Partnership for Quality Measurement
Powered by Battelle

Table of Contents

Logic Model

Attach a logic model and provide a description of the relationship between structures and processes and the desired outcome. *

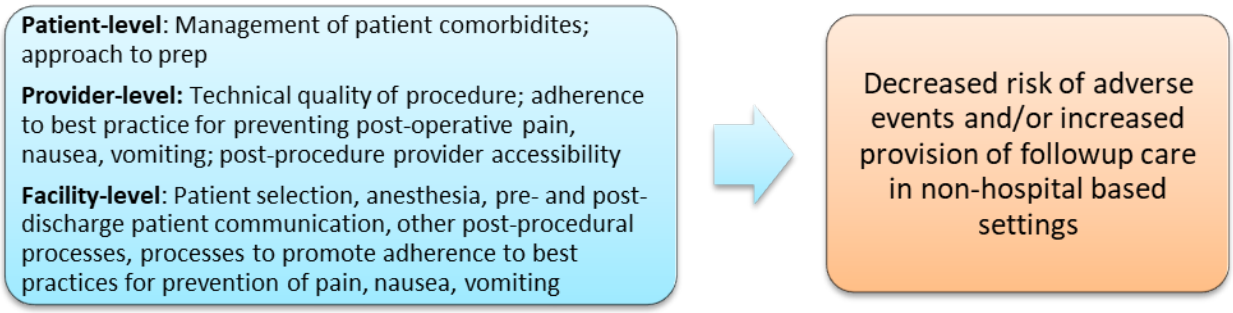
Briefly describe the steps between the health care structures and processes (e.g., interventions, or services) and the desired health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process, or outcome being measured.

Attachment (pdf, word)

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, outlined in Figure 1b, 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.

Figure 1b: Logic model for the HOPD Surgery measure showing the relationship between processes and outcomes.



Compilation of Tables and Figures- HOPD Surgery

Figure 1: Within- and Across- Facility Disparity Methods (Current Submission)

Figure 1b: Logic model for the HOPD Surgery measure showing the relationship between processes and outcomes (Current Submission)

Patient-level: Management of patient comorbidites; approach to prep

Provider-level: Technical quality of procedure; adherence to best practice for preventing post-operative pain, nausea, vomiting; post-procedure provider accessibility

Facility-level: Patient selection, anesthesia, pre- and post-discharge patient communication, other post-procedural processes, processes to promote adherence to best practices for prevention of pain, nausea, vomiting

Decreased risk of adverse events and/or increased provision of followup care in non-hospital based settings

Table 1: Distribution of measure scores (risk-standardized hospital visit ratio, or RSHVR) using the most recent testing data (2023 EM Dataset; CY2022 data) (Current Submission)

Characteristics	Value
Number of facilities	3,817
Mean RSHVR (standard deviation)	1.02 (0.24)
Minimum	0.44
25th percentile	0.89
50th percentile	0.99
75th percentile	1.10
Maximum	3.31

Table 2: HOPD Surgery Measure Score: Deciles of facility-level risk-standardized hospital visit ratio (RSHVR)* (Current Submission)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	1.02	0.44	0.70	0.82	0.89	0.94	0.98	1.00	1.05	1.10	1.20	1.53	3.31
Entities	3817	-	381	382	382	382	381	382	382	382	382	381	-
Total number of procedures	1,204,167	-	251,836	173,925	142,221	91,332	52,575	50,424	117,445	91,330	112,165	120,914	-

Table 3: HOPD Surgery Measure score performance categories using CY2022 data (EM 2023 Dataset) (Current Submission)

Performance Category	HOPDs	
	Number of facilities	% distribution
Better than Expected	220	5.8%
No different than Expected	2,427	63.6%
Worse than Expected	229	6.0%
Number of Cases Too Small (<30)	941	24.7%

Table 4: Dataset Descriptions (Prior 2020 Submission)

Dataset	Description of Dataset	Use and Section in the Testing Attachment
Dataset #1: Initial Development Dataset **Dataset #1a: Development dataset** **Dataset #1b: Validation 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. 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.	Validity testing (face validity) Identification and selection of risk-adjustment variables 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	Reliability Data Element & Measure Score Validity Testing of Measure Exclusion Selection of Social Risk Factors Meaningful Differences Predictive ability Statistical model discrimination statistics

Table 5: Dataset Descriptions (Current Submission)

Dataset Name	Description of Dataset	Types of analyses
2023 Endorsement Maintenance (EM) Dataset	Medicare-Fee-for Service claims from January 1, 2022 through December 31, 2022. Number of procedures: 1,204,167 Number of facilities: 3,817 Number of facilities with >= 30 procedures: 2,876 Mean age (SD): 74.8 % Female: 50.1	Reliability Data Element & Measure Score Validity Testing of Measure Exclusion Selection of Social Risk Factors Meaningful Differences Predictive ability Statistical model discrimination statistics

Table 6: Distribution of signal-to-noise reliability for all facilities (Current Submission)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.71	0.03	0.09	0.28	0.52	0.71	0.82	0.88	0.92	0.94	0.96	0.98	0.99
Entities	3,817	-	371	395	375	385	384	380	383	381	382	381	-
Total Number of Procedures	1,204,167	-	984	4,486	11,707	26,341	48,553	77,212	116,505	164,913	247,841	505,625	-

Table 6A. Deciles of signal to noise reliability for facilities with at least 30 procedures (January 1, 2022-December 31, 2022) (Current Submission)

	Overall	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Reliability	0.86	0.60	0.72	0.81	0.86	0.89	0.92	0.94	0.95	0.96	0.98
Entities	2,876	284	292	285	289	290	289	284	288	288	287
Total Number of procedures	1,194,500	11,735	21,279	33,119	48,786	68,508	91,608	116,928	158,407	221,304	422,826

Figure 2a: Relationship between HOPD Surgery measure score and HWR (surgery cohort) measure score (Prior 2020 Submission).

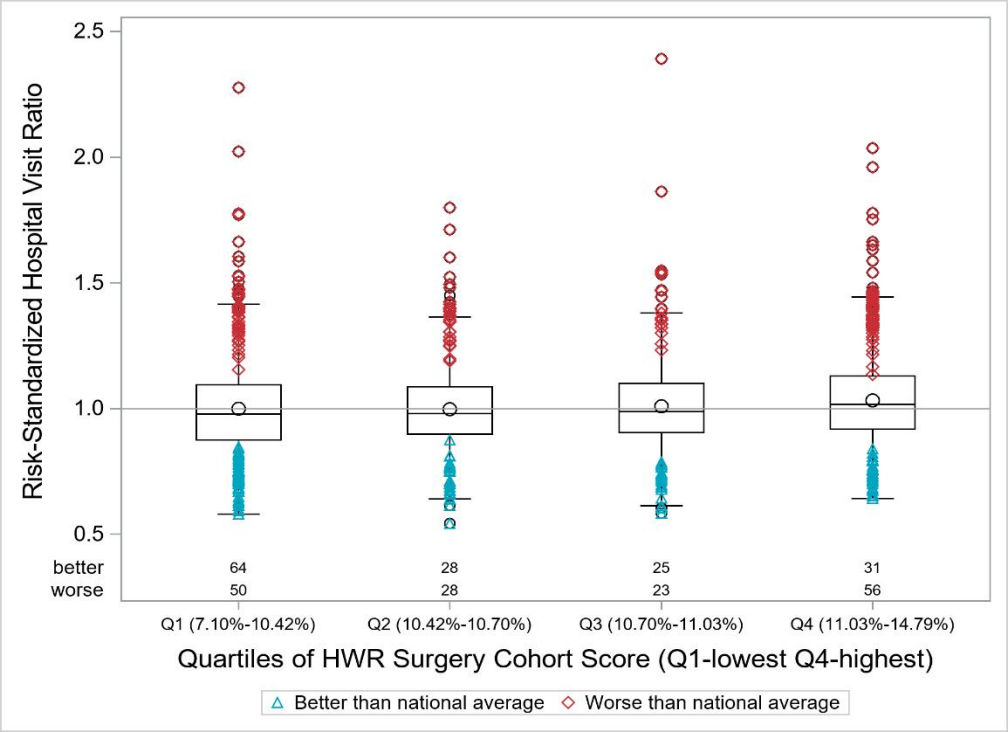


Figure 2b: Count of performance outliers on the HOPD Surgery measure within quartiles of the HWR measure surgery cohort (Prior 2020 Submission)

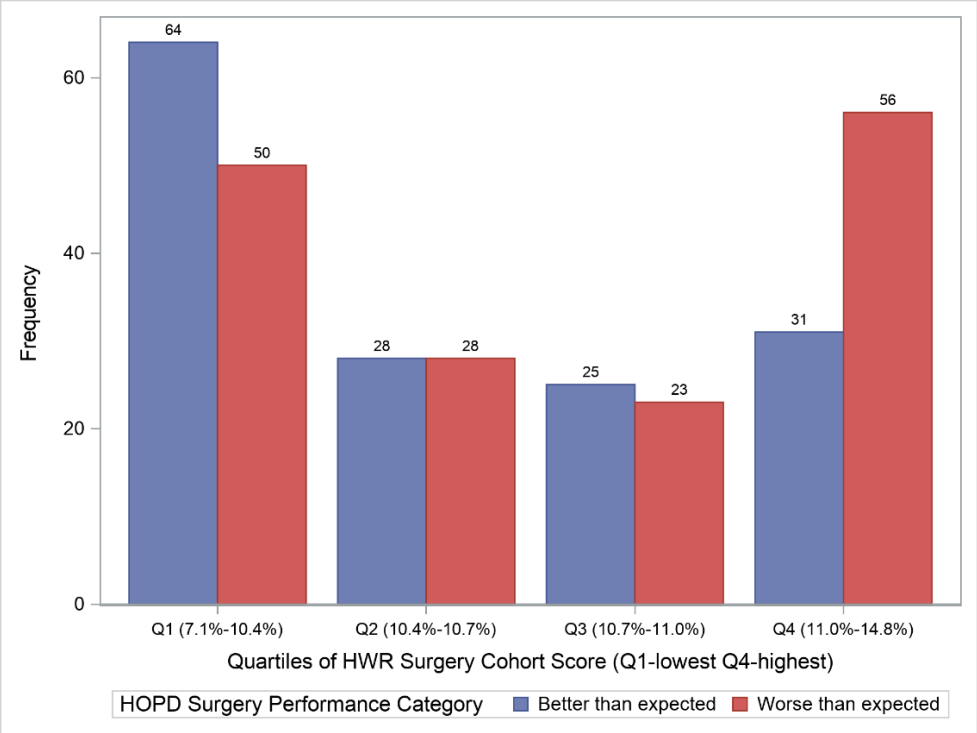


Figure 3: Distribution of RSHVR within deciles of facility procedural volume (January 1, 2022-December (Current Submission)

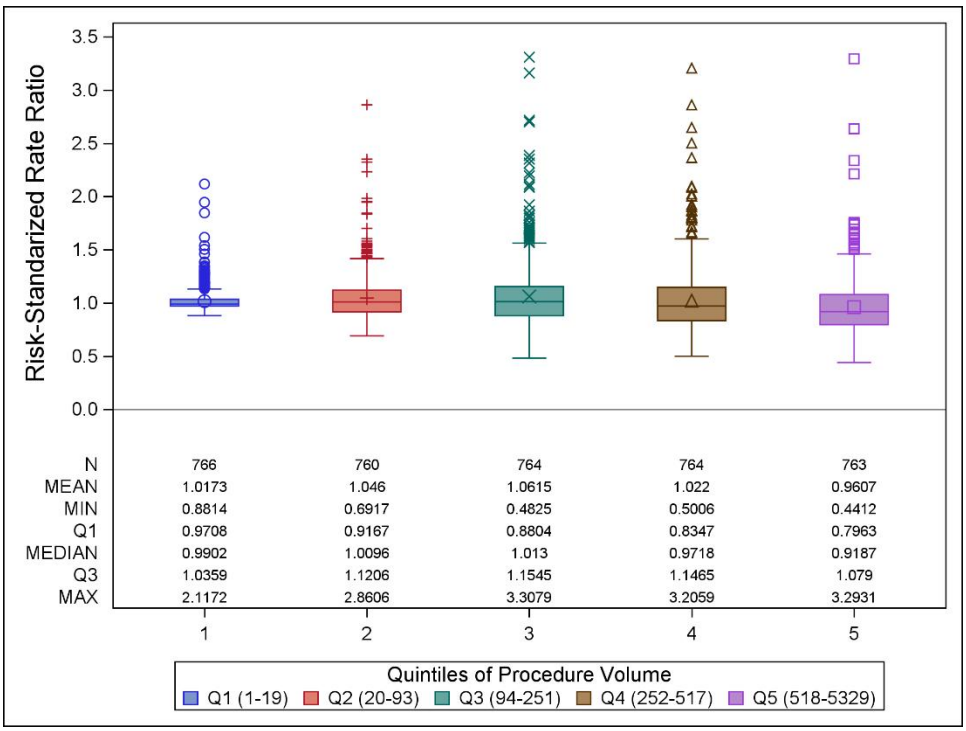


Table 7: Distribution of Dual Eligible (DE) and Area Deprivation Index (ADI) variables among HOPDs with procedures captured by the HOPD Surgery measure (January 1, 2022-December 31, 2022) (Current Submission)

Social risk variable	Min (%)	Min (N)	Median (%)	Median (N)	IQR (%)	IQR (N)	Max (%)	Max (N)
DE (Yes)	0%	0	3.43%	6	1.26% - 7.57%	1-18	100%	489
ADI (ADI >= 85)	0%	0	6.31%	7	0.71% - 18.67%	1-30	100%	529

Figure 4: Risk Decile Plot (all patients) (Prior 2020 Submission)

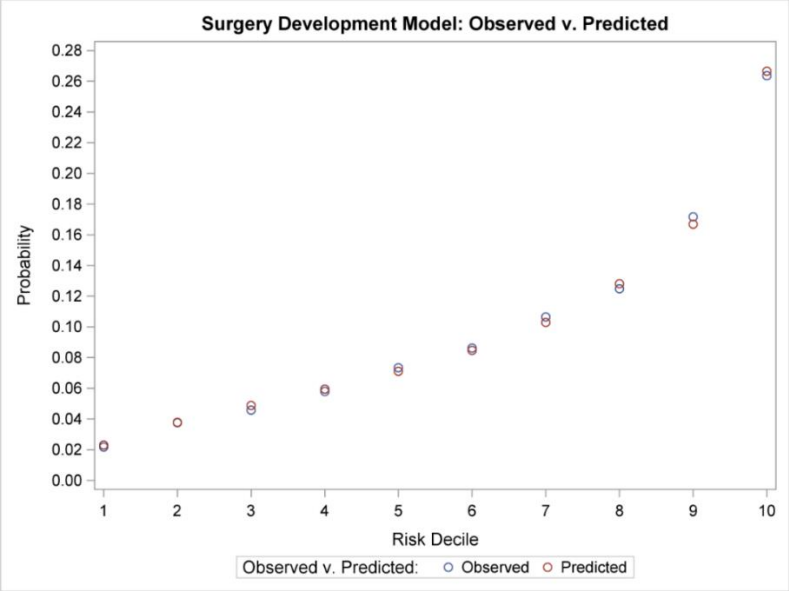


Figure 5: Risk decile plot: all patients (Current Submission)

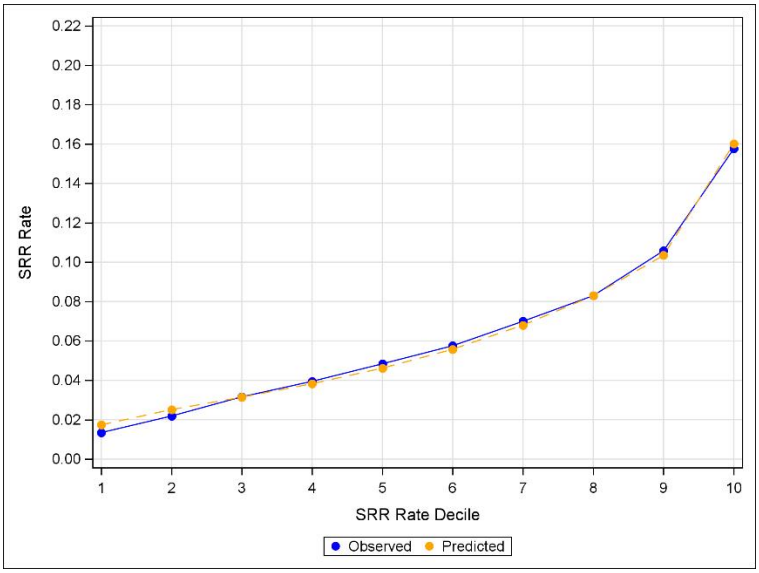


Figure 6: Risk decile plot: DE patients only (Current Submission)

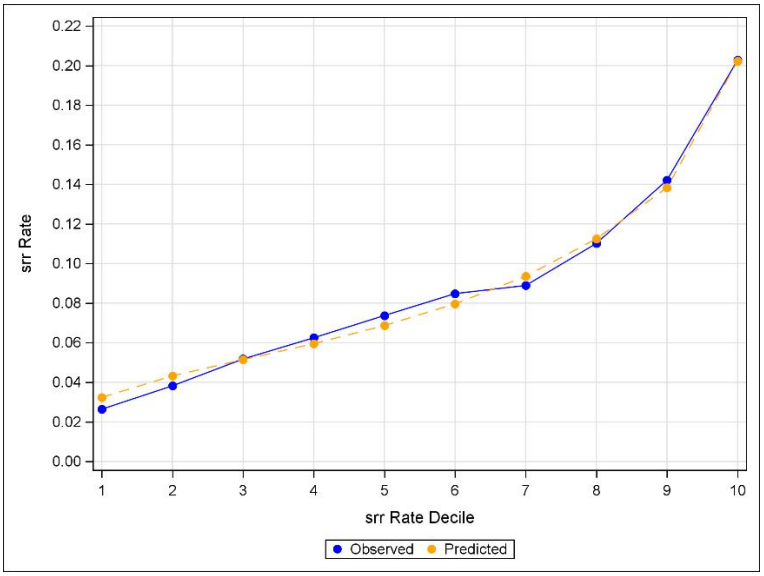


Figure 7: Risk decile plot: Patients with high ADI only (Current Submission)

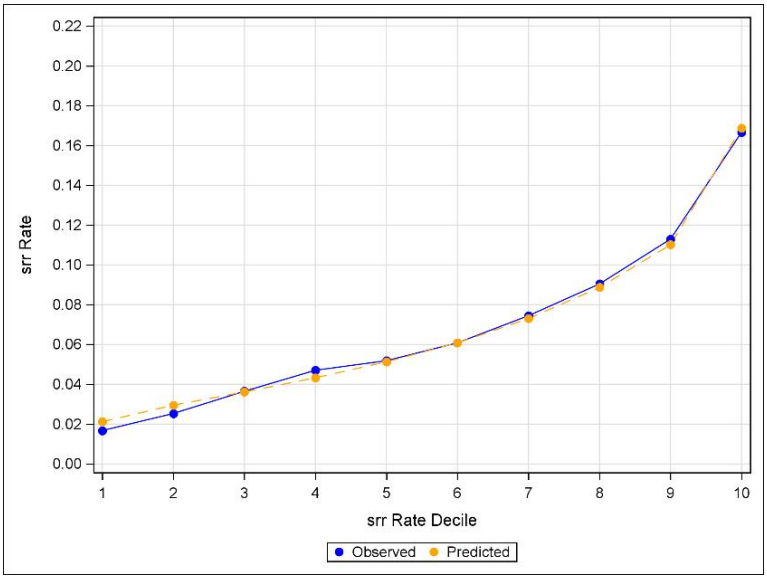


Figure 8: Correlation between measures scores calculated with and without DE in the risk model (Current Submission)

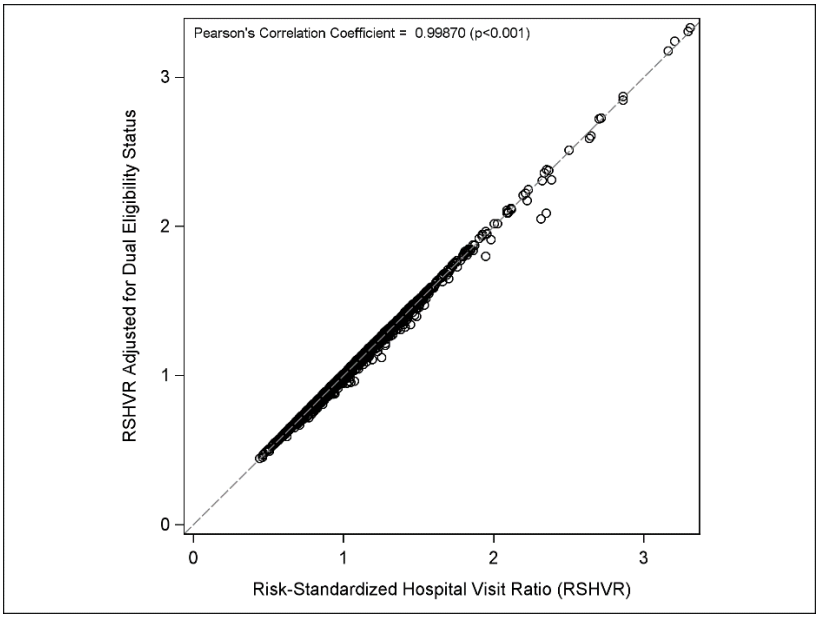


Figure 9: Correlation between measures scores calculated with and without high ADI in the risk model (Current Submission)

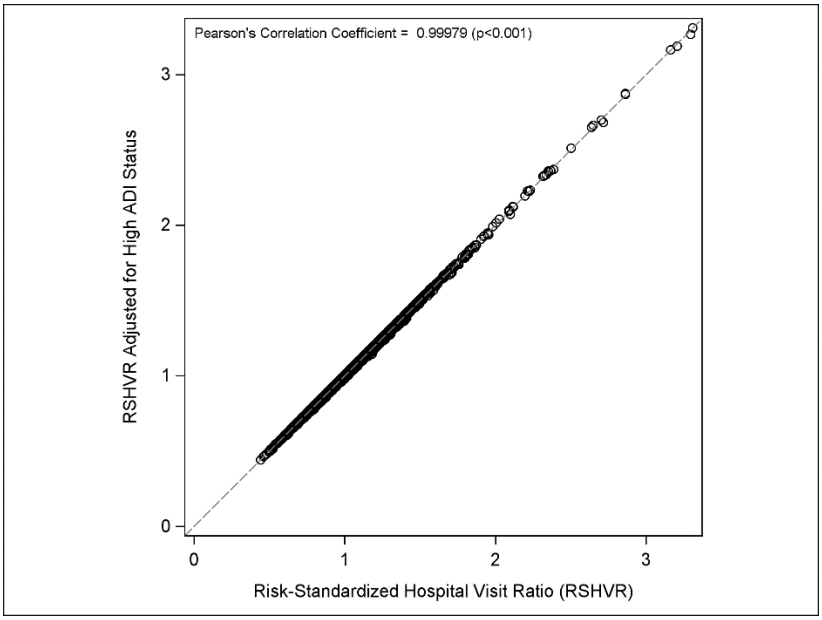


Figure 10: Spearman correlation between the facility proportion of patients with DE (top quartile) and measure scores (RSHVRs) (Current Submission)

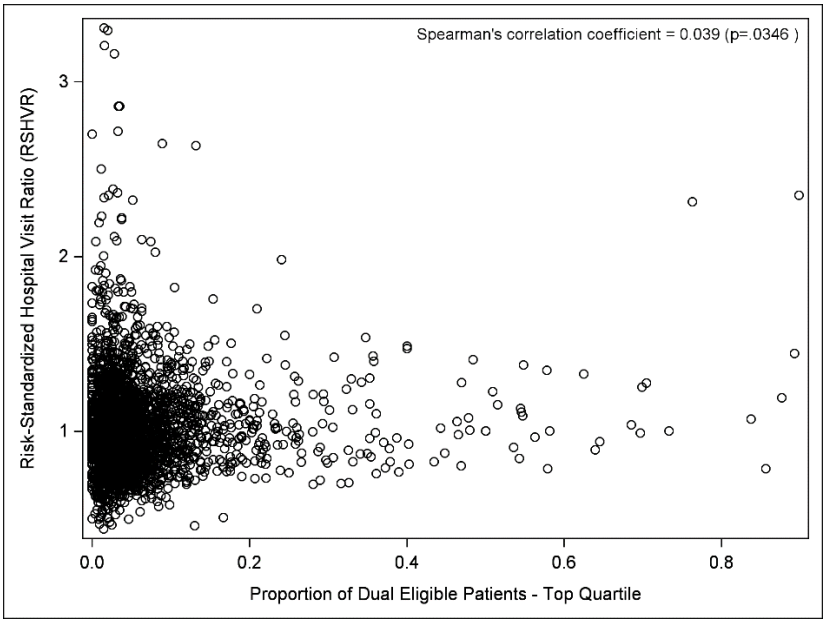
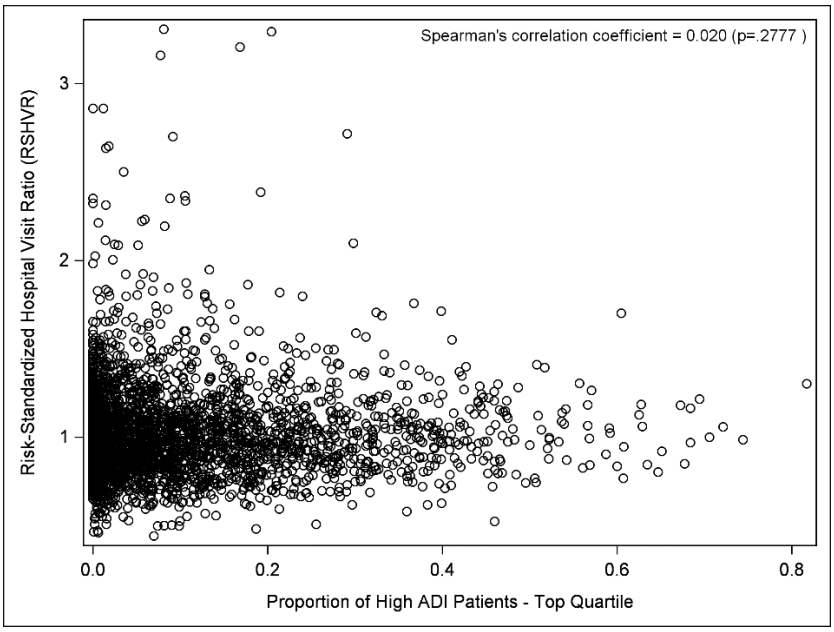


Figure 11: Spearman correlation between the facility proportion of patients with high ADI (top quartile) and measure scores (RSHVRs) (Current Submission)



Partnership for Quality Measurement
Powered by Battelle

Full Measure Submission to PQM

Instructions: You must complete all required fields (denoted by *) to submit your measure. You may save your progress as a draft prior to submitting your measure.

Some fields are required only if your measure is an electronic Clinical Quality Measure (eCQM), an initial (new) measure, or a maintenance measure. These are indicated at the beginning of the questions in brackets, e.g., *[For initial submissions only]*.

Measure Specifications

Note: *If you have changes to information submitted via the Intent to Submit, please edit the original content for the Full Measure Submission.*

Measure Name:

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.

If applicable, provide a rationale for why measured entities should report this measure with other measures to appropriately interpret results. *

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 them more visible to providers and patients. The measure score provides an assessment of quality that is publicly reported and informs quality improvement.

This measure is not a paired measure and does not require reporting with other measures to appropriately interpret results.

Provide a URL to a web page specific for this measure containing current detailed specifications, including code lists, risk model details, and supplemental materials. *

Do not enter a URL to a home page or to general information. If no URL is available, indicate "not available."

https://qualitynet.cms.gov/outpatient/measures/surgery/methodology

https://qualitynet.cms.gov/files/633b33c0a90aef00178477c3?filename=2022_OQR_MeasUpdt_Rpt_v1.0.pdf

---

[If the measure is an eCQM] If your measure is an electronic clinical quality measure (eCQM), please attach the zipped output from the Measure Authoring Tool (MAT). *

If you did not use the MAT, please contact PQM Support. Use the specification fields for the plain-language description of the specifications.

☐ MAT output attached
☐ MAT output not attached (explain)
If you select "MAT output not attached" a text box will open for you to provide an explanation.

Not applicable; this measure is not an eCQM.

Do you have a data dictionary, code table, or value sets (and risk model codes and coefficients, if applicable)? *

☐Yes
☐No

☒ Attached Excel or csv file -- attach file here if answered yes

Please put all information into one workbook. Excel formats are preferred (.xlsx or .csv).

☐If no, attest that all information will be provided in other fields in the submission.

Provide details needed to calculate the numerator. *

All information required to identify and calculate the cases from the target population (denominator) with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets. If your list of codes with descriptors is greater than will fit in this text box, you must attach an excel or csv file in the previous question. Please provide lists of individual codes with descriptors that exceed one page in an Excel or csv file in response to the field requesting the data dictionary, code table, or value sets.

Because this measure is an outcome measure, it does not have a traditional numerator. We use this section to describe the outcome.

The outcome for this measure 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://qualitynet.cms.gov/files/651b5fb570a30f001c388004?filename=2023_OQR_MeasureUpdatesRpt.pdf. The codes that define ED visits and observation stays are in the attached data dictionary, sheet "HOPD_Surgy__ED_Obs_Stay_Def"

Provide details needed to calculate the denominator. *

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. Please provide lists of individual codes with descriptors that exceed one page in an Excel or csv file in response to the field requesting the data dictionary, code table, or value sets.

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. The target population for this measure is Medicare Fee for Service (FFS) patients over age 65 who have had surgery performed in an HOPD during the performance period. We list the specific inclusion criteria below and note that this measure is procedure-based (not patient-based).

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.

Citations

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

---

Describe denominator exclusions. *

Briefly describe exclusions from the denominator cases, if any. Enter "None" if the measure does not have denominator exclusions.

The following surgeries are excluded from the denominator:

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 are 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.

Provide details needed to calculate denominator exclusions. *

Enter "None" if the measure does not have denominator exclusions. 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. If the lists of codes with descriptors exceeds one page in Word, then please provide these lists in an Excel or csv file in response to the field requesting the data dictionary, code table, or value sets.

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 are 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-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-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.

The 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.

Please select the most relevant type of score. *

☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g., average
☐ Count
☐ Rate/proportion
☐ Composite scale
☒ Other scoring method
Please specify (text box)
Ratio

Select the appropriate interpretation of the measure score. *
☐ Better quality = Higher score
☒ Better quality = Lower score
☐ Better quality = Score within a defined interval
☐ Passing score defines better quality
☐ N/A
Please specify (text box) For example, cost and efficiency measures

---

Diagram or describe the calculation of the measure score as an ordered sequence of steps. *

Identify the denominator, denominator exclusions, denominator exceptions, numerator,

numerator exclusions, time period of data collection, risk adjustment, and any other calculations.

Upload diagram if applicable (file types: PDF, visio, jpg, png)

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 HOPD's 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 2023 Hospital Visits after Hospital Outpatient Surgery Measure Update Report: https://qualitynet.cms.gov/files/651b5fb570a30f001c388004?filename=2023_OQR_MeasureUpdatesRpt.pdf

Below we provide the individual steps to calculate the measure score:

1. Identify surgeries meeting the inclusion criteria described in the denominator section above, and in Tab 1, "HOPD Surg Cohort," of the data dictionary.
2. Exclude procedures meeting any of the exclusion criteria described in the exclusion section above.
3. Identify a binary flag for an unplanned hospital visit within 7 days of index procedures as described above.
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 original measure development 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 most recent Hospital Visits after Hospital Outpatient Surgery Measure Update Report posted here: https://qualitynet.cms.gov/files/651b5fb570a30f001c388004?filename=2023_OQR_MeasureUpdatesRpt.pdf

---

Provide all information required to stratify the measure results, if necessary. *

Include 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. Please provide lists of individual codes with descriptors that exceed one page in an Excel or csv file in response to the field requesting the data dictionary, code table, or value sets.

This measure is currently stratified by dual eligibility (DE), reported confidentially to hospitals.

In the Calendar Year (CY) 2022 OPPS Proposed Rule, CMS described a plan to stratify reporting using two disparity methods, described below, in the HOPD setting and have identified the HOPD Surgery measure as one of six priority measures included in the Hospital Outpatient Quality Reporting (OQR) program for confidential disparity reporting stratified by patient DE.

The two stratification methods are:

1. **The Within-Facility Disparity Method**, highlights differences in outcomes for patient groups based on social risk factors within an HOPD and
2. **The Across-Facility Disparity Method**, illuminates variation in healthcare quality for patients with social risk factors across facilities.

The two methods are described in more detail below and visually shown in Figure 1. Details of the methodology can be found here: https://qualitynet.cms.gov/files/652fd45a8be3e0001c0b5141?filename=CY23_OP_32_35_36_DsprtySpecs.pdf

The Within-Facility Disparity Method reports differences in health outcomes between patient populations in the same facility. The goal of this method is to assess the difference in outcomes for two patients with the same condition and medical history but with different social risks. This method can answer the question: “Does a patient with a social risk factor experience similar health outcomes as a patient without that social risk factor when cared for at the same facility?”

The Across-Facility Disparity Method reports facility outcome rates for one patient population with a particular social risk factor across facilities. This method can answer the question: “How does the outcome rate for patients with a social risk factor at a specific facility compare to the outcome rate for patients with that social risk factor at an average facility?”

Please see the “Equity” section for results for the HOPD Surgery method when applying these two disparity methods.

Figure 1: Within- and Across- Facility Disparity Methods

Select the data sources for which you have tested and specified the measure. *

Select all that apply.

☒ Administrative Data
☒ Claims Data
☐ Electronic Health Records
☐ Other Electronic Clinical Data
☐ Paper Patient Medical Records
☐ Registries
☐ Standardized Patient Assessments
☐ Patient-Reported Data and/or Survey Data *(opens the questions noted below if selected)*
☐ Non-Medical Data
☐ Other Data Source
Please specify (text box)

If you selected Patient-Reported Data and/or Survey Data, you will see these questions:

Provide the survey, tool, questionnaire, or scale used as a data source for your measure.

☐ Available at measure-specific web page (provide the URL)
Please specify (text box)
☐ Attached

Please indicate the responder for your survey, tool, questionnaire, or scale.

☐ Patient
☐ Family or other caregiver
☐ Clinician

☐ Other
Please specify (text box)

Are proxy responses allowed?

☐ Yes
☐ No

If yes, please describe how. *
Required if checked yes above

For survey/patient-reported data, provide instructions for data collection and guidance on minimum response rate. Provide the data needed to calculate the response rates for reporting with performance measure results. *

Identify the specific data source or data collection instrument. *

For example, provide the name of the database, clinical registry, collection instrument, and describe how the measured entities will collect the data (e.g., the standard methods, modes, and languages of administration).

Medicare administrative claims and enrollment data

Indicate whether the measure has a minimum sample size to calculate the measure and provide any instructions needed for obtaining the sample and guidance on minimal sample size. *

There is no minimum sample size required to calculate the measure score. All facilities can benefit from receiving individual facility results. Programs using the measure should consider reliability when choosing sample sizes for measure use. For example, for the Hospital Outpatient Quality Reporting program, a facility must perform 30 qualifying procedures to receive a publicly reported measure score.

Importance

Attach a logic model and provide a description of the relationship between structures and processes and the desired outcome. *

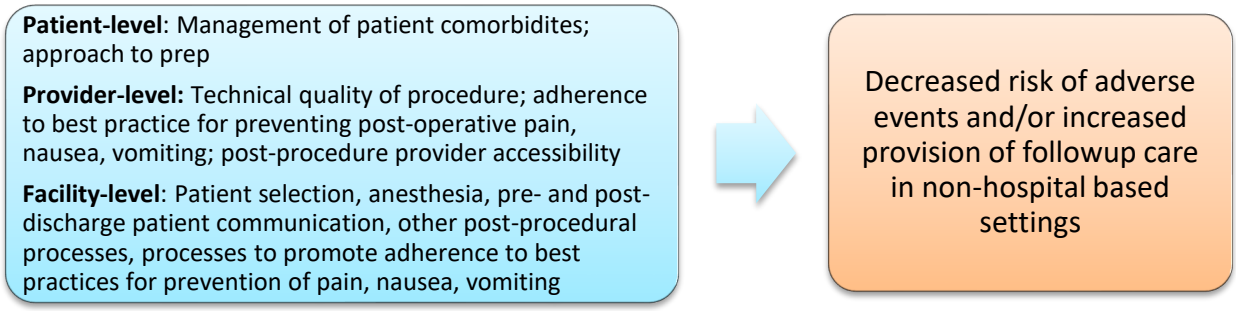
Briefly describe the steps between the health care structures and processes (e.g., interventions, or services) and the desired health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process, or outcome being measured.

Attachment (pdf, word)

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, outlined in Figure 1b, 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.

Figure 1b: Logic model for the HOPD Surgery measure showing the relationship between processes and outcomes.



Summarize evidence of measure importance from the literature, linking the structure/process/intermediate outcome to the desired health outcome. *

Please cite supporting evidence.

The outcome of unplanned hospital visits following outpatient same-day surgery is a widely accepted measure of outpatient surgical care quality. Differences across hospitals in risk-standardized, post-procedure unplanned hospital visits are likely related to quality of care rather than to pre-existing medical conditions or chance. Studies have consistently shown that post-operative complications and poorly controlled symptoms are the primary contributors to unexpected hospital visits following outpatient surgery (Desai et al., 2022). This measure provides the opportunity to improve the quality of care and to lower rates of adverse events leading to hospital visits after outpatient surgery.

A recent study, using Medicare FFS beneficiaries aged 65 years and older, across 4000+ HOPDs, showed nearly 8% of (overall) hospital outpatient surgeries were followed by an unplanned hospital visit within 7 days (Desai et al., 2022). Additionally, if a patient received the same procedure at a lower-quality versus higher-quality hospital, their risk of an unplanned hospital visit within 7 days would increase by 29%. (Desai et al., 2022).

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 and body system, 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 et al., 2019; De Oliveira

et al., 2015; Liu et al., 2018a; Liu et al., 2018b). For example, unadjusted 7-day outcome rates (of unplanned hospital visits) varied substantially by body system, ranging from a low of 4.4% for procedures of the ear to a high of 14.3% for male genital organ procedures, such as transurethral electrosurgical resection of prostate and laser vaporization of prostate (Desai et al., 2022).

Common causes of preventable return visits following outpatient surgery include surgical errors, post-operative pain, infection, nausea, urinary retention, and vomiting (Desai et al., 2022; 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 (De Oliveira et al., 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 Oliveira et al., 2015; Christian et al., 2019). In addition, clinical procedural factors can increase the risk, such as the type of anesthesia used, and longer operation time (Liu et al., 2018a; Mingus et al., 1997; Christian et al., 2019).

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) as well as routine multi-modal nausea and vomiting prophylaxis for all patients (Gan et al., 2020). 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).

Citations

A Proposed Rule by the Centers for Medicare & Medicaid Services on 07/31/2023. (2023, July 31). Federal Register; National Archives. https://www.govinfo.gov/content/pkg/FR-2023-07-31/pdf/2023-14768.pdf

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 R, Gordon DB, de Leon-Casasola OA, Rosenberg JM, Bickler S, Brennan T, Carter T, Cassidy CL, Chittenden EH, Degenhardt E, Griffith S, Manworren R, McCarberg B, Montgomery R, Murphy J, Perkal MF, Suresh S, Sluka K, Strassels S, Thirlby R, Viscusi E, Walco GA,

Warner L, Weisman SJ, Wu CL. Management of Postoperative Pain: A Clinical Practice Guideline From the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. J Pain. 2016 Feb;17(2):131-57.

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.

Desai MM, Zogg CK, Ranasinghe I, et al. Variation in Risk-standardized Rates and Causes of Unplanned Hospital Visits Within 7 Days of Hospital Outpatient Surgery. Ann Surg. 2022;276(6):e714-e720.

Gan TJ, Belani KG, Bergese S, Chung F, Diemunsch P, Habib AS, Jin Z, Kovac AL, Meyer TA, Urman RD, Apfel CC, Ayad S, Beagley L, Candiotti K, Englesakis M, Hedrick TL, Kranke P, Lee S, Lipman D, Minkowitz HS, Morton J, Philip BK. Fourth Consensus Guidelines for the Management of Postoperative Nausea and Vomiting. Anesth Analg. 2020 Aug;131(2):411-448.

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.

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

Schloemann DT, Sajda T, Ricciardi BF, Thirukumaran CP. Association of Total Knee Replacement Removal From the Inpatient-Only List With Outpatient Surgery Utilization and Outcomes in Medicare Patients. JAMA Netw Open. 2023;6(6):e2316769.

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.

[For initial endorsement] **If implemented, what is the measure's anticipated impact on important outcomes?** * *Please cite evidence to identify adverse events and costs avoided. Cite business case, if applicable.*

N/A-This is maintenance endorsement.

[*For maintenance review]* **Provide evidence of performance gap or measurement gap by providing performance scores on the measure as specified (current and over time) at the specified level of analysis.** *

Please include mean, standard deviation, minimum, maximum, interquartile range, and scores by deciles. Describe the data source including number of measured entities, number of patients, dates of data. If a sample, provide characteristics of the entities included. If performance scores are unavailable for the measure, please explain.

Previous 2020 Submission

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=3,874 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

Current 2023 Submission

There remains wide variation in measure scores among hospitals. Using updated data from CY 2022 (January 1, 2022-December 31, 2022) we found that among the 3,817 hospitals that performed at least one qualifying outpatient (HOPD) procedure, measure scores (risk-standardized hospital visit ratios, or RSHVRs) ranged from 0.44 (or 56% better than expected) to 3.31 (or more than 3 times worse than expected) (Table 1). Table 2 shows the measure score distribution by decile, and Figure 2a shows a histogram of facility measure scores.

We additionally provide evidence of variation by calculating the median odds ratio and calculating measure score outliers in the form of 95% confidence intervals, which are described below.

Measure score distribution

Table 1 shows the distribution of measure scores, the risk-standardized hospital visit ratio, or RSHVR, using the most recent testing data (2023 EM Dataset; January 1, 2022-December 31, 2022). A ratio of less than one indicates performance that is better than expected (better quality); a ratio of more than one indicates performance that is worse than expected (lower quality). HOPD Surgery RSHVRs range from 0.44 (or 56% better than expected) to 3.31, or more than 3 times worse than expected. The interquartile range (IQR; or the range between the 25th percentile and the 75th percentile) is 0.89 to 1.10.

Table 1 shows the distribution of measure scores in deciles. We note that the table provided by Battelle is not editable; the last row for this measure is the number of procedures, not the number of patients.

Table 1: Distribution of measure scores (risk-standardized hospital visit ratio, or RSHVR) using the most recent testing data (2023 EM Dataset; CY2022 data).

Characteristics	Value
Number of facilities	3,817
Mean RSHVR (standard deviation)	1.02 (0.24)
Minimum	0.44
25th percentile	0.89
50th percentile	0.99
75th percentile	1.10
Maximum	3.31

Table 2: HOPD Surgery Measure Score: Deciles of facility-level risk-standardized hospital visit ratio (RSHVR)*

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	1.02	0.44	0.70	0.82	0.89	0.94	0.98	1.00	1.05	1.10	1.20	1.53	3.31
Entities	3817	-	381	382	382	382	381	382	382	382	382	381	-
Total number of procedures	1,204,167	-	251,836	173,925	142,221	91,332	52,575	50,424	117,445	91,330	112,165	120,914	-

Median odds ratio

We provide further evidence of variation by calculating and interpreting the median odds ratio (MOR) (Merlo, et al., 2006). 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.

The median odds ratio for this updated analysis is 1.39. 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 1.39 indicates that a patient has a 39% 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.

Performance categories.

Finally, we provide evidence of variation by calculating statistical measure score outliers using 95% confidence intervals. Our results in Table 3 show that 450 facilities (or about 12%) of facilities with a measure score fall in the “better” or “worse” 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.

Table 3 shows the performance categories for facilities as measured by the HOPD Surgery score. A total of 220 facilities (5.8%) performed "Better than Expected," 2,427 facilities (63.6%) performed "No Different than Expected," and the remaining 229 facilities (6.0%) performed "Worse than Expected." There were 941 facilities (24.7%) with too few cases (<30).

Table 3: HOPD Surgery Measure score performance categories using CY2022 data (EM 2023 Dataset)

Performance Category	HOPDs	
	Number of facilities	% distribution
Better than Expected	220	5.8%
No different than Expected	2,427	63.6%
Worse than Expected	229	6.0%
Number of Cases Too Small (<30)	941	24.7%

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: https://qualitynet.cms.gov/files/5d0d3a7e764be766b0104644?filename=2016HOPDSurgeryTechReport.pdf.

Citation:

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.

[For initial endorsement] **Please explain why existing measures/quality improvement programs are insufficient for addressing this health care need.** *

Not applicable. This is a maintenance measure.

Provide evidence the target population (e.g., patients) values the measured outcome, process, or structure, and finds it meaningful. *
Please describe how and from whom you obtained input.

A hospital visit following same-day surgery is an unexpected and potentially preventable outcome for patients scheduled for elective 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. 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 (now *Care Compare*). Currently, there are no other publicly available quality reports of HOPDs that perform same-day surgery, and the recent migration of additional procedures to the outpatient setting (e.g., THA/TKA) (Burnett et al., 2023) underscores the measurement gap that would exist without this measure. 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.

During measure development, we asked our Technical Expert Panel (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. TEP members indicated their agreement (on a six-point scale) 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." 12 of the 13 indicated they moderately or strongly agreed. Two TEP members did not respond to the TEP survey.

Citation

Burnett, R. A., Barrack, T. N., Terhune, E. B., Della Valle, C. J., Shah, R. P., & Courtney, P. M. (2023). Over Half of All Medicare Total Knee Arthroplasty Patients Are Now Classified as an Outpatient-Three-Year Impact of the Removal From the Inpatient-Only List. The Journal of arthroplasty, 38(6), 992–997. https://doi.org/10.1016/j.arth.2022.12.029

Feasibility

[*For Initial Endorsement]* **Describe the feasibility assessment conducted showing you considered the people, tools, tasks, and technologies necessary to implement this measure**. If an eCQM, please also attach your completed feasibility scorecard. *

Please explain and upload the feasibility scorecard if applicable.

Describe how the feasibility assessment informed the final measure specifications, indicating any decisions made to adjust the measure in response to feasibility assessment. *

Because this is a claims-based measure the is no burden on the facility; rates are automatically calculated by CMS based on claims data submitted by facilities for payment.

Indicate whether your measure or any of its components are proprietary, with or without fees. *

☐ Proprietary measure or components (e.g., risk model, codes)
☐ Proprietary measure or components with fees
☒ Not a proprietary measure and no proprietary components

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). *

Required if checked in previous question that this is a proprietary measure or components (with or without fees)

There are no fees, licensing, or other requirements to use this measure as specified.

Scientific Acceptability

Describe the data or sample used for testing (include dates, source). If you used multiple data sources for different aspects of testing (e.g., reliability, validity, risk adjustment), identify how the data or sample are different for each aspect of testing. *

Previous 2020 Submission

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 (DE 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.

For updated measure testing provided in the 2020 (prior) 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 (DE 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.

Current 2023 Submission

For this 2023 endorsement maintenance submission we used paid, final action Medicare claims from January 1, 2022 to December 31, 2022 to identify procedures performed in the outpatient setting at Hospital Outpatient Departments (HOPDs), and subsequent hospital visits.

We use Medicare FFS claims to identify surgeries performed in the outpatient setting and subsequent hospital visits, as well as CMS enrollment and demographic data. Patient history is also assessed using claims data collected in the 12 months prior to the eligible same-day surgery. We identify outpatient surgeries using Medicare's list of covered ASC procedures. CMS reviews and updates this list of surgeries annually. The process includes a transparent public comment submission and review process for addition and/or removal of procedures codes. The lists are posted at: https://www.cms.gov/Medicare/Medicare-Fee-for-ServicePayment/ASCPayment/11_Addenda_Updates.html (refer to Addendum AA of the respective link). Procedures listed on Medicare's list of covered ASC procedures are defined using HCPCS and CPT® codes.

The measure attributes surgeries to an HOPD if a Part B physician claim is present and the claim can be linked to Medicare outpatient or inpatient institutional data. We first identify physician claims as Outpatient Hospital Department or Physician Office by the Line Place of Service Code in the Part B Carrier claims file. Place of Service coding is used to specify the entity where service(s) were rendered. We then link the physician claims to a hospital outpatient claim with the surgery indicated to identify the HOPD where the surgery took place. Physician claims with no match to a hospital outpatient claim are then matched to hospital inpatient claims with an inpatient admission date within zero to three days after the date of surgery, to capture

surgical procedures billed per the CMS 3-day payment window policy. The HOPD of the admitting hospital is where the case is attributed.

Please provide descriptive characteristics of measured entities included in the analysis (e.g., size, location, type). *

If you used a sample, describe how you selected measured entities for inclusion in the sample.

The number of measured entities (HOPDs) varies by testing type. Please see Table 5 in the section below for details.

Identify the number and descriptive characteristics (e.g., age, sex, race, diagnosis), of the level(s) of analysis, for example, patient, encounter or episode, separated by level of analysis and data source. *

If you used a sample, describe how you selected the patients for inclusion in the sample. If there is a minimum case count used for testing, you must reflect that minimum in the specifications.

The number of patients varies by testing type. Please see Table 5 in the section below for details.

If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), please identify how the data or sample are different for each aspect of testing. *

Previous 2020 Submission

Table 4 outlines the datasets used in each analysis.

Table 4: Dataset Descriptions

Dataset	Description of Dataset	Use and Section in the Testing Attachment
Dataset #1: Initial Development Dataset **Dataset #1a: Development dataset** **Dataset #1b: Validation 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. 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.	Validity testing (face validity) Identification and selection of risk-adjustment variables 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	Reliability Data Element & Measure Score Validity Testing of Measure Exclusion Selection of Social Risk Factors Meaningful Differences Predictive ability Statistical model discrimination statistics

Current 2023 Submission

Table 5 outlines the dataset used in this updated endorsement maintenance submission. We used Medicare Fee-For-Service claims from January 1, 2022 through December 31, 2022. This dataset included 1,204,167 procedures from 3,917 facilities. There were 2,876 facilities with at least 30 procedures (the public reporting threshold). Please see Table 3 for additional details.

Table 5: Dataset Descriptions

Dataset Name	Description of Dataset	Types of analyses
2023 Endorsement Maintenance (EM) Dataset	Medicare-Fee-for Service claims from January 1, 2022 through December 31, 2022. Number of procedures: 1,204,167 Number of facilities: 3,817* Number of facilities with >= 30 procedures: 2,876 Mean age (SD): 74.8 % Female: 50.1	Reliability Data Element & Measure Score Validity Testing of Measure Exclusion Selection of Social Risk Factors Meaningful Differences Predictive ability Statistical model discrimination statistics

---

Reliability

Select the level of reliability testing conducted. *
Please select all that apply.

☐ Patient or Encounter-Level (e.g., inter-abstractor reliability)
☒ Accountable Entity-Level (e.g., signal-to-noise analysis)

For each level of reliability testing conducted, 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? Provide the statistical analysis used.*

Previous 2020 Submission

Measure Score Reliability

We provide facility-level measure score reliability using the signal-to-noise method, using the formula presented by Adams and colleagues [Yu et al, 2013; Adams et al., 2010]. 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

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

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

Current Submission

We calculated facility-level (signal-to-noise) reliability using the same method described above from the prior submission.

Provide the statistical results from reliability testing. *

Previous 2020 Submission

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.

Current Submission

Table 6 shows the results from updated testing, using the 2023 EM dataset.

The median facility-level reliability (signal-to-noise reliability) for all facilities (N=3,817) was 0.853 (IQR 0.521-0.939); the median facility-level reliability for facilities with more than 30 procedures (n=2876) was 0.908 (IQR 0.808-0.953). The 2,876 facilities represent 1,194,500 procedures or 99.2% of the total 1,204,167 procedures. In Table 6 we present signal to noise reliability in deciles, as requested by Battelle, for all facilities. In table 6A (see attachment) we also provide signal to noise reliability for facilities with at least 30 procedures which is the public reporting threshold.

Table 6: Distribution of signal-to-noise reliability for all facilities

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.71	0.03	0.09	0.28	0.52	0.71	0.82	0.88	0.92	0.94	0.96	0.98	0.99
Entities	3,817	-	371	395	375	385	384	380	383	381	382	381	-
Total Number of Procedures	1,204,167	-	984	4,486	11,707	26,341	48,553	77,212	116,505	164,913	247,841	505,625	-

Table 6A. Deciles of signal to noise reliability for facilities with at least 30 procedures

	Overall	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Reliability	0.86	0.60	0.72	0.81	0.86	0.89	0.92	0.94	0.95	0.96	0.98
Entities	2,876	284	292	285	289	290	289	284	288	288	287
Total Number of procedures	1,194,500	11,735	21,279	33,119	48,786	68,508	91,608	116,928	158,407	22,1304	42,2826

Interpretation of Reliability Results

Provide your interpretation of the results in terms of demonstrating reliability. How do the results support an inference of reliability for the measure?

Previous 2020 submission

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).

Current submission

The median signal-to-noise reliability for the HOPD surgery measure is sufficiently high (above the CBE threshold of 0.6) for all facilities (0.85) and for facilities with at least 30 procedures (0.89) (the public reporting threshold).

Validity

Select the level of validity testing conducted. *

Please select all that apply.

☐ Patient or Encounter-Level (e.g., inter-abstractor reliability)
☒ Accountable Entity Level (e.g., signal-to-noise analysis)

Select the type of validity testing conducted. *
Please select all that apply.

☒ Empirical validity testing
☒ Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., the score is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance).

For each level of testing conducted, describe the method of validity testing and what it tests. *

Describe the steps, do not just name a method and what you tested (e.g., accuracy of data elements compared with authoritative source, relationship to another measure as expected). What statistical analysis did you use? Include analysis of missing data and any exclusions.

Previous 2020 Submission

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 scores 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 (Brooke et al., 2012).

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 (Steiner et al., 2014). 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 2a). 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 2b). 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 below. 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" (Krumholz et al., 2006; NQF, 2011).

Citations

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.

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.

National Voluntary Consensus Standards for Patient Outcomes 2009 A CONSENSUS REPORT Patient Outcomes. (2011). https://www.qualityforum.org/Publications/2011/07/National_Voluntary_Consensus_Standards_for_Patient_Outcomes_2009.aspx

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.

Current 2023 Submission

For our updated 2023 submission we provide additional validity testing in the form of the association of HOPD measure scores with procedural volume. We examined this association by plotting HOPD measure scores (RSHVRs) within quintiles of facility-level procedural volume.

Validity Testing Result
Provide the statistical results from validity testing for each level of validity testing conducted.

Previous 2020 Submission

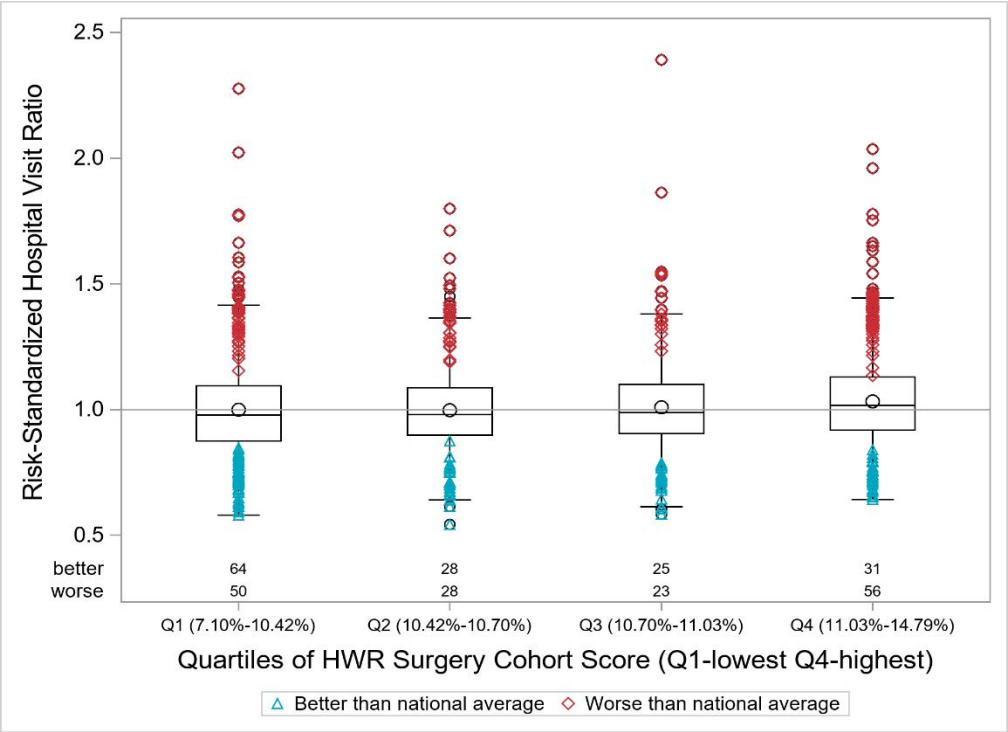
Empiric Validity Testing

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 2a). (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 2a and Figure 2b). In Figure 2a 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 below). In Figure 2b 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 2a: Relationship between HOPD Surgery measure score and HWR (surgery cohort) measure score



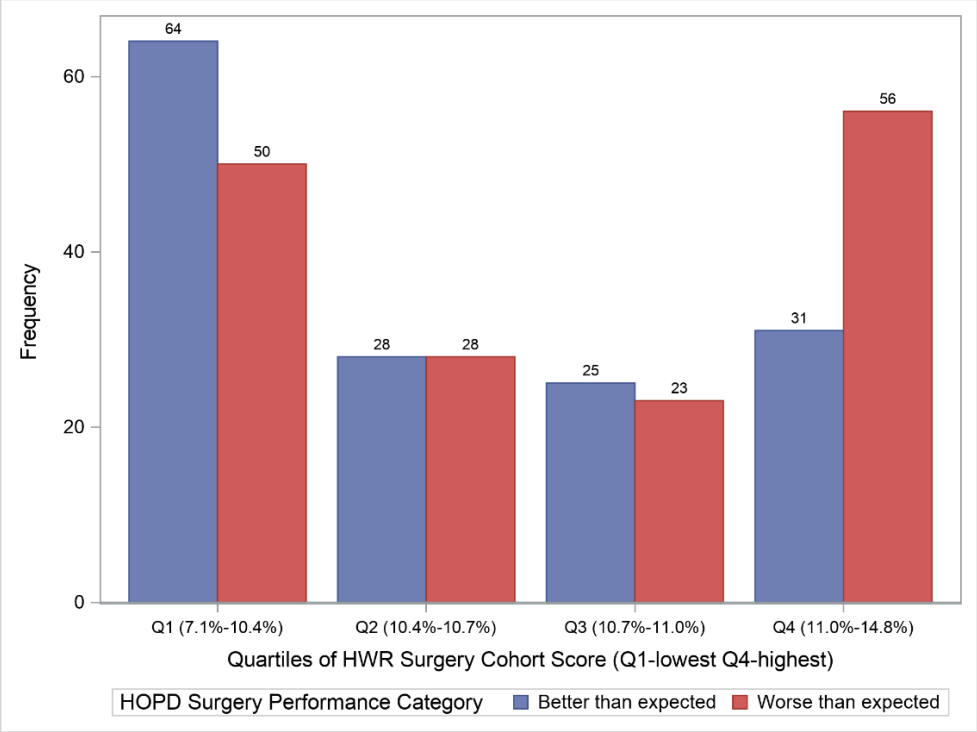
The results show a trend toward better performance on the HOPD Surgery measure with better performance on the comparator measure (HWR, surgery cohort). As shown in Figure 2a, better performance on the HOPD Surgery measure shows a small positive trend with better performance across quartiles of performance 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 2b). 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 2b) 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 2b: 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

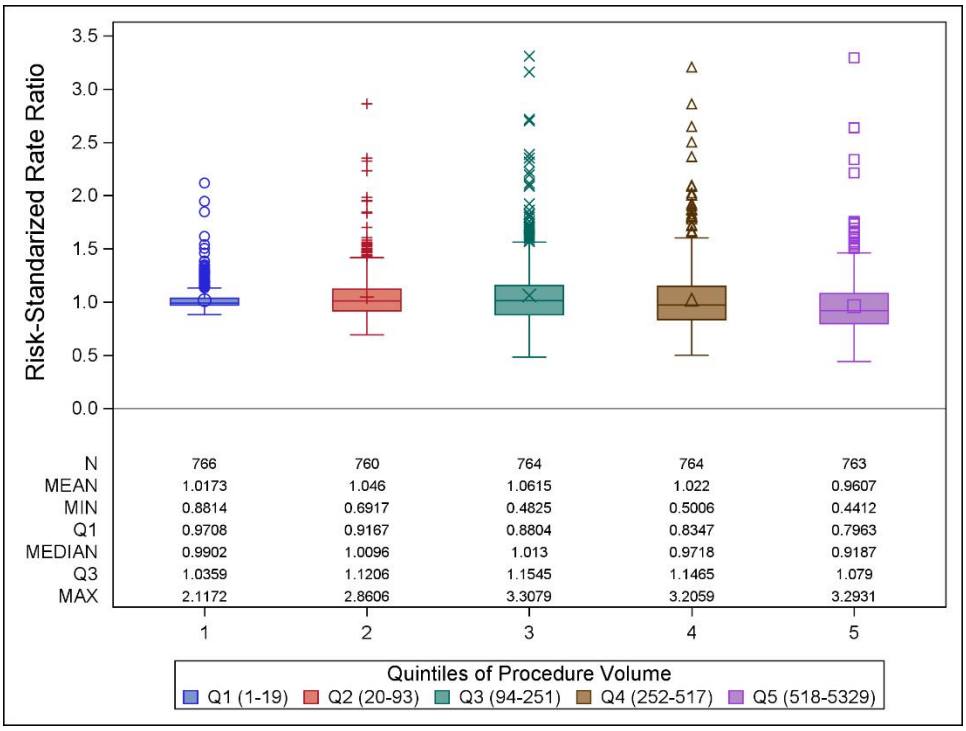
Rating	# (%) of Responses
1 (Strongly disagree)	0 (0)
2 (Moderately disagree)	0 (0)
3 (Somewhat disagree)	0 (0)
4 (Somewhat agree)	1 (7.7)
5 (Moderately agree)	8 (61.5)
6 (Strongly agree)	4 (30.8)

Current Submission

Association of HOPD Surgery measure scores with volume

Because there is evidence that outcomes for some surgical procedures are associated with higher volume (Levaillant et al., 2021; Brodeur et al., 2022), we examined the relationship between the HOPD surgery measure scores with facility-level procedural volume. Figure 3 shows that the there is an overall trend toward improved outcomes (lower RHSVRs) with increasing facility volume. Median RHSVRs decline across quintiles, in particular across the three highest-volume quintiles. For example, mean, median and IQR ranges of measure scores (RSHVRs) decline with each quintile when comparing the fourth and fifth quintiles of procedural volume to the third quintile.

Figure 3: Distribution of RSHVR within deciles of facility procedural volume.



Citations

Brodeur PG, Kim KW, Modest JM, Cohen EM, Gil JA, Cruz AI. Surgeon and Facility Volume are Associated With Postoperative Complications After Total Knee Arthroplasty. *Arthroplasty Today*. 2022;14:223-230.e1. doi:https://doi.org/10.1016/j.artd.2021.11.017

Levaillant M, Marcilly R, Levaillant L, et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. *BMC Medical Research Methodology*. 2021;21:204. doi:https://doi.org/10.1186/s12874-021-01396-6

Provide your interpretation of the results in terms of demonstrating validity. *
How do the results support an inference of validity for the measure?

Previous 2020 Submission

The combination of empiric and face validity results support the validity of this HOPD surgery measure. First, 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 "worse" than "better" HOPD Surgery outliers in the worst performing HWR quartile but the same number of "better" and "worse" outliers in the first (better) performing HWR quartile.

Current Submission

Our results from the current and past submissions provide additional evidence for the validity of the HOPD Surgery measure. With this current submission we provide additional evidence for measure score validity in the form of a volume-outcome relationship: as facility volume increases, we found that there is a trend toward lower (better) measure scores. These results are supported by literature that shows a volume/outcome relationship for some but not all procedures.

Finally, TEP survey results show high agreement (with 92 percent or 12/13 respondents "moderately or "strongly" agreeing) regarding the ability of the HOPD Surgery measure to distinguish between higher- and lower-quality facilities.

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

Risk Adjustment

Check all methods used to address risk factors *

☒ Statistical risk model with risk factors
Specify number of risk factors (27)
☐ Stratification by risk category
Specify number of categories (text box)
☐ Other
Specify other (text box)
☐ No risk adjustment or stratification.
If select no, this question appears
Is the measure an outcome or resource measure?
☐ Yes
☐ No

IF you select yes this question appears: **If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate there is no need to control for differences in patient characteristics (i.e., case mix) to achieve fair comparisons across measured entities.** *

The following questions are shown and required if the user selects Statistical risk model with risk factors, Stratification by risk category or Other above:

Attach a conceptual model that illustrates the pathway between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured outcome. Please explain the rationale for the model. *

Consider age, gender, race/ethnicity, urbanicity/rurality, Medicare/Medicaid dual eligibility status, indices of social vulnerability (e.g., Centers for Disease Control and Prevention Social Vulnerability Index), and markers of functional risk in the conceptual model. If social and/or functional risk factors are not available but are included in the conceptual model, consider potential bias in the risk model, and describe its direction and magnitude. Address the validity of the measure in light of this bias.

Attachments (word, pdf)

Previous 2020 submission

We developed and used the conceptual framework described below to identify potential social risk factors. Limited social risk factor data are available at this time, however, on Medicare beneficiaries (Department of Health and Human Services, 2016). We analyzed two well-studied social risk factors that could best be operationalized in data, outlined below. We note that this measure already adjusts for age and gender.

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 (Department of Health and Human Services, 2016)

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 (Bonito, 2008). 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

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.

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.

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 (NQF, 2015).

Since the measure was endorsed, we have updated the measure in response to feedback from stakeholders (as discussed in 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 (Mioton et al., 2014; Bhattacharyya, 2015; Al-Qurayshi et al., 2016; Dallas et al., 2017; Trivedi et al., 2014; Jha et al., 2011; Reames et al., 2014) 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) (HHS, 2016; NASEM, 2016).

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 (Mioton et al., 2014; Bhattacharyya et al., 2015).

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 (Al-Qurayshi et al., 2016). In one 2017 study, researchers 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 (Dallas et al., 2017).

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 (Trivedi et al., HHS, 2016). 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 (Jha et al., 2011; Reames et al., 2014). 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 (Al-Qurayshi et al., 2016).

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 (HHS, 2016). This causal pathway should be largely accounted for by current clinical risk-adjustment.

The social risk variables that we examined were:
- Dual-eligible status
- AHRQ-validated SES Index score

Citations

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.

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

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.

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.

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.

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.

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.

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.

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

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.

Current 2023 Submission

Conceptual model

For this 2023 submission we have reviewed our conceptual model related to social risk factors and have determined that the original conceptual model remains valid. With an updated focused literature search we found additional evidence for the impact of social risk factors on outcomes for patients undergoing surgery. For example, a 2019 study found that while patients with low income undergoing colectomy had higher rates of surgical-site infections compared with higher-income patients, there was no difference in surgical-site infection rates based on income for patients undergoing hysterectomy (Qi et al., 2019). A 2023 study in cancer patients undergoing surgery found that patients with psychosocial risk factors were more likely to experience complications following surgery (Leeds et al., 2019). Finally, a 2021 study found that for some procedures, people living in counties with high social vulnerability (SVI) were more likely to experience complications compared with patients who live in low SVI counties (Diaz et al., 2021).

Social risk factors analyzed

With this updated submission we have replaced analyses that previously used the AHRQ SES variable, described above, with the validated Area Deprivation Index (ADI) (Forefront Group, 2023). We made this change to align with other CMS work on social risk factors that now uses the ADI. We describe the ADI variable below.

Area Deprivation index (ADI): The ADI, initially developed by Health Resources & Services

Administration (HRSA), is based 17 measures across four domains: income, education, employment, and housing quality (Kind et al., 2018; Singh, 2003).

The 17 components are listed below:

- Population aged ≥ 25 y with < 9 y of education, %
- Population aged ≥ 25 y with at least a high school diploma, %
- Employed persons aged ≥ 16 y in white-collar occupations, %
- Median family income, $
- Income disparity
- Median home value, $
- Median gross rent, $
- Median monthly mortgage, $
- Owner-occupied housing units, % (home ownership rate)
- Civilian labor force population aged ≥16 y unemployed, % (unemployment rate)
- Families below poverty level, %
- Population below 150% of the poverty threshold, %
- Single-parent households with children aged < 18 y, %
- Households without a motor vehicle, %
- Households without a telephone, %
- Occupied housing units without complete plumbing, % (log)
- Households with more than 1 person per room, % (crowding)

ADI scores were derived using beneficiary's 9-digit ZIP Code of residence, which is obtained from the Master Beneficiary Summary File, and is linked to 2017-2021 US Census/American Community Survey (ACS) data. In accordance with the ADI developers' methodology, an ADI score is calculated for the census block group corresponding to the beneficiary's 9-digit ZIP Code using 17 weighted Census indicators. Raw ADI scores were then transformed into a national percentile ranking ranging from 1 to 100, with lower scores indicating lower levels of disadvantage and higher scores indicating higher levels of disadvantage. Percentile thresholds established by the ADI developers were then applied to ADI percentile to dichotomize neighborhoods into more disadvantaged (high ADI areas=ranking equal to or greater than 85) or less disadvantaged areas (Low ADI areas= ranking of less than 85).

Citations

Diaz, A., Hyer, J. M., Barmash, E., Azap, R., Paredes, A. Z., & Pawlik, T. M. (2021). County-level Social Vulnerability is Associated With Worse Surgical Outcomes Especially Among Minority Patients. *Annals of Surgery*, *274*(6), 881–891. https://doi.org/10.1097/SLA.0000000000004691

Kind AJH, Buckingham W. Making Neighborhood Disadvantage Metrics Accessible: The Neighborhood Atlas. *New England Journal of Medicine*, 2018. 378: 2456-2458. DOI: 10.1056/NEJMp1802313. PMCID: PMC6051533. AND University of Wisconsin School of Medicine Public Health. 2023 Area Deprivation Index v4.0. Downloaded from https://www.neighborhoodatlas.medicine.wisc.edu/.

Leeds, I. L., Meyers, P. M., Zachary Obinna Enumah, He, J., Burkhart, R. A., Haut, E. R., Efron, J. E., & Johnston, F. M. (2019). Psychosocial Risks are Independently Associated with Cancer

Surgery Outcomes in Medically Comorbid Patients. *Annals of Surgical Oncology*, *26*(4), 936–944. https://doi.org/10.1245/s10434-018-07136-3

Singh, G. K. (2003). Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. *American Journal of Public Health*, *93*(7), 1137–1143. https://doi.org/10.2105/ajph.93.7.1137

The Area Deprivation Index Is The Most Scientifically Validated Social Exposome Tool Available For Policies Advancing Health Equity. (2023). *Forefront Group*. https://doi.org/10.1377/forefront.20230714.676093

Qi AC, Peacock K, Luke AA, Barker A, Olsen MA, Joynt Maddox KE. Associations Between Social Risk Factors and Surgical Site Infections After Colectomy and Abdominal Hysterectomy. JAMA Netw Open. 2019;2(10):e1912339. doi:10.1001/jamanetworkopen.2019.12339

Provide descriptive statistics on the distribution across the measured entities of the risk variables identified in the conceptual model. *

Table 7 shows the distribution of social risk factors identified in the conceptual model for the HOPD Surgery measure. The facility median proportion of patients with the DE and ADI variables is 3.4% and 6.3%, respectively.

We note that this measure is also reported to hospitals confidentially, stratified by dual eligibility.

Table 7: Distribution of Dual Eligible (DE) and Area Deprivation Index (ADI) variables among HOPDs with procedures captured by the HOPD Surgery measure (Dataset: EM 2023)

Social risk variable	Min (%)	Min (N)	Median (%)	Median (N)	IQR (%)	IQR (N)	Max (%)	Max (N)
DE (Yes)	0%	0	3.43%	6	1.26% - 7.57%	1-18	100%	489
ADI (ADI >= 85)	0%	0	6.31%	7	0.71% - 18.67%	1-30	100%	529

If using statistical risk models, provide detailed risk model specifications (query or algorithm), including the risk model method, risk factor data sources, and equations. Please attach an excel file providing the risk factors, coefficients, codes with descriptors, and definitions. *

Attachment (excel)

Previous 2020 Submission

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 (Normand et al., 2007; Krumholz et al. 2006; NQF, 2019).

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, which we assume follows a normal distribution with mean and variance, the between-facility variance component. The following equations define the HGLM:

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

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

$$i=1...I; j=1...n_i$$

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 = Zij = (Z1ij , Z2ij ,Zpij) is a set of p patient-specific covariates derived from the data; and I denotes the total number of HOPDs; and ni the number of surgeries performed at HOPD i. The facility-specific intercept of the i-th HOPD, defined above, comprises, the adjusted average intercept over all HOPDs in the sample, and, the facility specific intercept deviation from. A point estimate of, 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 and clinical comorbidities, in addition to 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 the attached Excel document. 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).

Citations

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.

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

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

Current Submission

Updated model risk variables (using data from January 1, 2022 through December 31, 2022) and their frequencies as well as odd ratios are in the attached Excel spreadsheet.

Detail the statistical results of the analysis used to test and select risk factors for inclusion in or exclusion from the risk model/stratification. *

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)
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
Surgical Body System:
Miscellaneous diagnostic and therapeutic procedures
Cardiovascular
Digestive
Ear
Endocrine
Female Genitalia
Hemic-Lymphatic
Skin & Breast
Male Genitalia
Musculoskeletal
Nervous
Nose-Throat-Pharynx
Respiratory
Urinary

Provide the approach and results of calibration and discrimination testing. Describe any over- or under-prediction of the model for important subgroups. Please attach results of calibration and discrimination testing. *

Attachment (pdf, jpg, png)

CORE's measures undergo an annual measure reevaluation process, which ensures that the risk-standardized models are continually assessed and remain valid, given possible changes in clinical practice and coding standards over time. Modifications made to measure cohorts, risk models, and outcomes are informed by a review of the most recent literature related to measure conditions or outcomes, feedback from various stakeholders, and empirical analyses, including assessment of coding trends that reveal shifts in clinical practice or billing patterns. Input is solicited from a workgroup composed of up to 20 clinical and measure experts, inclusive of internal and external consultants and subcontractors.

To assess model performance, we computed three summary statistics for the HOPD Surgery measure: two discrimination statistics (the C-statistic, predictive ability) and one calibration statistic (overfitting) (Harrell et al, 2001). In addition, we provide risk-decile plots.

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.

Previous 2020 Submission
C-statistic: 0.684
The c-statistic of 0.684 indicate good model discrimination.

Current Submission
C-statistic: 0.693
The c-statistic of 0.694 indicates continued 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.

(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.

Previous 2020 Submission

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

Current Submission
Predictive Ability, % (lowest decile - highest decile): 1.74-16.01
The model continues to show a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics (from original measure development)

(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. Our results, shown below, 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.

CORE notes that after initial measure development we do not re-test our risk models for overfitting using a dataset that is external to the testing sample. In our risk models, coefficients are updated each time the measure is calculated; we refit the model with new data each time the measure is calculated. Therefore, random statistical fluctuations in model coefficients across repeated reporting cycles are part of the overall random error in the facility performance estimates. CORE believes that this approach is not a validity issue for this type of model, unlike the case of a static risk model.

2010 Development Sample results:

Calibration: (0,1)

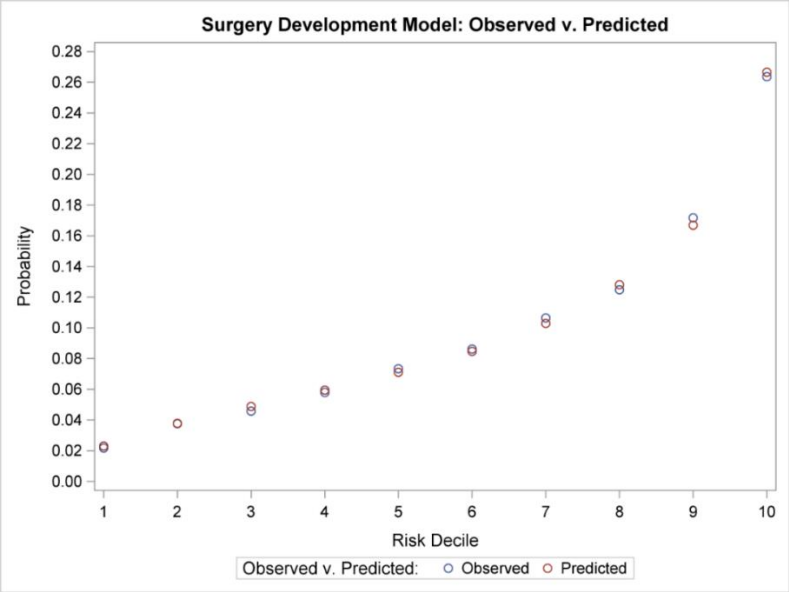
2010 Validation Sample results:
Calibration: (-0.05, 0.96)

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 below indicates good discrimination of the model and good predictive ability.

Previous 2020 Submission

Figure 4: Risk Decile Plot



Current Submission

We provide updated risk decile plots for all patients, and for patients with DE and high ADI, below.

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which continue to show good calibration of the model. The risk decile plot indicates continued good discrimination of the model and good predictive ability, for all patients, and for patients with DE and high ADI, separately.

Figure 5: Risk decile plot: all patients

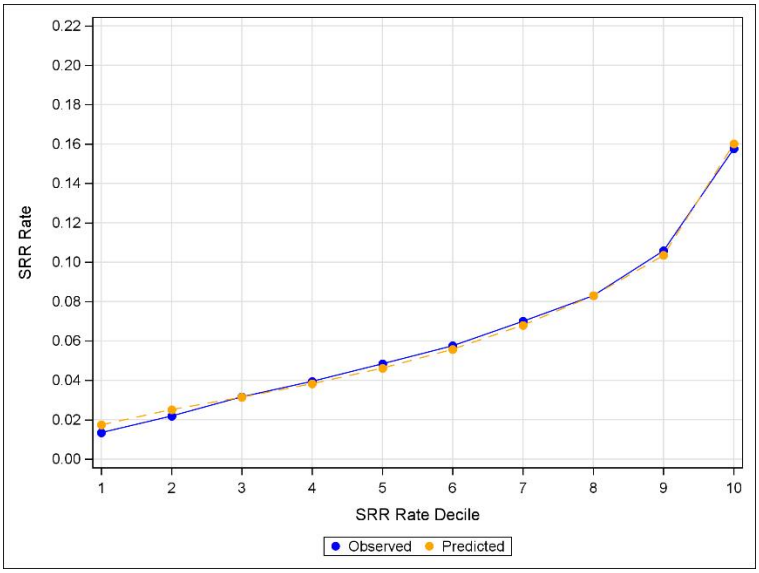


Figure 6: Risk decile plot: DE patients only

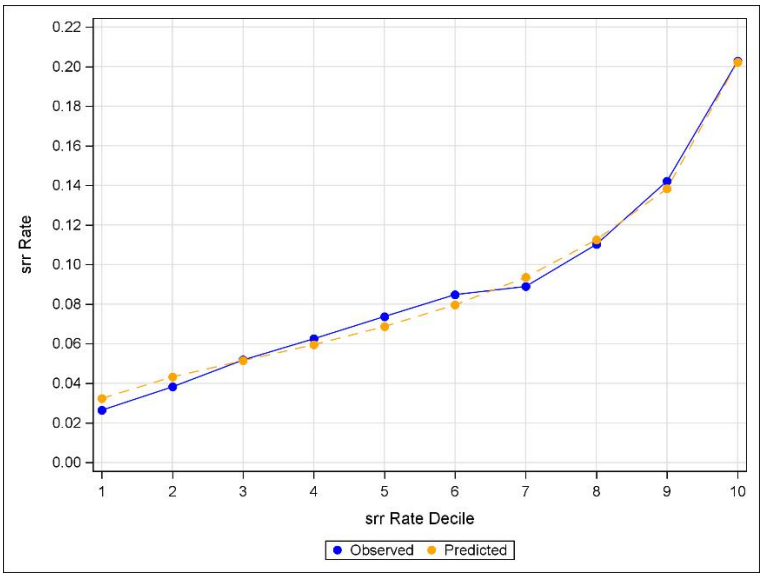
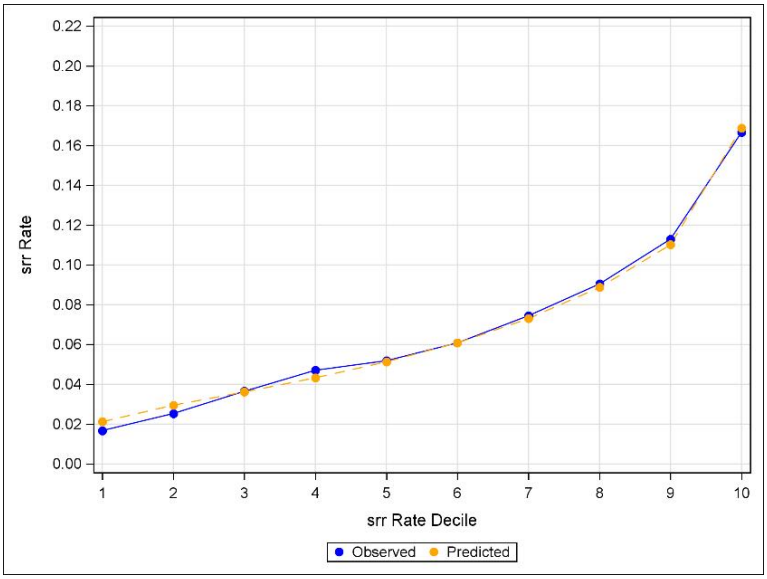


Figure 7: Risk decile plot: Patients with high ADI only



Citation

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.

Equity

Describe how this measure contributes to efforts to advance health equity (*optional*).
Provide a description of your methodology and approach to empirical testing of differences in performance scores across multiple socio-contextual variables (e.g., race, ethnicity, urbanicity/rurality, socio-economic status, gender, gender identity, sexual orientation, age). Provide an interpretation of the results, including interpretation of any identified differences and consideration of negative impact or unintended consequences on subgroups.

At the patient level, we know that patients with social risk factors (present in our conceptual model) may have higher unadjusted outcomes (hospital visit rates) following outpatient surgery, but differences vary depending on the social risk factor. For example, using CY2022 data (Jan 1, 2022-Dec 30, 2022) we found that patients with dual eligibility (DE) have an unadjusted hospital visit rate of 8.8%, compared with 6.2% for patients without DE. In contrast, however, patients with high ADI have visit rates that are only slightly higher than patients with Low ADI (6.8% vs. 6.3% respectively).

Measure Stratification

Due to these observed disparities in outcomes and the desire to shed light on and improve outcomes for all patients, CORE has developed for CMS (as described in the “stratification” section of this CBE submission) a disparities stratification methodology. Please see the “stratification” section for more information about the methodology; here we provide the most recent results for both the within-facility disparities method (which compares care within a hospital, comparing their DE and non-DE patients) and the across facility method (which compares facility-level outcomes for DE patients to the national average for all DE patients).

The full methodology and results are available in more detail in this report:

https://qualitynet.cms.gov/files/652fd45a8be3e0001c0b5141?filename=CY23_OP_32_35_36_DsprtySpecs.pdf

For the within-hospital approach, we found that more hospitals have worse outcomes for their DE patients compared with their non-DE patients. Using data from January 1, 2022-December 31, 2022, we characterized performance using the within-facility approach at the facility level, using rate difference cutoffs of: “better” outcomes as rate differences of less than -1%; “worse” outcomes as rate differences of >1%, and “similar” outcomes for rate differences between -1% and +1%. Using these categories, we found that 160 hospitals were characterized as “better,” 326 hospitals were characterized as “worse” and the remaining 167 were characterized as “similar” for their outcomes for DE patients compared with non-DE patients.

For the across-facility disparities approach, we found that slightly more hospitals perform better than the national rate, compared with worse than the national rate: 152 hospitals had outcomes for DE patients that were better than the national rate, compared with 106 that were worse, and 207 that were no different than the national rate. However, 2,676 of 3,141 eligible hospitals (85%) had insufficient data to be categorized.

Publicly reported measure

The version of the HOPD Surgery measure that is publicly reported is not adjusted for social risk factors. We performed two analyses to explore the impact of adding either of two social risk factors (DE, and ADI) to the model, on measure scores. We found that adding either social risk factor to the model did not result in major impacts on measure scores, suggesting that the variables in the risk model account for some of the differences we see in unadjusted patient-level outcome rates. Therefore, in this pay-for-reporting program, providers will not be unfairly profiled when assessed by the HOPD surgery measure. We describe the analyses and results below.

To examine the impact of social risk factors on measure scores, we first examined correlations (Pearsons) between measure scores with and without either social risk factor and found that correlations were near 1 (0.999, and 0.999, respectively) (Figure 8 and 9). Second, we examined the association between the facility proportion of patients with each social risk factor and measure scores, focusing on the quartile of facilities with the highest proportion of patients with social risk factors (Figure 9 and 10). We found that there is a very weak but significant correlation (r=0.039, p=0.03) between the proportion of patients with DE and the measure score for the fourth quartile of facility-proportion of patients with DE. However, there is no significant correlation for the high ADI variable (r=0.02, p=0.28). We concluded therefore, that there is little to no impact of adding social risk factors on measure scores for this HOPD Surgery measure. As described above, however, CMS has implemented confidential reporting of the measure, stratified for DE.

Figure 8: Correlation between measures scores calculated with and without DE in the risk model.

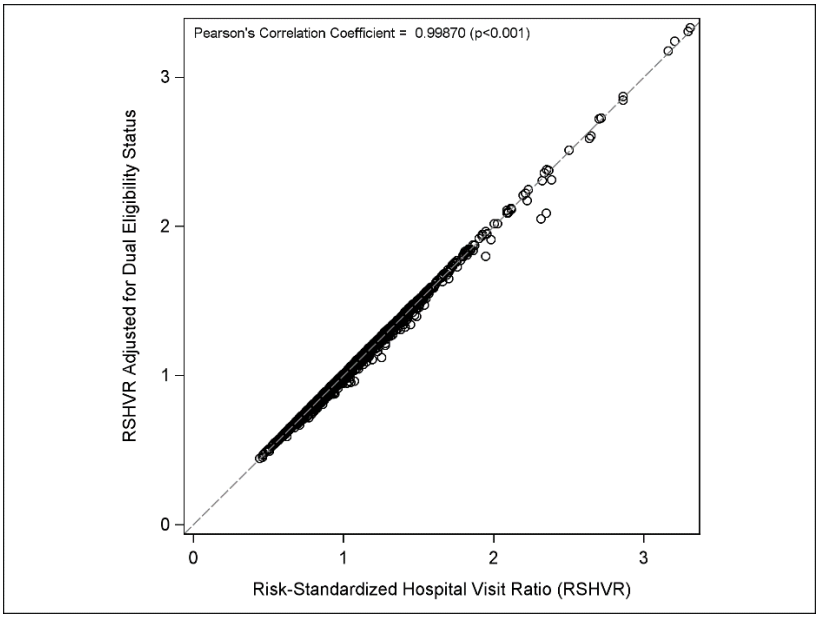


Figure 9: Correlation between measures scores calculated with and without high ADI in the risk model.

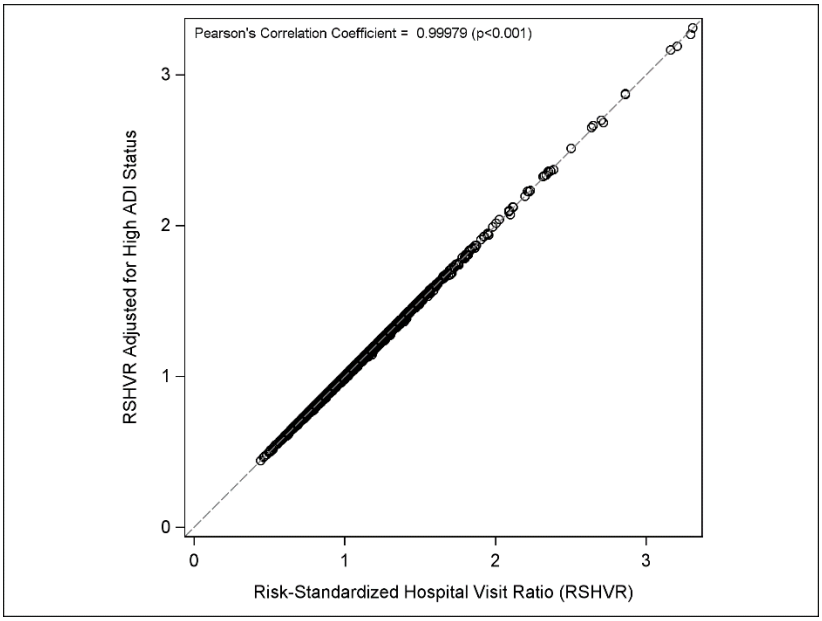


Figure 10: Spearman correlation between the facility proportion of patients with DE (top quartile) and measure scores (RSHVRs).

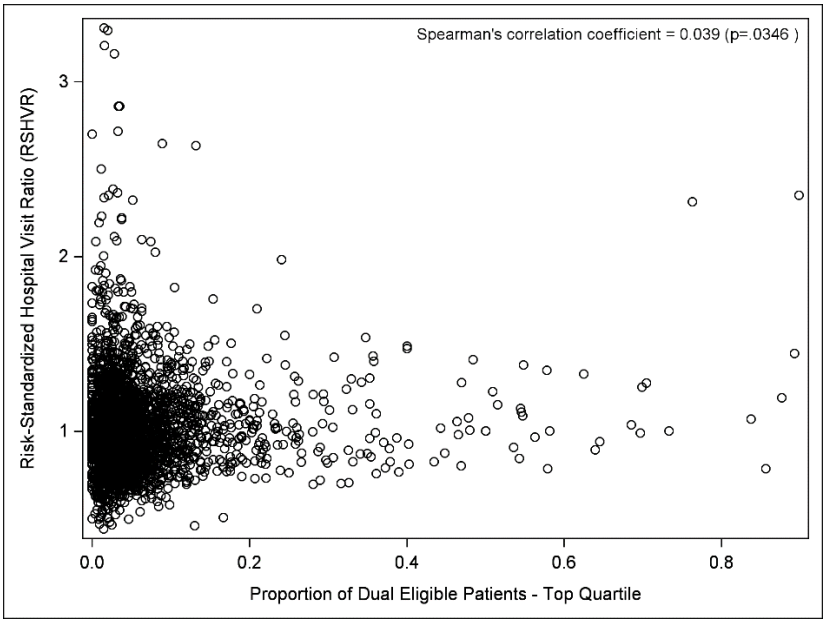
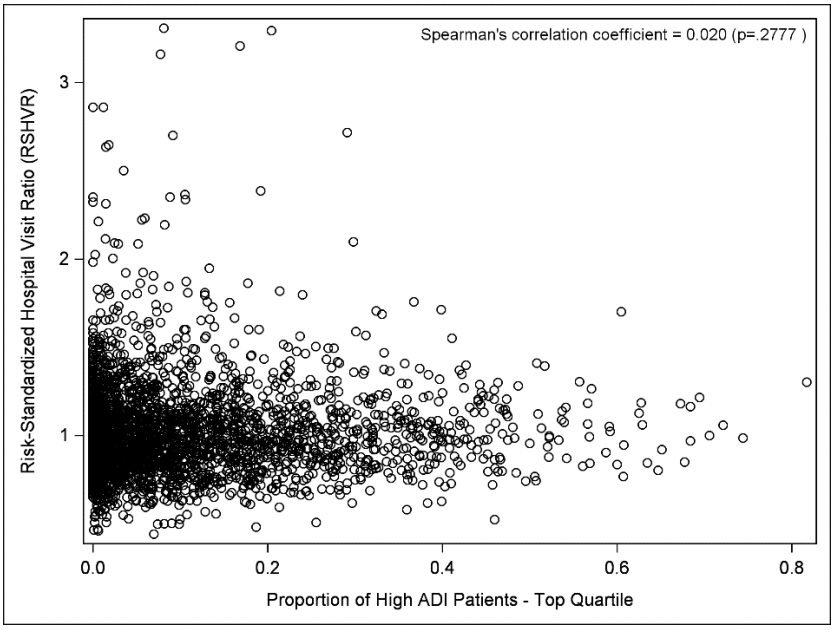


Figure 11: Spearman correlation between the facility proportion of patients with high ADI (top quartile) and measure scores (RSHVRs).



Use & Usability

Use

[For initial endorsement] **Check all current or planned uses** *

☐ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☐ Other
Please specify (text box)

[For maintenance review] **Check all current uses:** *

☒ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☒ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☐ Other
(Please specify (text box)
☐ Not in use
Please provide more information as to why the measure is not in use (text box)

[For maintenance review] **Please provide the following information describing the program(s) in which the measure is used**: *

Name of the program and sponsor *(text box)*
Sponsor: Hospital outpatient quality reporting program (HOQR), CMS

URL *(text box)*

https://qualitynet.cms.gov/outpatient/oqr

Purpose *(text box)*

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.

Geographic area and percentage of accountable entities and patients included *(text box)*

Geographic area: national; For the final cohorts from January 1, 2022 – December 31, 2022, there were 1,204,400 procedures performed in 3,818 facilities, representing 93.7% of included procedures.

Level of analysis and care setting. *(text box)*

The level of measurement is the facility; the setting is the Hospital Outpatient Department.

You may add additional programs or sponsors

Usability

What are the actions measured entities must take to improve performance on this measure? How difficult are those actions to achieve? *

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 aged 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.

[For maintenance only] **Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how you obtained feedback.** *

CMS receives feedback on all of its measures through the publicly available Q&A tool on Quality Net. Through this tool, we have received, since the last submission, only basic questions about the measure, including the cohort definition, the outcome definition, and specific questions about a facility's data. We did not receive any suggestions for changes to the HOPD surgery measure.

[For maintenance only] **Describe how you considered the feedback when developing or revising the measure specifications or implementation, including whether you modified the measure and why or why not. ***

We have made no major changes to the HOPD Surgery measure since it was last endorsed in 2020.

Minor measure updates include:

- Annual updates (2020, 2021, 2022) to ICD-10 codes that are used to specify the measure.

Each year, as part of reevaluation of the measure, CMS reviews the measure's existing code set as well as updates to ICD-10, CPT®, and HCPCS coding guidelines to ensure that the measure's code set is up to date.

- Update to coding for ED visits by shifting from the previously used 'claim from date' on the claim, to the 'minimum ED revenue center date' on the claim. This Aligns with changes made in the prior year to exclude cases based on this date.

[For maintenance only] **Discuss any progress on improvement (trends in performance results, including performance across sub-populations if available, number and percentage of people receiving high-quality healthcare, geographic area, number and percentage of accountable entities and patients included). If use of the measure demonstrated no improvement, provide an explanation. ***

The measure score is a risk-standardized hospital visit ratio (RSHVR) for each HOPD. It is calculated by computing the ratio of the number of predicted unplanned hospital visits (numerator) to the number of expected unplanned hospital visits (denominator). The numerator (predicted visits) is the number of unplanned hospital visits the HOPD is predicted to have, representing the observed unplanned hospital visit rate, the number of surgeries performed at the HOPD, and the HOPD's case mix. The denominator (expected rate) is the number of unplanned hospital visits the HOPD is expected to have based on the nation's performance with that HOPD's case mix and surgical procedure mix.

This measure captures an ever-changing mix of procedures based on procedures that have been added (or removed) from the ASC-covered procedure list, which is the basis for inclusion of procedures for this measure. For example, Total Knee Arthroplasty (TKA) was removed from the CMS inpatient-only (IPO) list in January 2018, allowing TKAs to be performed in the inpatient or outpatient hospital setting; TKA was added to the Ambulatory Surgery Center (ASC) Covered Procedures List in January 2020 and THA was added in January 2021. HOPD volumes of THA/TKA procedures have been steadily increasing; after the onset of COVID, the proportion of THA/TKA procedures performed (for Medicare FFS patients) in the outpatient setting exceeded those performed in the inpatient setting (data not shown); other than this HOPD Surgery measure, there are currently no active/implemented outpatient performance measures that capture complications following THA/TKA procedures performed at HOPDs. (CMS, has, however proposed the adoption of the THA/TKA PRO-PM for the HOPD and ASC settings (Federal Register, 2023)).

Because this measure includes procedures defined by the ASC covered procedures list, when new procedures, such as THA/TKA, are included, we expect this to impact measure scores across facilities which impacts our ability to track improvements in outcomes. In fact, with this most recent data update, we found that there are more statistical outliers (both better and worse than national rates) compared with the prior 2020 submission, suggesting that performance variation has widened with the expansion of the cohort to include additional outpatient surgical procedures.

Facilities, however, receive (from CMS) facility-specific information to support quality improvement. For example, they receive detailed patient-level details that indicate which patients experienced an unplanned hospital visit, what type of visit it was (e.g., inpatient admission, ED visit, observation stay) and the principal diagnosis code associated with the visit. They also receive summary information that shows their unadjusted performance by body system (e.g., musculoskeletal, urinary tract, etc.) in comparison with state and national benchmarks.

[For maintenance only] **Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.** *

There have been no unexpected findings during implementation. However, disruptions to the healthcare system due to COVID likely accelerated the migration of procedures to the outpatient space, resulting in changes to the case-mix of procedures captured by this measure (for example, more THA/TKA procedures).