected by chance alone (p=0.0331).

More specifically,

- There are 64 HOPD Surgery "better than expected" outliers (blue bar in Figure 2) in the first or best quartile (Q1) of HWR performance. There are also more "better than expected" HOPD Surgery outliers (blue bar) than "worse than expected" (red bar) (64 vs. 50).
- There are 56 HOPD Surgery "worse than expected" outliers (red bar) in the fourth (Q4) or worst performing quartile of HWR. There are also more "worse than expected" (red bar) HOPD Surgery outliers than "better than expected" (blue bar) outliers (56 vs. 31).

Figure 2: Count of performance outliers on the HOPD Surgery measure within quartiles of the HWR measure surgery cohort.



Validity as assessed by the TEP

The results of the TEP rating of agreement with the validity statement were as follows:
N=13
Mean rating=5.2

All TEP members who responded to the survey indicated they agreed with the statement that "The risk-standardized hospital visit ratios obtained from the outpatient surgery measure as specified can be used to distinguish between better and worse quality facilities." 12 of the 13 indicated they moderately or strongly agreed. Two TEP members did not respond to the TEP survey.

Frequency of Ratings of Agreement

Rating	# (%) of Responses
1 (Strongly disagree)	0 (0)
2 (Moderately disagree)	0 (0)
3 (Somewhat disagree)	0 (0)
4 (Somewhat agree)	1 (7.7)
5 (Moderately agree)	8 (61.5)
6 (Strongly agree)	4 (30.8)

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)

The results of the external empiric validation analysis suggest that there is a positive, although very weak, relationship between the HOPD Surgery measure score and the measure score for the surgery cohort of the HWR measure. However, we did observe a significant relationship between outliers identified in the HOPD Surgery measure and the performance score quartiles of the HWR surgery cohort. This relationship showed more "better than expected" HOPD Surgery outliers in the best performing quartile of HWR, and more "worse than expected" HOPD Surgery outliers in the worst performing HWR quartile. This finding contributes evidence of validity, since it suggests the HOPD Surgery measure is accurately measuring the shared underlying domains of quality (safe surgical practices, care coordination at discharge) shared by both outcome measures.

Survey results from the TEP indicate high agreement (12/13 respondents "moderately or "strongly" agreed) regarding the face-validity of the HOPD Surgery measure.

Taken together, these results support the validity of the HOPD Surgery measure.

2b2. EXCLUSIONS ANALYSIS

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

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

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion. Rationales for the exclusions are detailed in section S.8 of the Submission/Intent to Submit form (Denominator Exclusions).

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 (See section S.7 of the Intent to Submit/Submission form) resulted in an initial cohort of 1,249,013 procedures (Dataset #2). We then applied the following exclusion criteria (see the Intent to Submit Form, sections S.8 and S.9, for exclusion rationale) with the following number and percent of excluded surgeries as a percent of the initial cohort):

Exclusion	N	%
[All included procedures]	[1,249,013]	[100%]
Procedures for patients who lack continuous enrollment in Medicare FFS Parts A & B in the 7 days after the procedure	727	0.06
Surgeries that occur on the same day and at the same hospital as an ED visit that was billed on a different claim than the index surgery, unless the ED visit has a diagnosis indicative of a complication of care.	11,907	0.95
Surgeries that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care	8,575	0.69
Surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit	6,166	0.49
Surgeries that are billed on the same hospital outpatient claim as an observation stay	62,214	4.98
Final Cohort	1,172,087	93.84

The final cohort, after applying all inclusion and exclusion criteria, includes 1,172,087 procedures.

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 pCitation 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*)

The exclusions for this measure are narrowly targeted and the rationale for each exclusion is presented in section S.9 of the ITS/submission form.

The largest exclusion (about 5% of surgeries) removes surgeries that occur on the same claim as an observation stay. We remove these surgeries because it is not possible to use claims data to determine whether the surgery was the cause of, subsequent to, or during the observation stay.

In total, exclusions remove a small number (about 7%) of surgeries.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 21 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

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.

Risk Model:

To calculate a HOPD RSHVR, the measure uses a two-level hierarchical logistic regression model (see details below). We model the log-odds of the outcome from an index outpatient surgery as a function of the patient demographic, procedure, and clinical characteristics, and a random outpatient facility-specific intercept. This strategy accounts for within-facility correlation of the observed outcome and sample size differences, and accommodates the assumption that underlying differences in quality across HOPDs lead to systematic differences in outcomes. This approach is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines [1-3].

We fit a hierarchical generalized linear model (HGLM), which accounts for the clustering of observations within HOPDs. We assume the outcome is a known exponential family distribution and relates linearly to the covariates via a known link function, h. For our model, we assumed a binomial distribution and a logit link function. Further, we accounted for the clustering within HOPDs by estimating a facility-specific effect, α_i, which we assume follows a normal distribution with mean μ and variance τ^2, the between-facility variance component. The following equations define the HGLM:

$$(1)\quad h\left(\Pr\left(Y_{ij}=1|\boldsymbol{Z}_{ij,}\omega_i\right)\right)=\log\left(\frac{\Pr\left(Y_{ij}=1|\boldsymbol{Z}_{ij,}\omega_i\right)}{1-\Pr\left(Y_{ij}=1|\boldsymbol{Z}_{ij,}\omega_i\right)}\right)=\alpha_i+\boldsymbol{\beta Z}_{ij}$$

$$where\ \alpha_i=\mu+\omega_i;\ \omega_i\sim N(0\tau^2)\Big/$$

$$i=1\ldots I; j=1\ldots n_i$$

Where Yij denotes the outcome (equal to 1 if patient has one or more qualifying hospital visits within 7 days, 0 otherwise) for the *j*-th patient who had a surgical procedure at the *i*-th HOPD; Zij = $\boldsymbol{Z}_{ij}$ = $(Z_{1ij}, Z_{2ij}, Z_{pij})$ is a set of *p* patient-specific covariates derived from the data; and I denotes the total number of HOPDs; and n_i the number of surgeries performed at HOPD *i*. The facility-specific intercept of the *i-th* HOPD, α_i, defined above, comprises μ, the adjusted average intercept over all HOPDs in the sample, and ω_i, the facility specific intercept deviation from μ. A point estimate of ω_i, greater or less than 0, determines whether HOPD performance is worse or better compared to the adjusted average outcome.

Risk Variables:

The risk-adjustment model includes 25 patient-level variables, including age, clinical comorbidities, and indicators of surgical complexity obtained from both Part A and B inpatient, outpatient and carrier claims 12 months prior to index procedure. Data dictionary tab "HOPD_Surg_Risk_Factors_CCs" presents the definition of these variables, based on CMS's hierarchical condition categories (CCs). The selection of risk factors was informed by the peer-reviewed literature, an open review process including comments from stakeholders and the public, and empirical analyses. CORE also convened, through a public process, a national technical expert panel (TEP) consisting of patients, surgeons, methodologists, researchers, and providers.

The risk-adjustment methodology does not include specific acute conditions if they occur only during the index procedure because they could be consequences of care (also called the complication-of-care variables); please see data dictionary "HOPD_Surg_RF_CoC" tab for a summary of these diagnoses.

The odds ratios for the risk variables in the final model are shown below in Table 1. For a detailed description of the development and refinement of the risk-adjustment model, see the original measure development methodology report: (direct link: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Hospital-Visits-after-Hospital-Outpatient-Surgery-Measure.pdf).

Table 1: Logistic Regression Model Variable Odds Ratios (January 1, 2018-December 31, 2018; Dataset #2)

Parameter	Odds Ratio	95% CI
Age minus 65 (years above 65)	1.02	1.02-1.03
Comorbidities:		
Cancer (CC 8-14)	1.02	1.00-1.03
Diabetes and DM Complications (CC 17-19, 122, 123)	1.15	1.13-1.17
Disorders of Fluid/Electrolyte/Acid-Base (CC 24)	1.15	1.13-1.17
Intestinal Obstruction/Perforation (CC 33)	1.17	1.13-1.17

Parameter	Odds Ratio	95% CI
Inflammatory Bowel Disease (CC 35)	1.07	1.00-1.13
Bone/Joint/Muscle Infections/Necrosis (CC 39)	1.37	1.32-1.44
Hematological Disorders Including Coagulation Defects and Iron Deficiency (CC 46, 48, 49)	1.12	1.10-1.14
Dementia or Senility (CC 51-53)	1.18	1.15-1.21
Psychiatric Disorders (CC 57-63)	1.15	1.13-1.17
Hemiplegia, Paraplegia, Paralysis, Functional Disability (CC 70, 71, 73, 74, 103-105, 189-190)	1.18	1.14-1.22
Other Significant CNS Disease (CC 77-80)	1.18	1.14-1.22
Cardiorespiratory Arrest, Failure and Respiratory Dependence (CC 82-84)	1.06	1.03-1.09
Congestive Heart Failure (CC 85)	1.13	1.10-1.15
Ischemic Heart Disease (CC 86-89)	1.14	1.12-1.16
Hypertension and Hypertensive Disorders (CC 94, 95)	1.08	1.06-1.10
Arrhythmias (CC 96, 97)	1.13	1.11-1.15
Vascular Disease (CC 106-109)	1.14	1.12-1.16
Chronic Lung Disease (CC 111-113)	1.13	1.11-1.15
UTI and Other Urinary Tract Disorders (CC 144, 145)	1.14	1.12-1.15
Pelvic Inflammatory Disease and Other Specified Female Genital Disorders (CC 147)	0.90	0.86-0.93
Chronic Ulcers (CC 157-161)	1.10	1.06-1.13
Cellulitis, Local Skin Infection (CC 164)	1.17	1.14-1.19
Prior Significant Fracture (CC 169-171)	1.41	1.37-1.45
Morbid Obesity (CC 22)	1.15	1.12-1.19
Work Relative Value Units	1.12	1.11-1.12
Body System Operated On:		
Cardiovascular	1.99	1.77-2.23
Digestive	3.34	2.98-3.74
Ear	Reference	
Endocrine	1.86	1.64-2.12
Female Genitalia	2.85	2.51-3.24
Hemic-Lymphatic	2.10	1.78-2.48
Skin & Breast	1.47	1.31-1.65
Male Genitalia	3.75	3.34-4.21
Miscellaneous Procedures	1.03	0.42-2.52
Musculoskeletal	2.39	2.13-2.68
Nervous	2.99	2.67-3.36
Nose-Throat-Pharynx	2.59	2.27-2.95
Respiratory	2.56	1.99-3.29

Parameter	Odds Ratio	95% CI
Urinary	3.89	3.47-4.36

Citations

1. Normand S-LT, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. . *Stat Sci.* 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.* Jan 24, 2006;113(3):456-462.

3. National Quality Forum. Measure Evaluation Criteria and Guidance on Evaluation. September 2019. http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/2018_Measure_Evaluation_Criteria_and_Guidance.aspx

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?

Selecting Risk Variables (done during measure development; Dataset #1)

The measure adjusts for HOPD case mix and surgical procedure mix differences based on patient demographic and clinical characteristics, and surgical procedural complexity. Risk adjustment is necessary to ensure that variation in the measure score among providers is due to quality of care rather than differences in case mix or surgical procedure mix.

When CMS originally developed the measure, CORE considered candidate variables for risk-adjustment that had an association with adverse surgical outcomes or hospital visits following surgery as identified in the literature and through expert clinical input and statistical testing. These included:

•Patient age, sex, and comorbidity variables
For candidate variables, based on the literature and clinical input, CORE identified variables of interest that were both clinically relevant and had a documented or clinically plausible relationship with the outcome. The candidate patient variables are shown in Appendix E of the original methodology report (direct link: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-

Instruments/HospitalQualityInits/Downloads/Hospital-Visits-after-Hospital-Outpatient-Surgery-Measure.pdf).

•Surgical procedural complexity
We tested two candidate variables of procedural complexity:

Work Relative Value Unit (RVU) of the procedure: Work RVUs are assigned to each CPT procedure code and approximate procedure complexity by incorporating elements of physician time and effort. Surgeries with increasing complexity are assigned a higher Work RVU. For patients with multiple concurrent CPT procedure codes, we risk adjust for the CPT code with the highest Work RVU value.

Anatomical body system group using Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS). We use the body system variable, in addition to the Work RVU of the procedure, to account for organ-specific difference in risk and complications that are not adequately captured by the Work RVU alone. This approach to accounting for procedural complexity is similar to that described in the literature and used for risk adjustment in the American College of Surgeons' National Surgical Quality Improvement Program (NSQIP)[1].

Final Variable Selection
To select the final variables to include in the risk-adjustment model, we fitted an initial logistic regression model with all candidate variables to predict the outcome of hospital visits within 7 days. The Development Dataset (Dataset #1a) was a randomly selected split sample of our 2010 cohort. To develop a parsimonious model, we then iteratively removed non-significant variables from the initial model using a stepwise purposeful selection approach described by Hosmer and Lemeshow [2]. We retained all variables significant at $p<0.05$ in the final model.

Social Risk Factors for Disparities Analyses

CMS submitted the HOPD Surgery measure for NQF endorsement in January of 2015, prior to the NQF Sociodemographic Status (SDS) trial. Therefore, according to NQF guidance, results of social risk factor testing were not considered in the risk adjustment for this measure. However, during NQF public comment during initial endorsement, two stakeholders noted their concern regarding the lack of social risk factor adjustment. Accordingly, in response to public comment, we provided NQF with the results of social risk factor testing results that had been completed, which were consistent with the updated testing provided below. The Standing Committee voted to endorse the measure without adjustment for social risk factors, and NQF's Consensus Standards Approval Committee (CSAC) voted to uphold the Standing Committee's endorsement, following a discussion about social risk factor adjustment [3].

Since the measure was endorsed, we have updated the measure in response to feedback from stakeholders (as discussed in section S.3.2 of the Measure Submission/ITS form). CMS initiated a dry run in 2017 in preparation for 2020 public reporting, but did not receive any feedback that resulted in re-examination of risk variables, including social risk factors. (Note that hospitals received their confidential facility-level scores in November 2019; CMS will report facility-level measure scores to the public on Hospital Compare in January 2020.)

For this re-endorsement application, we re-analyzed the effects of social risk factors on the models, incorporating the evolution in both policy and technical approaches from the past few years. CMS reviewed these results, and after careful consideration within the context of the conceptual model

outlined below in this section, decided not to adjust the measure for social risk factors. The details regarding the methods, results, and interpretation of results are in section 2b3.4b, below.

We selected social risk factor variables 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 social risk factors may influence risk of the outcome.

Causal Pathways for Social Risk Variable Selection

Our conceptualization of the pathways by which patients' social risk factors affect the outcome was informed by the literature [4-10] and IMPACT Act–funded work by the National Academies of Sciences, Engineering and Medicine (NASEM) and the Department of Health and Human Services Assistant Secretary for Planning and Evaluation (ASPE) [11-13].

Literature Review of Social Risk Variables and Ambulatory Surgery Post-Procedure Hospital Visits

To inform a conceptual model for the relationship of social risk factors to the outcome we performed a literature search during development of the original measure in 2016 that included articles that contained key words in the title or abstract related to outpatient surgeries or procedures, socioeconomic and sociodemographic disparities, and hospital visits (emergency department, observation, or hospital admission). We excluded any 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 social risk factors and hospital visits after outpatient 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 outpatient surgery setting [4,5]. An updated literature search performed in November of 2019 identified two additional studies. In a 2016 study, authors found that patients in "high-risk" communities undergoing outpatient thyroidectomy were more likely to be operated on by low-volume surgeons, and that patients in these communities were more likely to have worse post-operative outcomes, including a higher risk for hospital admission [6]. In one 2017 study, reserachers found that Medicaid status was independently associated with an increase in the odds of an unplanned hospital admission following urethral sling placement, and that the increase remained after controlling for patient comorbidities, demographics, and facility characteristics [7].

Conceptual Pathways for Social Risk Factor Variable Selection

Although there is limited literature linking social risk factors and adverse outcomes, we identified the following potential pathways through which social risk factors may influence the outcome of 7-day visits following outpatient surgery, based on the specific clinical consideration of the procedure and the broader social risk factor literature.:

1.**Differential care within a facility or unmet differential needs**. One pathway by which social risk factors may contribute to post-surgical hospital visit risk is that patients may not receive equivalent care within a facility [8,11]. However, as noted above, studies in the outpatient surgery setting are lacking. Moreover, patients with social risk factors, such as lower education, may require differentiated care – e.g., provision of information at a lower health literacy level – to achieve outcomes comparable to those of patients without social risk factors. Facilities that do not identify the need for and provide such care could have worse outcome rates for their patients with social risk factors.

2. Use of lower-quality facilities. Patients may differentially obtain care in lower quality facilities. With respect to inpatient 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 [9,10]. In the outpatient setting, as described above, there is evidence that patients with social risk factors may receive services at facilities that have surgeons with less experience, putting patients at higher risk of a post-surgical visit [6].

3. Influence of social risk factors on hospital visit risk outside of facility quality. Some social 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 facility. For instance, while a surgeon and/or a facility may make appropriate care decisions and provide tailored care and education, we hypothesized that a lower-income patient may still have a worse outcome post-procedure due to factors such as a limited understanding of the discharge plan, or a lack of home support, transportation or other resources for following discharge instructions. These factors, however, can be anticipated and addressed for outpatient elective surgeries more readily than in more emergent care contexts.

4. Relationship of social risk factors with patients' health at admission. Patients with lower income/education/literacy for unstable housing may have worse general health status and may present for their procedure with greater severity of underlying illness [11]. This causal pathway should be largely accounted for by current clinical risk-adjustment.

As indicated in section 1.8, the social risk variables that we examined were:

- Dual-eligible status
- AHRQ-validated SES Index score

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. 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 ratios at the facility 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 coding updates. In addition, changes to ICD-10 codes are continually monitored for their potential impact on this measure, and updates are made accordingly.

Citations

1. 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. Dec 2010;211(6):715-723.

2. Hosmer D, Lemeshow S. *Applied Logistic Regression.* New York: John Wiley & Sons; 2000.

3. NQF-Endorsed Measures for Surgical Procedures, 2015, Final Report. December 23, 2015. http://www.qualityforum.org/Projects/s-z/Surgery_Measures_2014/Final_Report.aspx. Accessed December 18, 2019.

4. Mioton LM, Buck DW, 2nd, Rambachan A, Ver Halen J, Dumanian GA, Kim JY. Predictors of readmission after outpatient plastic surgery. Plast Reconstr Surg. 2014;133(1):173-180.

5. Bhattacharyya N. Healthcare disparities in revisits for complications after adult tonsillectomy. Am J Otolaryngol. 2015 Mar-Apr;36(2):249-253.

6. Al-Qurayshi Z, Randolph GW, Srivastav S, et al. Outcomes in thyroid surgery are affected by racial, economic, and healthcare system demographics. Laryngoscope. 2016 Sep;126(9):2194-9.

7. Dallas KB, Sohlberg EM, Elliott CS, Rogo-Gupta L, et al. Racial and Socioeconomic Disparities in Short-term Urethral Sling Surgical Outcomes. Urology. 2017 Dec; 110:70-75.

8. Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. New Engl J Med. 2014; 371:2298-2308.

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 Aff. 2011; 30:1904-1911.

10. Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA Surg. 2014; 149:475-481.

11. 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.

12. 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.

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

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

☒ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

The final list of risk of clinical, procedural and demographic variables was selected during development and is shown here and defined in the data dictionary in tab “HOPD_Surg_Risk_Factor_CCs”.

Age minus 65 (years above 65) – Mean (SD)
Cancer (CC 8-14)
Diabetes and DM Complications (CC 17-19, 122, 123)
Disorders of Fluid/Electrolyte/Acid-Base (CC 24)
Intestinal Obstruction/Perforation (CC 33)
Inflammatory Bowel Disease (CC 35)
Bone/Joint/Muscle Infections/Necrosis (CC 39)
Hematological Disorders Including Coagulation Defects and Iron Deficiency (CC 46, 48, 49)
Dementia or Senility (CC 51-53)
Psychiatric Disorders (CC 57-63)
Hemiplegia, Paraplegia, Paralysis, Functional Disability (CC 70, 71, 73, 74, 103-105, 189, 190)
Other Significant CNS Disease (CC 77-80)
Cardiorespiratory Arrest, Failure and Respiratory Dependence (CC 82-84)
Congestive Heart Failure (CC 85)
Ischemic Heart Disease (CC 86-89)
Hypertension and Hypertensive Disorders (CC 94, 95)
Arrhythmias (CC 96, 97)
Vascular Disease (CC 106-109)
Chronic Lung Disease (CC 111-113)
UTI and Other Urinary Tract Disorders (CC 144, 145)
Pelvic Inflammatory Disease and Other Specified Female Genital Disorders (CC 147)
Chronic Ulcers (CC 157-161)
Cellulitis, Local Skin Infection (CC 164)
Prior Significant Fracture (CC 169-171)
Morbid Obesity (CC 22)
Work Relative Value Units – Mean (SD)
Body System Operated On:
Cardiovascular
Digestive
Ear
Endocrine
Female Genitalia
Hemic-Lymphatic
Skin & Breast
Male Genitalia
Musculoskeletal
Nervous
Nose-Throat-Pharynx
Respiratory
Urinary

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, we evaluated two indicators of social risk: Medicaid dual-eligibility (DE), and AHRQ SES Index. Our goal for these analyses were to:

- Examine whether these factors were associated with increased risk of the outcome after adjusting for other risk factors;
- Evaluate the impact of including social risk factors on model performance, and
- Compare facilities' measure scores calculated with and without social risk factor adjustment.

To answer these questions we completed seven analyses described below. All analyses were performed with data from January 1, 2018-December 31, 2018 (Dataset #2).

Analysis #1. Distribution of social risk factors across measured entities

To assess the extent to which any effects of social risk factors may differentially influence the scores of a subset of providers, we examined how the proportion of patients with each social risk factor varied across HOPDs.

The prevalence of social risk factors varied across measured entities as shown in Table 2.

Table 2. Percent and count of patients with social risk factors, per facility

Social risk variable	Min (%)	Min (N)	Median (%)	Median (N)	Inter-quartile range (%)	Inter-quartile range (N)	Max (%)	Max (N)
DE (Yes)	0%	0	5.23%	8	2.26% - 10.00%	2-23	100%	365
AHRQ SES Index (lowest quartile)	0%	0	11.95%	17	4.57% - 24.90%	4-48	100%	701

The distribution was skewed; among the facilities in the top quartile of the distribution, the proportion of patients with social risk factors ranged from >10% to 100% for the dual eligible variable, and from >24.9% to 100% for the AHRQ SES Index. We therefore also analyze this group separately in Analyses #6 and #7 (see pages 30-33).

Analysis #2. Patient-level observed hospital visit rates for patients with social risk factors

To evaluate the association of these risk factors with the outcome, we first quantified the overall observed hospital visit rate for each social risk factor group (dual-eligible: yes vs. no, AHRQ SES Index: lowest quartile of SES Index vs. all others).

The outcome rate for patients with dual-eligible (DE) status and low AHRQ SES was higher than the outcome rate for patients who do not have the social risk factor (DE: 10.56% vs. 6.93%, p-value<0.0001; AHRQ SES: 8.23% vs. 6.99%, p<0.0001). The outcome rate for all patients was 7.17%.

Analysis #3. Strength and significance of each of the social risk factors in the context of a multivariable model for each division.

We examined the strength and significance of the social risk variables in a bivariate model (examining just the social risk factor and its relationship to the measure outcome) compared with a multivariable model (adding the social risk factor into the model with all other model variables).

In the bivariate models, both social risk factors have an odds ratio greater than one, indicating patients with the social risk factor have an increased risk of the outcome (Table 3). When we include these variables in a multivariable model that includes all of the final risk model variables, the odds ratios for both the dual eligible and AHRQ SES variables in the multivariable model were lower than the odds ratio for the bivariate association (Table 3; DE: OR 1.59 vs. 1.26; AHRQ SES: OR 1.19 vs. 1.12). This indicates that some of the relationship between hospital visits and social risk is accounted for by the final risk model variables, including clinical comorbidities. However, after the addition of the final model variables, odds ratios for both social risk factors remain significantly above 1.

Table 3. Odds ratios for DE and AHRQ SES SRFs in a bivariate vs. multivariate model

	Bivariate			Multivariate		
	Odds ratio	**95% CI**	**p-value**	**Odds ratio**	**95% CI**	**p-value**
DE (Yes vs No)	1.59	1.63 - 1.55	<0.0001	1.26	1.30 - 1.23	<0.0001
AHRQ SES Index (lowest quartile vs. all others)	1.19	1.22 - 1.17	<0.0001	1.12	1.15 - 1.10	<0.0001

Analysis #4. Model performance with and without each social risk factor

To understand the effect of each risk factor on the performance and predictive ability of the risk adjustment model, we compared the c-statistic with and without the addition of each of the social risk factors. The results shown below in Table 4 indicate that entering the DE and low AHRQ SES Index variables into the risk-adjustment model did not meaningfully improve model performance.

Table 4. Comparing c-statistics for risk adjustment models with and without social risk factors

	C-statistic (model with social risk factor)	C-statistic (model without social risk factor)
DE	0.685	0.684
AHRQ SES Index	0.684	

Analysis #5: Impact of social risk factors on measure scores

To evaluate how social risk factors affect the measure score of individual facilities, we compared RSHVRs calculated for each facility with and without each social risk factor included in the model. For these analyses we calculated Pearson correlation coefficients for the paired scores. We also show scatter plots for these same analyses. We limited these analyses to facilities with at least 30 procedures, which is the public reporting cut-off; only facilities that have at least 30 procedures during the performance period have a publicly-reported RSHVR (discussed earlier on page 9).

The results (Figures 3A and 3B, below) show that entering either of these variables into the risk-adjustment model did not substantially change hospital-level measure scores (RSHVRs). Correlation coefficients between RSHVR with and without adjustment for these factors were near 1 (0.998 for dual-eligible, 0.998 for low SES patients). This indicates that including the DE and AHRQ SES Index social risk factors in the model resulted in limited differences in HOPD's measure scores after accounting for other factors (age, comorbidities) included in the risk model.

Figure 3A: Correlation of RSHVRs with and without adjustment for dual-eligibility

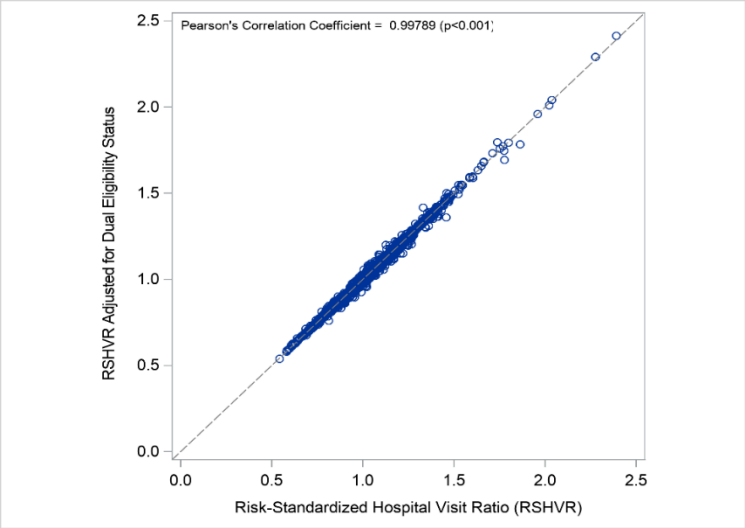
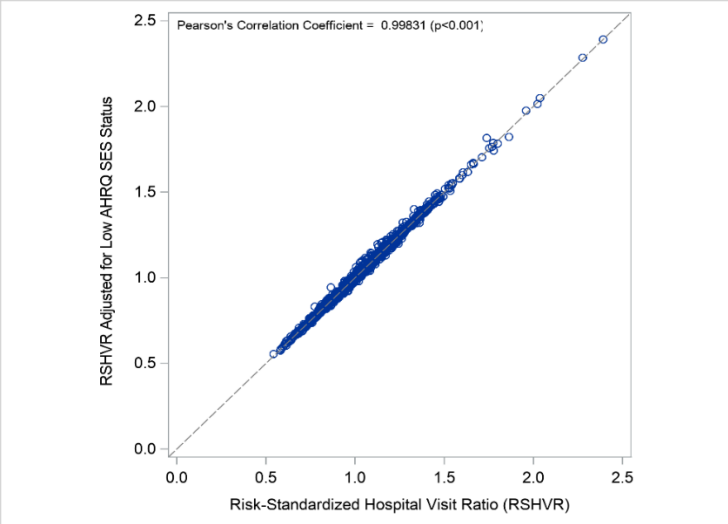


Figure 3B. Correlation of RSHVRs with and without adjustment for low AHRQ SES status



<u>Analysis #6. Comparison of measure scores between facilities with highest and lowest proportion of patients with social risk factors</u>

Distributions of the measure score (RSHVR) for facilities with a low proportion of patients with social risk factors (1st quartile) and high proportion of patients with social risk factors (4th quartile) for each social risk factor are shown in Table 5. The results showed higher measure scores for the 4th quartile (facilities with higher proportions of patients with the social risk factors) compared to the 1st quartile, but the distributions largely overlapped. The median RSHVR was slightly higher between the 1st and 4th quartiles for both variables (DE: 0.98 vs. 1.01; AHRQ SES: 0.99 vs. 1.00).

Table 5. Comparison of HOPD Surgery measure scores (RHSVR) across the distribution, between 1st and 4th quartile of the proportion of patients with the social risk factor (DE and Low AHRQ SES).

	Dual Eligible		Low AHRQ SES	
	1st Quartile for Proportion of Patients with DE status (<=3.06%)	**4th Quartile for Proportion of Patients with DE status (>9.14%)**	**1st Quartile for proportion of patients with low AHRQ SES (<=5.88%)**	**4th Quartile for proportion of patients with low AHRQ SES (>22.16%)**
Number of HOPDs	745	745	744	745
Number of patients	324,240	187,685	309,758	184,928
Maximum RSHVR	2.39	2.02	2.04	1.86
90th	1.20	1.22	1.20	1.21
75th	1.09	1.12	1.10	1.10
Median	0.98	1.01	0.99	1.00
25th	0.88	0.92	0.90	0.91
10th	0.80	0.86	0.81	0.85
Minimum RSHVR	0.62	0.54	0.54	0.58

Analysis #7. Relationship between RSHVR and percent of patients with social risk factors in facilities in the highest quartile for proportion of patients with the social risk factor

Finally, for the quartile of facilities with the highest proportion of patients with social risk factors, we plotted the relationship between the proportion of a facilities' patients with each risk factor (x-axis) and risk-standardized hospital visit ratios (RSHVRs) (y-axis) in a scatter plot, and calculated the strength of the relationship between the facility-level measure score and the facility's proportion of patients with social risk factors using the unweighted Spearman's correlation coefficient (Figures 4A and 4B, below). The results show a weak correlation between the proportion of patients at the facility with either the DE or low SES status and the measure score.

Figure 4A and 4B. Relationship between the proportion of patients with dual-eligible status (A) or low AHRQ SES (B) and risk-standardized hospital visit ratios (RSHVRs) (for facilities in the highest quartile for the proportion of patients with the social risk factor; for facilities with at least 30 cases)

Figure 4A. Dual Eligible variable

Spearman correlation coefficient: 0.093

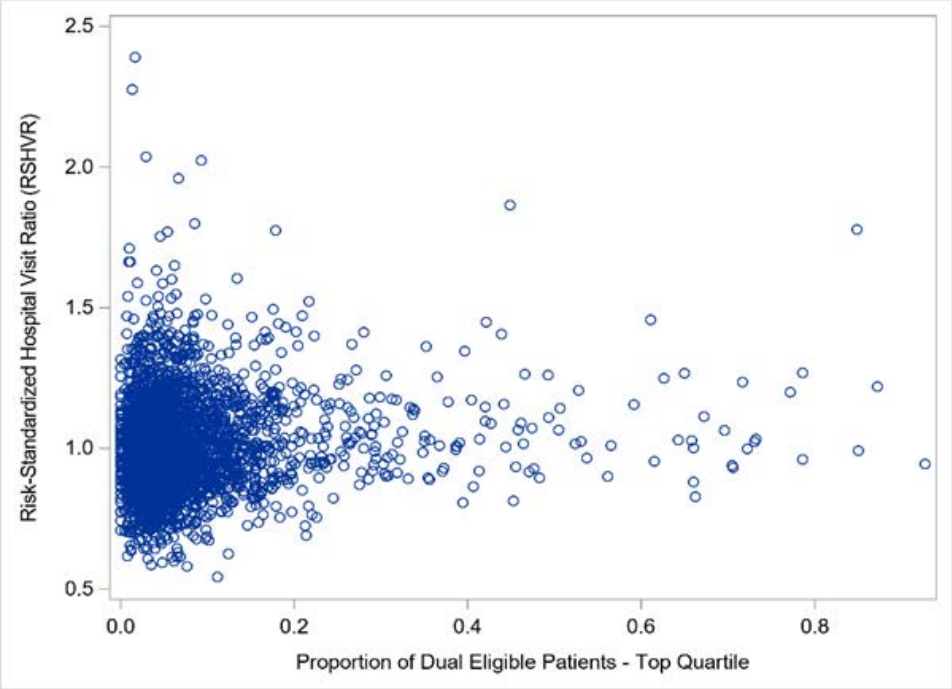
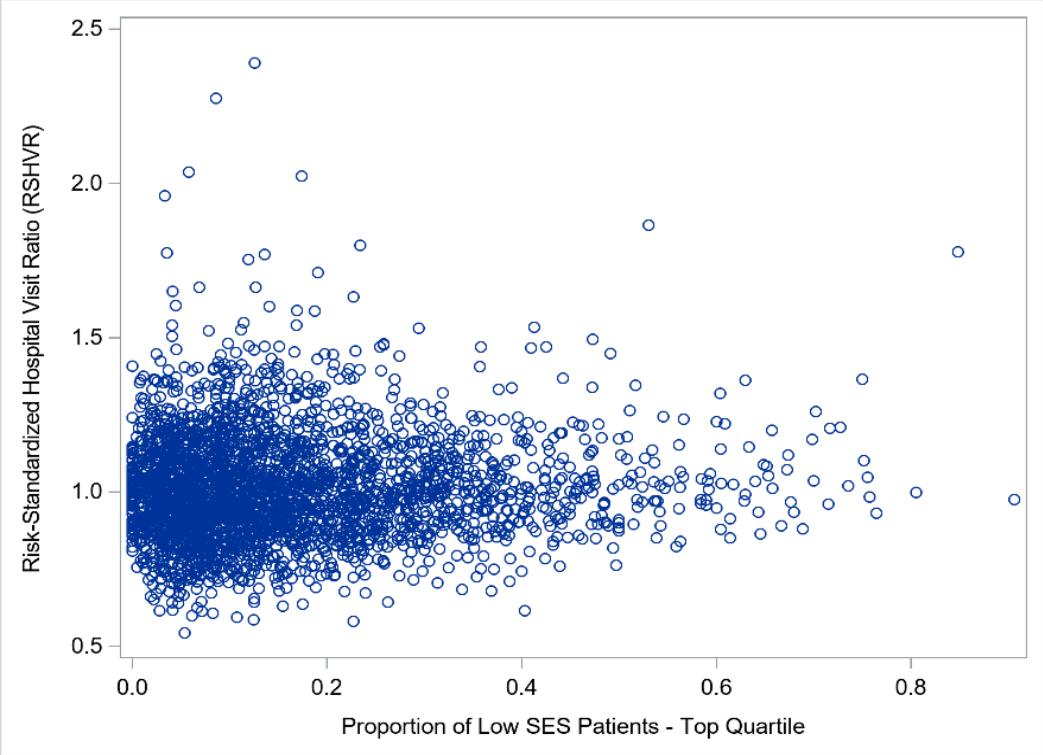


Figure 4B. AHRQ Low SES variable

Spearman correlation coefficient: 0.046



Conclusion: Social Risk Factors

The analyses above show that DE patients and patients identified as low-SES using the AHRQ SES Index are at increased risk of post-surgical hospital visits within seven days, even after adjusting for other risk factors in a multivariable model. However, the measure scores estimated for facilities with and without either social risk factor are highly correlated. And the increase in the average measure score among facilities with the highest proportion of patients with social risk factors versus those with the smallest proportion of patients with social risk factors is small, generally 0.02 or lower across the distribution with a large amount of overlap. That indicates many hospitals in the quartile with the higher number of patients with social risk factors are providing high quality care to these patients.

Nevertheless, the residual risk suggests the need to consider whether to add the two variables as risk adjusters to the measure's risk model to ensure fairness to providers care for such patients. As presented in the conceptual model (section 2b3.3a), the relationship may reflect that patients with social risk factors are receiving differential care within facilities, that facilities are missing opportunities to mitigate social risk factors they can address, that patients with these social risk factors disproportionately get care at lower quality facilities, or that patient factors that are difficult for facilities to address are driving differences in the outcome. The extent to which each of these or other factors are contributing to the measured relationship is unknown.

In making the decision about whether or not to risk adjust for these factors, CMS considered the potential unintended consequence of adjusting, and the fairness to patients and providers that care for patients with social risk factors of the unadjusted measure score. If the relationship is driven by poorer quality, adjusting will mask the disparity in care. In contrast, an unadjusted measure will illuminate quality differences and create an incentive to mitigate them. Not adjusting, however may disadvantage providers who care for patients with social risk factors, and unintentionally create an incentive for providers to care for fewer patients with social risk factors, potentially reducing access to outpatient surgery. CMS considers this risk limited, given that the correlations between the measure scores and facilities' proportions among the facilities with the most low-SES patients (as defined by DE and the AHRQ SES Index) are weak.

Given the testing results, CMS decided that on balance, the benefits of a measure that can illuminate the potential disparities for beneficiaries with the two social risk factors outweigh the concerns of fairness or unintended consequences of not adjusting for these. CMS therefore has decided not to adjust this measure for either DE or the AHRQ SES Index. CMS, however, plans to test approaches to

stratifying this measure by social risk factors under the IMPACT Act and will continue to assess the issue in measure reevaluation.

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

We computed three summary statistics for assessing model performance for the HOPD Surgery measure [1]:

Discrimination Statistics

(1) Area under the receiver operating characteristic (ROC) curve (c-statistic)

The c-statistic is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome. To calculate the c-statistic, observed hospital visit ratios were compared to predicted hospital visit probabilities across predicted rate deciles. We used Dataset #2 for this analysis.

(2) Predictive ability

Discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, for a model with good predictive ability we would expect to see a wide range in hospital visit ratios between the lowest decile and highest decile. To calculate the predictive ability, we calculated the range of observed hospital visit ratios between the lowest and highest predicted deciles. We used Dataset #2 for this analysis.

Calibration Statistics

(3) Over-fitting indices

Over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset 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. We used Dataset #1 for this analysis.

Citations

1. Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

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

If stratified, skip to 2b3.9

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

C-statistic: 0.684

Predictive Ability, % (lowest decile - highest decile): 2.26-18.02

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

We present the calibration statistics from original measure development (Datasets #1a and #1b). Please note that while the model is recalibrated yearly, coefficients remain similar.

2010 Development Sample results:

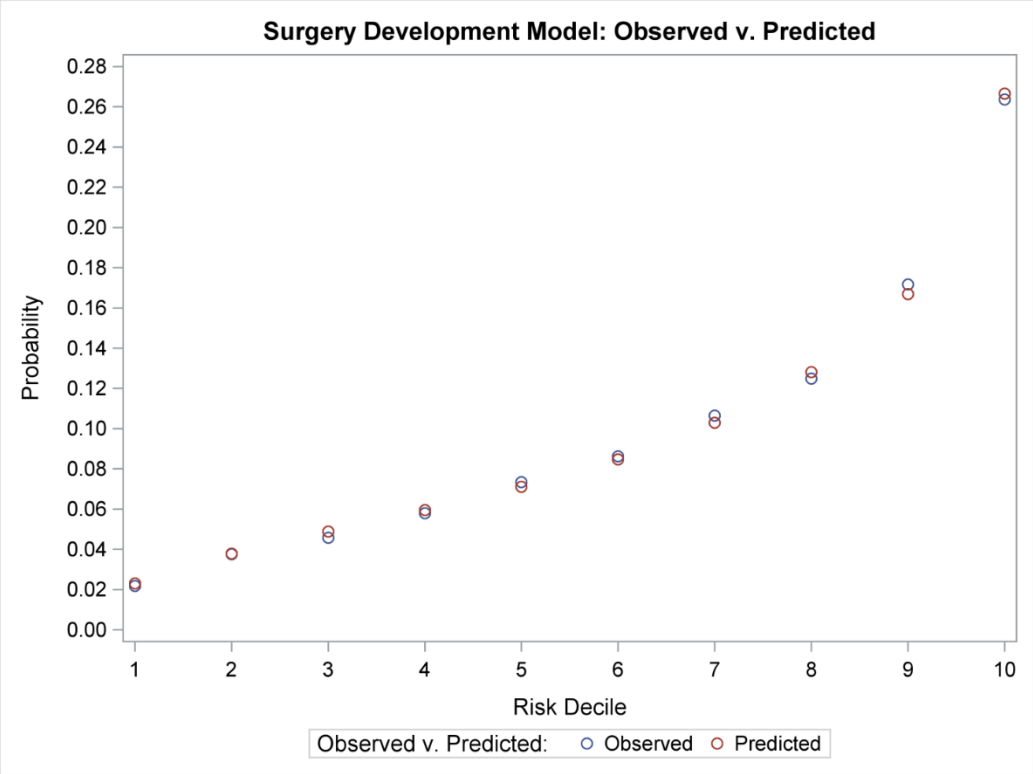
Calibration: (0,1)

2010 Validation Sample results:

Calibration: (-0.05, 0.96)

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

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing predicted values for risk deciles (2010 Medicare 20% FFS Development Sample; Dataset #1a).



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*)

Discrimination Statistics

The c-statistic of 0.684 indicate good model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. Our results show a calibration value of close to 0 at one end and close to 1 to the other end indicating good calibration of the model.

Risk Decile Plots

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. The risk decile plot shown in 2b3.8 indicates good discrimination of the model and good predictive ability.

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*)

Not applicable.

_____

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 a facility-level risk-standardized hospital visit ratio (RSHVR). The RSHVR is calculated as the ratio of the predicted to the expected number of post-surgical unplanned hospital visits among an HOPDs patients. For each HOPD, the numerator of the ratio is the number of hospital visits predicted for the HOPD's patients, accounting for its observed rate, the number and complexity of the procedures performed at the HOPD, and the patient mix. The denominator is the number of hospital visits expected nationally for the HOPDs case/procedure mix. To calculate an HOPD'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 HOPD-specific intercept. A ratio greater than one indicates that the HOPD's patients and have more visits than expected, compared to an average HOPD with similar patient and procedural complexity. A ratio less than one indicates that the HOPD's patients have fewer post-surgical visits than expected, compared to an average HOPD with similar patient and procedural complexity. For details on the measure calculation, please see the 2018 methodology report (direct link: https://www.qualitynet.org/files/5d0d367e764be766b01006a7?filename=HOPD_Surg_MsrUpdtRpt_2018.pdf).

We characterize the degree of variation by:

1) Providing the median odds ratio (MOR) [1]. The MOR represents the median

increase in odds of a hospital visit if a procedure on a single patient was performed at a higher risk

HOPD compared to a lower risk HOPD. It is calculated by taking all possible combinations of HOPDs, always comparing the higher risk HOPD to the lower risk HOPD. The MOR is interpreted as a traditional odds ratio would be.

2) Reporting the distribution of the RSHVR.

3) Presenting performance categories. 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 a HOPD is performing better than, worse than, or no different than expected. A HOPD is considered as better than expected if their entire confidence interval falls below 1, and considered worse if the entire confidence interval falls above 1. They are considered no different if the confidence interval overlaps 1.

More specifically, we use a bootstrapping procedure to compute 95% confidence intervals. Because the theoretical-based standard errors are not easily derived, and to avoid making unnecessary assumptions, we use the bootstrap to empirically construct the sampling distribution for each facility-level risk-standardized ratio. The bootstrapping algorithm is described in Appendix D of the original measure development methodology report (direct link: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Hospital-Visits-after-Hospital-Outpatient-Surgery-Measure.pdf).

All analyses were performed with Dataset #2.

Citations

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*)**

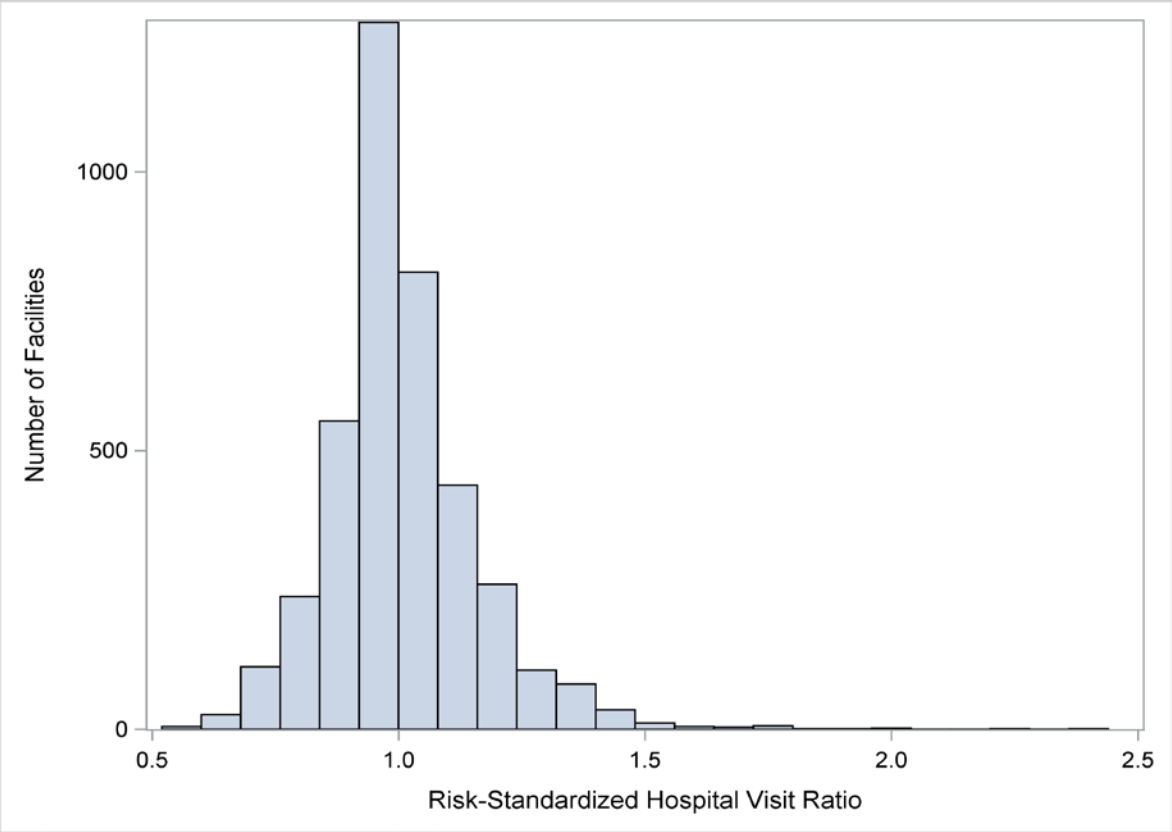
1. The median odds ratio is 1.28.

2. Percentiles of distribution for the overall measure score (RSHVR) are shown in Table 6. The distribution of the measure score is shown in Figure 5. Dataset #2 was used for this analysis.

Table 6. Distribution of RSHVR (all facilities)

Number of facilities	3974
Mean RSHVR (SD)	1.01 (0.15)
Range (min - max)	0.54 - 2.39
25th percentile	0.93
50th percentile (median)	0.99
75th percentile	1.07

Figure 5. Distribution of the measure score (RSHVR) (N=3974 facilities)



3. Table 7 shows the performance categories for facilities as measured by the HOPD Surgery score. A total of 150 facilities (3.77%) performed "Better than Expected," 2,671 facilities (67.21%) performed "No Different than Expected," and the remaining 158 facilities (3.98%) performed "Worse than Expected."

Table 7. Distribution of Facilities by Performance Category

Performance Category	HOPDs	
	Number of facilities	% distribution
Better than Expected	150	3.77%
No different than Expected	2,671	67.21%
Worse than Expected	158	3.98%
Number of Cases Too Small	995	25.04%

Note: Performance category "Number of Cases Too Small" indicates that a facility had fewer than 30 procedures, and thus its RSHVR was not reported.

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 HOPD compared to a lower risk HOPD. A value of 1.28 indicates that a patient has a 28% increase in the odds of a hospital visit if the same procedure was performed at higher risk HOPD compared to a lower risk HOPD indicating the impact of quality on the outcome rate is substantial.

The range of performance on the HOPD Surgery measure (RSHVR min-max of 0.54-2.39) demonstrates that there is a significant quality gap. Specifically, the best-performing HOPD (RSHVR of 0.54) is performing 46% better than average, whereas the worst-performing HOPD (RSHVR of 2.39) is performing 139% worse than the average. Furthermore, our outlier analysis identified about 300 or about 8 percent of HOPDs as outliers (3.77% significantly better and 3.98% significantly worse than expected). Note that the that average performer refers to an HOPD with the same case and service-line mix, performing at the average.

Overall, our results suggest that there is substantial need to both reduce the expected rate and the variation in rates across HOPDs, 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.***

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*)

Not applicable.

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*)**

Not applicable.

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*)

Not applicable.

_____

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 electronic claims

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.

This is a claims-based measure, data is generated during the course of billing. There have been no difficulties regarding data collection, availability of data, missing data, etc. Because completion of claims is required for hospital reimbursement, there is little missing data. The measures do not require any additional data collection and offer no data collection burden to facilities.

This measure has been through a confidential reporting period, as well as three years of public reporting. There have been no reports of difficulties with data collection from stakeholders during this time.

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)*.

There are no fees, licenses or other requirements.

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)
Not in use	Public Reporting Hospital Outpatient Quality Reporting (HOQR) Program https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalOutpatientQualityReportingProgram Payment Program Pay-for-reporting in HOQR. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalOutpatientQualityReportingProgram

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

Program Name, Sponsor: Hospital outpatient quality reporting program (HOQR), CMS
Implemented by CMS for outpatient services, the Hospital OQR is a national pay-for-quality-data-reporting program mandated by the Tax Relief and Health Care Act of 2006. This act requires hospitals to submit data on measures on the quality of care furnished by hospitals in outpatient settings. The HOQR program provides hospitals with a financial incentive to report their quality of care measure data and CMS with data to help Medicare beneficiaries make more informed decisions about their health care. The level of measurement is the facility; the setting is the Hospital Outpatient Department.
For the final cohorts from January 1, 2018 – December 31, 2018, there were 1,172,087 procedures performed in 3974 facilities, representing 93.8% of included procedures.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable; this measure is in use.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not applicable; this measure is in use.

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.

CMS created and distributed measure scores through "facility-specific reports" (FSRs), described below, for all 4,023 HOPDs that were open and had at least one qualifying outpatient surgery case. CMS successfully uploaded the reports to 3,961 HOPDs that had QualityNet account holders with the requisite role designation to receive FSRs. These were facilities with at least one qualified user having an active QualityNet Secure Portal account. Of these, 2,805 (70.8%) had at least one QualityNet user successfully download their report.

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.

To prepare HOPDs for public reporting, CMS conducted a confidential national reporting (dry run) of measure results from September 1 to September 30, 2017. The objectives of the dry run were to educate HOPDs and other stakeholders about the measure, allow facilities to review their measure results and data prior to public reporting, answer questions from facilities and other stakeholders, test the production and reporting process, and identify potential changes to the measure specifications.

The goals of the dry run were to:

•Educate HOPDs about the measure before its use in the Hospital OQR program;

•Allow facilities to review their measure results and data prior to public reporting;

•Answer questions and respond to comments from facilities and other stakeholders;

•Collect feedback from facilities to inform measure reevaluation activities;

•Test the data production and reporting process; and

•Identify potential upgrades to the measure specifications.

To achieve these goals, CMS took the following steps:

1.Announced the dry run to HOPDs and other stakeholders via email blasts on CMS listservs beginning in August 2017 (prior to the start of the dry run). The announcements included information on the dry run timeline, process, and key activities, along with the process for accessing a Facility Specific Report (FSR), described below in #3. Facilities were encouraged to participate and were provided with 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:

•The technical report entitled "Hospital Visits after Hospital Outpatient Surgery: 2016 Measure Updates and Specifications Report" documenting the measure's methodology;

•A measure fact sheet highlighting key information about the measure's dry run;

•A Frequently Asked Questions document responding to common questions;

•A mock FSR Excel file, populated with national data and simulated facility and state data;

•An FSR User Guide containing measure information and instructions for interpreting FSRs; and

•Additional resources (e.g., condition category to ICD-9-CM and ICD-10-CM code crosswalks).

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 with other facilities in the same state and in the United States; an FSR User Guide containing an overview of the measure's methodology and instructions for interpreting results found in the FSRs.

CMS provided these files confidentially through the QualityNet Secure Portal on September 1, 2017. Quality Improvement Organizations (QIOs) 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 before 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 .

•CMS responded to each email received.

5. Conducted a National Provider Call to present the measure's methodology, dry run process, plans for measure implementation, and answer stakeholder questions:

•CMS hosted the call on September 14, 2017.

•CMS informed stakeholders of the National Provider Call in each of its email notifications about the dry run. CMS also posted information about the call on the QualityNet website.

•Each of the calls consisted of a review of the logistics of the dry run and a presentation of the measure's methodology and quality improvement objectives, followed by a Q&A session. CMS posted a recording and transcript for the call, as well as the call 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.

CMS took public comments on the measure during development and during rulemaking, and answered questions through a question and answer (Q&A) email inbox before and during the dry run, as well as a Q & A session as part of the National Provider Call.

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

Below we provide feedback obtained from those being measured during and after the dry run.

Email Q&A Period

CMS received 91 inquiries (including follow-up questions) via the surgery measure email inbox before and during the dry run. Facilities inquired about interpretation of their patient-level data, the measure's methodology, and, in a small number of cases, flagged findings in their report that seemed inconsistent with the methodology. Most of the questions CMS received were submitted by HOPDs (97.5%).

National Provider Call

Feedback from measured entities raised during the national provider call were similar to those received through the Q&A inbox. The majority of the questions asked during the National Provider Call were about the measure's methodology, such as the definition of the eligible same-day surgery, the definition of an unplanned hospital visit, risk adjustment methodology, and inclusion/exclusion criteria. Other questions pertained to the implementation of the measure in the Hospital OQR program, timing for the distribution of the FSRs with the updated data, and eligibility of the Ambulatory Surgery Centers to participate in the dry run. The transcript and recording of the National Provider Call are available on the QualityNet website.

Questions and Comments Received During the Dry Run

CMS received and responded to a variety of questions via the measure inbox during the dry run period. The most common types of questions were inquiries about specific cases in facilities' data (36.7%), followed by requests for information on the measure inclusion criteria (21.5%), and questions about the dry run process or the national provider calls (10.1%).

Stakeholders identified the following situations for the developer to consider in refinements of the measure algorithm during detailed review of the patient data in their FSRs.

1.Surgery included in the measure outcome, while performed after the patient was in inpatient status. The measure identified some surgeries as hospital outpatient surgeries that were followed by a direct inpatient admission, but the hospital identified these surgeries as having been done on an inpatient basis.

2. An "inpatient-only" procedure performed as part of the hospital encounter in conjunction with an eligible hospital outpatient surgery. In this case, the facility indicated a procedure on the list of inpatient-only procedures was performed in conjunction with a procedure on the ASC procedures list used to identify eligible outpatient surgeries for the measure.

3. Some hospital visits counted in the measure outcome considered planned by the hospital. Stakeholders reported cases they considered to be planned follow-up hospital visits (occurring on days 2-7 after the hospital outpatient surgery), but which were counted in the measure outcome. These included situations where the admission was: (1) for treatment to address an issue found during the outpatient surgery (for example, irrigation procedure after urology surgery such as trans-urethral resection of a bladder tumor (TURBT)); (2) for an inpatient procedure (such as lithotripsy) for which the earlier outpatient surgery was a preoperative procedure; and (3) a planned admission unrelated to surgery (such as cancer care).

4.Surgery performed in the ED and not in the hospital outpatient department, or after an emergency department (ED) visit. One stakeholder reported a situation where an outpatient surgery was performed in the ED and was followed by inpatient admission directly after the surgery for a second procedure to complete treatment. Another stakeholder reported a similar situation where a hospital outpatient surgery was performed after an ED visit and was followed by the inpatient admission directly after surgery. The outpatient surgery was not excluded from the measure.

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

Email Q&A Period

CMS received two questions from professional associations (2.5%). CMS did not receive any questions from QIOs.

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.

CMS worked directly with the 45 facilities that asked about specific cases in their patient-level data and responded to every inquiry received.

In response to the measure-specific feedback received, described above in 4a2.2.2., the developer made several changes to the measure. These changes are described in section S.3.2 and are outlined below.

[Note that Issue #1 outlined in 4a2.2.2 relates to corrections that need to be made to billing practices, not issues that should be addressed through changes to the measure.]

Issue #2 (from 4a2.2.2) An "inpatient-only" procedure performed as part of the hospital encounter in conjunction with an eligible hospital outpatient surgery. In this case, the facility indicated a procedure on the list of inpatient-only procedures was performed in conjunction with a procedure on the ASC procedures list used to identify eligible outpatient surgeries for the measure.

Developer note: These cases were identified as hospital outpatient surgeries because the "place of service" field on a professional services claim for the surgery indicated "outpatient". The measure includes surgeries that have an outpatient professional services claim which matches to an inpatient claim for an admission within 3 days of the procedure to capture those surgeries subject to CMS's 3-day payment window policy. The index surgery was included because it met the eligibility criteria for the measure and was not billed concurrently with any "high-risk" procedures.

"High-risk" procedures are defined by the measure as major and minor Outpatient Prospective Payment System (OPPS)/Inpatient Prospective Payment System (IPPS) procedures that are not included in the ASC same-day surgery procedure list. Specifically, the measure identifies "high-risk" procedures through the following process: i) using Addendum B of the OPPS rule, it selects procedures with status indicators of C, S, T, and Qx; ii) from this selection of procedures, it keeps those not included in the ASC procedure list (OPPS or IPPS-only procedures); and iii) it keeps those where the procedure global surgical package codes are '010' or '090' from the Medicare Physician Fee Schedule.

Change to the measure: In response to this feedback, the developer updated the approach for identifying concurrent high-risk procedure to (1) not require specific GSI values for procedures on CMS's Hospital Outpatient Prospective Payment System "inpatient only" procedures list, (2) exclude cases with a high-risk procedure identified on the outpatient or inpatient facility claim that are matched to a professional services claim having an eligible same-day surgery

Rationale: This improves the measure's ability to exclude all cases with concurrent high-risk surgery.

Issue #3: Some hospital visits counted in the measure outcome considered planned by the hospital. Stakeholders reported cases they considered to be planned follow-up hospital visits (occurring on days 2-7 after the hospital outpatient surgery), but which were counted in the measure outcome.

Developer note: The measure does not intend to count planned hospital visits in the outcome because these are not a signal of quality of care. Planned admissions are those planned by providers for anticipated medical treatment or procedures that must be provided in the inpatient setting. CMS developed an algorithm that identifies planned readmissions and applied this algorithm to the hospital outpatient surgery measure. The algorithm uses procedure codes and principal discharge diagnosis codes on each hospital claim to identify admissions that are typically planned and may occur after hospital outpatient surgery. A few specific, limited types of care are always considered planned (for example, major organ transplant, rehabilitation, or maintenance chemotherapy). Otherwise, a planned admission is defined as a non-acute admission for a scheduled procedure (for example, total hip replacement or cholecystectomy). Admissions for an acute illness or for complications of care, as well as all ED and observation stay hospital visits, are never considered

planned. Admissions occurring on the day of the surgery. Day 0 and day 1 post-surgery are considered "unplanned" as the vast majority of these admissions are inpatient admissions directly following surgery.

Changes to the measure: CMS's planned admissions algorithm (PRA) is used for several CMS performance measures and is assessed and updated regularly. See section S.2.3 for details about updates to the PAA made in 2019.

Rationale: These changes improve the accuracy of the algorithm by updating the algorithm for coding changes and decreasing the number of hospital visits that the algorithm mistakenly designated as unplanned or planned.

Issue #4: Surgery performed in the ED and not in the hospital outpatient department, or a procedure performed as an inpatient procedure, after an emergency department (ED) visit. One stakeholder reported a situation where an outpatient surgery was performed in the ED and was followed by inpatient admission directly after the surgery for a second procedure to complete treatment. Another stakeholder reported a similar situation where a hospital outpatient surgery was performed after an ED visit and was followed by the inpatient admission directly after surgery. The outpatient surgery was not excluded from the measure.

Developer note: The measure does not include surgeries that are billed on the same hospital outpatient claim as an ED visit because the timing of events is unclear. In the cases noted by stakeholders, the measure included a hospital outpatient surgery that was billed on an inpatient claim, while an emergency room revenue code was also present on the inpatient claim.

Changes made to the measure:

•Update to exclusion for surgeries that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.

Rationale: In these situations, it is not possible to use claims data to determine whether the surgery was the cause of, subsequent to, or during the ED visit. However, if the ED visit is coded with a diagnosis for a complication, the assumption is that it occurred after the surgery.

Prior to this update, surgeries billed on the same claim as an ED visit were excluded from the measure, unless the claim had a diagnosis indicating a complication of care occurred. This update further refines this exclusion to exclude surgeries that occur on the same day and on the same claim as the index surgery, unless there is a diagnosis of complication of care indicated on the claim. Additionally, we expand the exclusion criteria to exclude surgeries that are billed on the same hospital outpatient claim, but occur after the ED visit, regardless of whether complications of care are billed or not.

•Update to exclusion for surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit.

Rationale: In these situations, we assume that the surgery was subsequent to the ED visit and may not represent a routine surgery. Timing of the ED visits is determined using revenue center dates from the outpatient claim.

•Update to the exclusion for surgeries billed on the same claim as an ED visit, where the measure continues to exclude surgeries billed on the same hospital outpatient claim as an ED visit unless the primary diagnosis on the facility claim is indicative of a complication of care

Rationale: This improves accuracy of capturing the outcome by including same-day, same-claim ED visits indicative of a complication 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.

Below we present data from the 2017 dry run (data from October 1, 2015-Septebmer 30, 2016), using 2016 specifications compared with 2018 performance data (data from January 1, 2018-Decembr 31, 2018), using 2019 specifications. Note that there were changes to the measure between these two timeframes.

Characteristic=2017 value//2018 value

N facilities=4060//3874

Mean RSHVR (standard deviation) = 1.01 (0.17)//1.01 (0.15)

Minimum=0.46//0.54

25th percentile= 0.92//0.93

50th percentile=0.99/0.99

75th percentile=1.08//1.07

Maximum=2.08//2.39

No improvement in performance was seen between the 2017 dry run results and the 2018 performance data. Note that changes in the measure occurred during this timeframe.

We would not necessarily expect improvement in measure scores to date because public reporting for this measure only began in January of 2020.

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.

We have encountered no unexpected findings during implementation, including unintended impacts on patients. CMS regularly surveys providers about its quality programs, in part regarding unintended consequences of implementing its quality measures. As this measure was first publicly reported in January 2020, CMS does not yet have feedback from providers on unintended consequences of this specific measure.

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

We have identified no unexpected benefits.

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)

0697 : Risk Adjusted Case Mix Adjusted Elderly Surgery Outcomes Measure

1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

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

3366 : Hospital Visits after Urology Ambulatory Surgical Center Procedures

3470 : Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures

3490 : Admission and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

The measures are harmonized to the extent possible with other CMS claims-based measures. The HOPD Surgery measure is a claims-based measure, therefore any differences in measure specifications create no burden to facilities as the measures are calculated from data produced during the billing process. We identified the following related NQF-endorsed measures: 1. NQF 3357: Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at ASCs (ASC General Surgery) 2. NQF 3470: Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures (ASC Orthopedic) 3. NQF 3366: Hospital Visits after Urology Ambulatory Surgical Center Procedures (ASC Urology) 4. NQF 3490: Admission and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy (Chemotherapy) 5. NQF 2539: Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy 6. NQF 1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR). 7. NQF 0697: Risk Adjusted Case Mix Adjusted Elderly Surgery Outcomes Measure The outcome in measures #1-5 are the same as the outcome of CMS's HOPD Surgery

measure presented in this re-endorsement application; an unplanned hospital visit is defined as an emergency department (ED) visit, observation stay (for NQF 3357, 3470, 3366, 2539), or unplanned inpatient admission. Hence, these related measures target the same quality domains as the HOPD Surgery measure. The patient cohort is also similar in that the related measures target Medicare Fee-For-Service (FFS) patients aged 65 years and older. For those measures that focus on the same facility (HOPD) as the target, the procedure/clinical cohorts however, differ. For example, the HOPD Surgery measure includes patients undergoing general surgery at an HOPD, but not colonoscopy procedures; the chemotherapy measure includes patients undergoing chemotherapy treatment at an HOPD, but not surgery or colonoscopy. The ASC-related measures have a different target (ASCs instead of HOPDs), and are divided into separate measures for general surgery, orthopedic surgery, and urologic surgery. The HWR measure and HOPD Surgery include overlapping but distinct surgeries (inpatient vs. outpatient) and overlapping but distinct patient outcomes (hospital visits within 7 days vs. readmissions within 30 days). They address a similar patient cohort (Medicare FFS patients 65 years of age and older). NQF 0687 overlaps with the HOPD Surgery measure in terms of target (patients over 65) and has an overlapping outcome. However, NQF 0687 includes all surgeries (in- and out-patient) and is not limited to outpatient surgeries. In addition, the outcomes that are part of NQF 0687 include complications that may result in an ED visit, observation stay, or inpatient admission (such as sepsis, surgical site infection, wound disruption, and urinary tract infection). It also includes mortality as an outcome, which is not included in the HOPD Surgery measure.

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. None of the measures are competing measures.
The measures selected in the drop down are related, but not competing.

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:** Surgery_Measure_Appendix_01-14-15_v1.0_FINAL.pdf

Contact Information

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

Co.2 Point of Contact: Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-

Co.3 Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

Co.4 Point of Contact: Faseeha, Altaf, Faseeha.Altaf@yale.edu, 203-764-5700-

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 of clinicians, patients, purchasers, and experts in quality improvement to provide input on key methodological decisions.

TEP Members:

-David Chang, PhD, MPH, MBA— Massachusetts General Hospital (Associate Professor of Surgery, Department of Surgery; Director of Healthcare Research and Policy Development, Codman Center for Clinical Effectiveness in Surgery); Boston, MA

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

-Martha Deed, PhD—Member of the public; North Tonawanda, NY

-Richard Dutton, MD, MBA—Anesthesia Quality Institute (Executive Director); Park Ridge, IL

-Nestor Esnaola, MD, MPH, MBA—Temple University School of Medicine (Professor of Surgery; Chief, Surgical Oncology); Philadelphia, PA

-Charles Goldfarb, MD—Washington University School of Medicine (Associate Professor of Orthopaedic Surgery); St Louis, MO

-Lisa Ishii, MD, MHS—Johns Hopkins School of Medicine (Associate Professor, Department of Otolaryngology-Head & Neck Surgery); Baltimore, MD

-Sandra Koch, MD—Carson Medical Group (OB/GYN surgery); Carson City, NV

-Tricia Meyer, PharmD, MS—Scott & White Memorial Hospital (Associate Vice-President, Department of Pharmacy); Texas A&M University College of Medicine (Associate Professor, Department of Anesthesiology); Texas A&M Rangel College of Pharmacy (Adjunct Associate Professor, Department of Anesthesiology); Temple, TX

-Linda Radach, BA—Member of the public; Lake Forest Park, WA

-Danny Robinette, MD—Surgery Center of Fairbanks (General Surgeon; Medical Director); Fairbanks, AK

-Suketu Sanghvi, MD—The Permanente Medical Group, Kaiser Permanente (Ophthalmologist; Associate Executive Director); Oakland, CA

-Christopher Tessier, MD—Manchester Urology Associates (Urologist); Manchester, NH

-Thomas Tsai, MD, MPH—Brigham and Women's Hospital (General Surgery Resident; Administrative Chief Resident for Research); Harvard School of Public Health (Postdoctoral Fellow, Department of Health Policy and Management); Boston, MA

-Katherine Wilson, RN, MHA—AmSurg Corp (Vice President, Quality); Nashville, TN

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

CORE Measure Development Team:

-Susannah Bernheim, MD, MHS—Director of CMS Projects; Clinical Investigator, CORE

-Kanchana Bhat, MPH—Senior Project Manager, CORE

-Tasce Bongiovanni, MD, MPP—Clinical Investigator, CORE

-Elizabeth Drye, MD, SM—Project Director, CORE

-Harlan Krumholz, MD, SM—Director, CORE

-Zhenqiu Lin, PhD—Supporting Analyst, CORE

-Julia Montague, MPH—Research Project Coordinator II, CORE

-Craig Parzynski, MS—Lead Analyst, CORE

-Isuru Ranasinghe, MBChB, MMed, PhD—Project Lead, CORE

-Joseph Ross, MD, MHS—Clinical Investigator, CORE

-Rana Searfoss, BA—Research Associate, CORE

-Sharon-Lise Normand, PhD, MSc—Statistical Consultant, Professor of Biostatistics, Department of Health Care Policy, Harvard University

Surgical Consultants:

-Kevin Bozic, MD, MBA—William R. Murray Professor, M.D. Endowed Chair in Orthopaedic Surgery, and Professor and Vice Chair of the Department of Orthopaedic Surgery; University of California, San Francisco

-Simon Kim, MD, MPH— Assistant Professor of Urology, Urological Institute, University Hospitals Case Medical Center, Cleveland, OH; Smilow Cancer Hospital at Yale-New Haven Hospital ; Center for Outcomes and Public Policy Effectiveness Research (COPPER) Center, Yale University

-Sharon Sutherland, MD, MPH—Clinical Assistant Professor of Surgery, Case Western Reserve University (CWRU) Cleveland Clinic Lerner College of Medicine; Quality Improvement Officer, Cleveland Clinic

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

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

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