**National Quality Forum—Measure Testing (subcriteria 2a2, 2b1-2b6)**

**Measure Number** (*if previously endorsed*)**:** 2539

**Measure Title**: Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy

**Date of Submission**: 1/6/2020

**Type of Measure:**

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

|  |
| --- |
| **Instructions**   * Measures must be tested for Fall the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form. * **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.** * **For outcome and resource use measures**, section **2b3** also must be completed. * If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed. * Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed. * If you are unable to check a box, please highlight or shade the box for your response. * Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.*** * Contact NQF staff regarding questions. Check for resources at [Submitting Standards webpage](http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx). * For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment. |

|  |
| --- |
| **Note:** The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF’s evaluation criteria for testing.  **2a2.** **Reliability testing** [**10**](#Note10) demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.  **2b1.** **Validity testing** [**11**](#Note11) demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.    **2b2.** **Exclusions** are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [**12**](#Note12)  **AND**  If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [**13**](#Note13)  **2b3.** **For outcome measures and other measures when indicated** (e.g., resource use):   * **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [**14**](#Note14)**,**[**15**](#Note15) and has demonstrated adequate discrimination and calibration   **OR**   * rationale/data support no risk adjustment/ stratification.   **2b4.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [**16**](#Note16) **differences in performance**;  **OR**  there is evidence of overall less-than-optimal performance.  **2b5.** **If multiple data sources/methods are specified, there is demonstration they produce comparable results**.  **2b6.** Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.  **Notes**  **10.** Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).  **11.** Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.  **12.** Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.  **13.** Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.  **14.** Risk factors that influence outcomes should not be specified as exclusions.  **15.** With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers. |

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

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

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

|  |  |
| --- | --- |
| **Measure Specified to Use Data From:**  **(*must be consistent with data sources entered in S.17*)** | **Measure Tested with Data From:** |
| abstracted from paper record | abstracted from paper record |
| claims | claims |
| registry | registry |
| abstracted from electronic health record | abstracted from electronic health record |
| eMeasure (HQMF) implemented in EHRs | eMeasure (HQMF) implemented in EHRs |
| other: Enrollment database files | other: Enrollment database files; Master Beneficiary Summary File (MBSF) Database, Census Data/American Community Survey |

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

We use paid, final action Medicare claims to identify colonoscopies performed in the outpatient setting at Ambulatory Surgical Centers (ASCs) and Hospital Outpatient Departments (HOPDs), and subsequent hospital visits. In addition, we use the Center for Medicare and Medicaid Services (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 colonoscopy procedure. The measure is calculated separately for HOPDs and ASCs, and the results in this form are presented separately by facility type.

For all derived cohorts:

a. Datasets used to define the cohort:

* All cohort, outpatient colonoscopy procedures performed at ASCs or HOPDs were identified using the full set of Medicare beneficiaries’ claims from the Carrier non-institutional claims, which included the ASC facility claims and physician bills for hospital outpatient services. HOPD claims were linked to the outpatient institutional colonoscopy claims or inpatient institutional colonoscopy 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 colonoscopy procedures 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 colonoscopy were used to identify comorbidities for risk adjustment for these patients.

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

The dates of the data vary by testing type as described in detail in Section 1.7.

**1.4. What levels of analysis** **were tested**? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

|  |  |
| --- | --- |
| **Measure Specified to Measure Performance of:**  **(*must be consistent with levels entered in item S.20*)** | **Measure Tested at Level of:** |
| individual clinician | individual clinician |
| group/practice | group/practice |
| hospital/facility/agency | hospital/facility/agency |
| health plan | health plan |
| other: ASCs and hospital outpatient facilities | other: ASCs and hospital outpatient facilities |

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

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

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

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

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

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Description of Dataset** | **Use and Section in the Testing Attachment** |
| Dataset #1:  Initial Development Dataset  Dataset #1a: Development dataset  Dataset #1b:  Validation dataset | We used two claims datasets for measure development.  To develop and test the patient-level model, CORE used 2009-2011 claims data from Medicare inpatient, outpatient, and carrier (Part B Physician) Standard Analytical Files (SAF). Specifically, we identified outpatient colonoscopies using 20% of Medicare FFS beneficiaries’ claims from the carrier SAF consisting of physician claims from ASCs, HOPDs and physician office settings.  For measure development and testing, we randomly split the 2010 data into Development and Validation Samples (each sample containing approximately 50% of colonoscopies contained in the 2010 data). For patients in these samples, we used data from 2009 to derive comorbidities for risk adjustment. We derived a cohort of colonoscopies in 2011 for temporal validation of the model (2011 Validation Sample), using 2010 data for risk adjustment.  **Dataset #1:**  Number of facilities (HOPDs and ASCs combined): 8,142  Number of procedures: 332,391  Percent female: 54.4%  Mean age: 74.2 years.  **Dataset #1a (development split sample)**  Number of facilities (HOPDs and ASCs combined): 7,475  Number of procedures: 166,196  **Dataset #1b (validation split sample)**  Number of facilities (HOPDs and ASCs combined): 7,475  Number of procedures: 166,196 | * Section 2b1 Validity testing (face validity) * Section 2b3.3a Identification and selection of risk-adjustment variables * Section 2b3.7 Risk model calibration statistics |
| Dataset #2:  Endorsement Maintenance Dataset | Final action Medicare claims (100%) were used identify colonoscopies performed in the outpatient setting at Hospital Outpatient Departments (HOPDs), and Ambulatory Surgical Centers (ASCs), 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 inpatient and outpatient claims data collected in the 12 months prior to the outpatient surgery.  Dates of data for the outcome: All analyses for this endorsement maintenance application were performed in data from the January 1, 2016 – December 31, 2018 performance year period.   |  |  | | --- | --- | | HOPDs - Number of procedures | 2,258,661 | | N-Facilities | 4034 | | N-Facilities >= 30 procedures during the performance period | 3583 | | Mean Age | 72.634 | | % Female | 53.4 |  |  |  | | --- | --- | | ASC - Number of procedures | 2,524,898 | | Number of facilities | 2261 | | Number of facilities >= 30 procedures during the performance period | 2073 | | Mean Age | 72.273 | | % Female | 53.8 | | * Section 2a.2 Reliability * Section 2b2 Testing of Measure Exclusion * Section 2b3.4b Selection of Social Risk Factors * Section 2b4 Meaningful Differences * Section 2b3.6 Predictive ability * Section 2b3.6 Statistical model discrimination statistics |

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

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

1. Medicare-Medicaid dual-eligibility status

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

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

We selected the AHRQ-validated SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas [2]. It is a widely used index that summarizes area-level measures of employment, income, education, and housing from the American Community Survey (ACS). Each of the index components is available at the census block level, which we then used to link to patient’s residence using 9-digit ZIP code. The AHRQ SES index score summarizes the following variables:

• Percentage of people in the labor force who are unemployed,

• Percentage of people living below poverty level,

• Median household income,

• Median value of owner-occupied dwellings,

• Percentage of people ≥25 years of age with less than a 12th grade education,

• Percentage of people ≥25 years of age completing ≥4 years of college, and

• Percentage of households that average ≥1 people per room.

The AHRQ SES Index’s value as a proxy for patient-level information is dependent on having the most granular level data with respect to communities that patients live in. In this submission, we present analyses using the census block group-level, the most granular level possible using ACS data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients’ 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, we adjusted the median income and median property value components of the AHRQ SES Index by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. We identify patients at risk due to social factors if they are in the bottom 25th percent of the ARHQ SES distribution.

Citations

1. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation. Report to Congress: Social Risk factors and Performance Under Medicare’s Value-based Payment Programs. 2016; <https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs>. Accessed December 8, 2019.
2. Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;2.

**2a2. RELIABILITY TESTING**

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

**2a2.1. What level of reliability testing was conducted? (may be one or both levels)**  
 Critical data elements used in the measure (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)  
 Performance measure score (e.g., signal-to-noise analysis)

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

Measure Score Reliability

We tested facility-level measure score reliability using the signal-to-noise method, using the formula presented by Adams and colleagues [1,2]. Specifically, for each facility we calculate the reliability as:

Reliability=(σ\_(facility-to-facility)^2)/(σ\_(facility-to-facility)^2+ (σ\_(facility error variance)^2)/n)

Where facility-to-facility variance is estimated from the hierarchical logistic regression model, n is equal to each facility’s observed case size, and the facility error variance is estimated using the variance of the logistic distribution (pi^2/3).

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 (using Dataset #2) for all facilities, and for facilities with a volume cutoff of 30 procedures. Our rationale for this is described below.

In general, CMS sets the volume cutoff for publicly reporting facility measures scores based on two considerations. CMS considers the empiric results of reliability testing conducted on the dataset used for public reporting. CMS also considers the volume cutoff for score reporting used for related measures. CMS has empirically determined that measure scores for facilities with 30 or more procedures are reliable. Regardless of the score reporting volume cutoff, all facilities and their cases are used in calculating the measure scores. In the dry run and in public reporting CMS typically reports scores for facilities with fewer procedures than the volume cutoff as having “too few cases” to support a reliable estimate. In summary, the measure specifications do not prejudge the ideal volume cutoff. The minimum sample size for public reporting is a policy choice that balances considerations such as the facility-level reliability testing results on the reporting data and consistency across measures for consumers.

Citations

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

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

**2a2.3. For each level of testing checked above, what were the statistical results from reliability testing**? (e*.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Measure score reliability (signal-to-noise reliability) for HOPDs and ASCs is shown in Table 1.

**Table 1: Signal-to-noise reliability for HOPDs and ASCs**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **All HOPDs** | **HOPDs with >=30 procedures** | **All ASCs** | **ASCs with >=30 procedures** |
| **Median signal-to- noise Reliability** | 0.744 | 0.782 | 0.864 | 0.883 |
| **Interquartile range (IQR)** | 0.489 - 0.883 | 0.596 - 0.892 | 0.628 - 0.938 | 0.714 - 0.942 |

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability?** (i.e., what do the results mean and what are the norms for the test conducted?)

**HOPDs**

Using three years of performance data, the median facility-level reliability score is **0.744** (IQR, 0.489 - 0.883) for all HOPDs and **0.782** (IQR, 0.596 - 0.892) for HOPDs with at least 30 cases, representing high reliability (“substantial agreement”) [1].

**ASCs**

Using three years of performance data, the median reliability is **0.864** (IQR, 0.628 - 0.938) for all ASCs and **0.883** (IQR, 0.714 - 0.942) for ASCs with at least 30 cases, representing high reliability (“almost perfect agreement”) [1].

These results indicate that there is sufficiently high reliability in the measure scores for ASCs and HOPDs.

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

< 0 – Less than chance agreement;

0 – 0.2 Slight agreement;

0.21 – 0.39 Fair agreement;

0.4 – 0.59 Moderate agreement;

0.6 – 0.79 Substantial agreement;

0.8 – 0.99 Almost Perfect agreement; and

1 Perfect agreement

Citations

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

**2b1. VALIDITY TESTING**

**2b1.1. What level of validity testing was conducted**? (*may be one or both levels*)  
 **Critical data elements** (*data element validity must address ALL critical data elements*)

**Performance measure score**

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

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

Face validity assessed based on facility feedback, dry run, and public reporting results:

We conducted a national confidential reporting period for HOPDs and ASCs in 2015, during which facilities received their measure results and facility-specific data used in measure calculation. During the dry run we solicited feedback from facilities on the measure specifications and results, and as a result, revised measure specifications. We also review feedback from facilities prior to beginning annual re-evaluation analyses in order to determine whether the measure continues to be valid.

The colonoscopy measure went into public reporting in December 2017. While the rate of hospital visits following colonoscopies for ASCs remained similar in 2019 vs 2018 public reporting (national rate of hospital visits per 1,000 colonoscopies for ASCs was 12.5 in 2018 and 12.3 in 2019), performance for HOPDs showed improvement. Compared to data from the prior year, performance on the colonoscopy measure for HOPDs showed improvement. The national rate of hospital visits per 1,000 colonoscopies among HOPDs declined from 16.4 in 2018 reporting to 14.8 in 2019 reporting, and the distribution of risk-standardized rates also declined; the interquartile range of rates for 2019 reporting lie completely below the 2018 interquartile range. We surmise that this decline reflects quality improvement as there were no specification changes to the measure for 2019 reporting that would impact rates, nor were there noticeable differences in patient mix.

Face validity as assessed during measure development:

We demonstrated measure validity through 1) use of established measure development guidelines, 2) assessment by external groups, and 3) systematic assessment of measure face validity by a technical expert panel (TEP) of national experts and stakeholder organizations.

Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input through holding regular discussions with the external experts in our working group, consulting our national TEP, and holding a 30-day public comment period.

Yale New Haven Health Services Corporation—Center for Outcomes Research and Evaluation (CORE) clinicians as well as two national clinical leaders in the field of gastroenterology comprised the working group. Through regular in-person meetings and teleconferences, the working group discussed all aspects of measure development, including the cohort and outcome definitions, and risk adjustment.

In addition to the working group and in alignment with the CMS Measures Management System, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives including clinicians, patients, and individuals with experience in quality improvement, performance measurement, and healthcare disparities. We held three structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We made minor modifications to the measure specifications (e.g., outcome definition) based on TEP feedback on the measure.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site: https://www.CMS.gov/MMS/17\_CallforPublicComment.asp. We made refinements to the measure in response to public comment.

Face Validity as Determined by TEP

We also asked our TEP, made up of 17 members including patient representatives, expert clinicians, methodologists, researchers, and providers, to formally assess the measure’s face validity.

List of TEP Members

1) Joel Brill, MD; Predictive Health LLC (Chief Medical Officer); Fair Health (Medical Director)

2) Zahid Butt, MD; Medisolv Inc. (CEO)

3) David Chang, PhD, MPH, MBA; University of California San Diego (Director of Outcomes Research, Assistant Professor, Department of Surgery)

4) Richard Dutton, MD, MBA; Anesthesia Quality Institute (Executive Director)

5) Brian Fennerty, MD; Oregon Health and Science University (Professor of Medicine, Department of Internal Medicine, Section of Gastroenterology)

6) Terry Golash, MD; Aetna, Inc. (Senior Medical Director)

7) Claudia Gruss, MD; Arbor Medical Group, a division of ProHealth (Physician Partner)

8) Cynthia Ko, MD, MS; University of Washington (Associate Professor, Division of Medicine; Adjunct Associate Professor, Department of Health Services)

9) David Lieberman, MD; Oregon Health and Science University (Professor of Medicine; Chief, Division of Gastroenterology and Hepatology)

10) Keith Metz, MD, JD, MSA; Great Lakes Surgical Center (Medical Director)

11) Michael Morelli, MD, CPE; Indianapolis Gastroenterology and Hepatology (President)

12) Philip Schoenfeld, MD, MSEd, MSc; University of Michigan (Professor of Medicine, Division of Gastroenterology)

13) Anthony Senagore, MD, MS, MBA; Central Michigan University, School of Medicine (Chair, Surgical Disciplines)

14) Joan Warren, PhD; Applied Research Program, NIH, National Cancer Institute (Epidemiologist)

15) Jennifer Weiss, MD, MS; University of Wisconsin School of Medicine and Public Health (Assistant Professor, Department of Medicine – Division of Gastroenterology & Hepatology)

16, 17) Two patients

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting TEP members’ agreement with the following statement: *“The risk-standardized hospital visit rates obtained from the colonoscopy measure as specified can be used to distinguish between better and worse quality facilities.”*

The 14 TEP members who responded to the survey 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

External Empiric Validity

Stewards of NQF-endorsed measures going through the re-endorsement process are required to demonstrate external validity testing at the time of maintenance review, or if this is not possible, justify the use of face validity only. To meet this requirement for the Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy measure (CMS colonoscopy measure), we would need to identify and assess the measure’s correlation with other measures of HOPD or ASC colonoscopy quality that target the same domain of quality (e.g. complications, safety, or post-procedure utilization) for the same or similar populations. If such measures exist, a positive correlation between the other measure scores and the CMS colonoscopy measure score at the facility level would strengthen the evidence of the CMS colonoscopy measure’s validity. When the measure was developed and initially endorsed, it filled a gap and no such similar measures existed. However, relevant measures may have been developed since the CMS colonoscopy measure’s endorsement. We therefore searched for and considered similar measures that we could use to further test the CMS measure’s validity.

We first considered CMS’s two related NQF-endorsed measure, **Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at ASCs (ASC General Surgery),** and **Hospital Visits after Hospital Outpatient Surgery (HOPD Surgery).** The outcome of both measures is nearly identical to that of the colonoscopy measure; an unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission. Hence, the measures target the same quality domains as the CMS colonoscopy measure. The patient cohort is also somewhat similar in that the measures target Medicare Fee-For-Service (FFS) patients aged 65 years and older. The cohorts, however, have no overlap with the colonoscopy measure, because they include patients undergoing general surgery, not colonoscopy procedures.

Nevertheless, it could be hypothesized that HOPDs or ASCs that perform both general surgery procedures and colonoscopy procedures might have correlated measure scores for these two groups, given that patients in the two groups may to some extent share post-operative care, discharge planning services, and facility-wide policies that affect patient care. However, many ASCs specialize in a single procedure (in 2017, more than 60 percent of ASCs were single-specialty), and gastroenterology is one of the most common single-specialty facility types [1]. Therefore, one would not expect that ASCs performing colonoscopies to be the same facilities that would be measured in the ASC General Surgery measure. While HOPDs are typically not single-specialty, they are unlikely to share the same procedural suites or providers that are captured by the HOPD surgery measure. We therefore concluded these measures cannot be used for validity testing of the CMS colonoscopy measure.

**Colonoscopy-related Measures Endorsed by NQF**

To identify non-CMS measures against which to validate, we first searched NQF’s Quality Positioning System (QPS) for measures related to colonoscopy and colorectal cancer screening and identified three colonoscopy-related measures that are endorsed by NQF. These measures assess the proportion of patients that received colorectal cancer screenings. Each measure is classified as a process measure.

**1. Colorectal Cancer Screening (electronic clinical quality measure [eCQM]):**

Identifies the proportion of patients in the recommended age group for colonoscopy screenings (50-75) who have had the procedure.

**2. Appropriate Follow-Up Interval for Normal Colonoscopy in Average Risk Patients**

Identifies the percentage of patients who have received a screening colonoscopy and have a regular recommended follow-up of ten years. This measure excludes patients who are older than 66 or who have a life expectancy of fewer than ten years, as the follow-up colonoscopy is no longer deemed beneficial. This measure is also not risk-adjusted.

**3. Colonoscopy Interval for Patients with a History of Adenomatous Polyps – Avoidance of Inappropriate Use**

Measures the percent of patients who appropriately receive a colonoscopy greater than three years after a previous colonoscopy. This measure is designed to track procedures that are inappropriately done within three years, and excludes procedures that occur within three years, but have a documented reason for the interval. This measure is not risk-adjusted.

**Assessment**

The three measures described above do not assess the domains of quality measured by the CMS colonoscopy measure. The facility-level scores for these measures would therefore not be expected to correlate with facilities’ 7-Day Risk-Standardized Hospital Visit rate and cannot be used to externally validate the CMS measure.

In summary, none of the measures that we identified meet the criteria for a comparator measure that could be used for external validation. We therefore present face validity results for this measure as meeting the requirements for validity.

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

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

•ICD-10 diagnosis codes used to define diverticulitis of the colon and inflammatory bowel disease (IBD) were identified from ICD-10-CM codes using the ICD-9-CM to ICD-10-CM General Equivalence Mapping (GEM) files made available by CMS.

•Similarly, procedure codes used to define total colectomy were identified from the ICD-10 PCS codes using the General Equivalence Mapping (GEM) files made available by CMS.

•ICD-10 codes were searched separately to ensure capture of all relevant ICD-10-CM and PCS codes.

One of the physicians on our team created the initial crosswalk of ICD-9-to-ICD-10 codes following the process above. A second physician performed an initial review of the list. Then the measure’s two working group external experts reviewed the list. Following a review of the proposed crosswalk, our working group experts confirmed that the proposed ICD-10 codes and crosswalk were appropriate.

Citations

1. MedPAC’s Report to Congress, Chapter 5, Ambulatory Surgical Center Services, March 2019. <http://www.medpac.gov/docs/default-source/reports/mar19_medpac_ch5_sec.pdf?sfvrsn=0>; Accessed December 9, 2019.

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

**Validity as Assessed by External Groups**

The distribution of the responses is shown below:

Mean rating=4.6.

**Frequency of Ratings of Agreement**

**Rating # (%) of Responses**

1 (Strongly disagree) 0 (0)

2 (Moderately disagree) 1 (7.1)

3 (Somewhat disagree) 1 (7.1)

4 (Somewhat agree) 2 (14.3)

5 (Moderately agree) 8 (57.1)

6 (Strongly agree) 2 (14.3)

Of the 14 TEP members who responded to the survey, 12 (86%) indicated they somewhat, moderately, or strongly agreed with the validity statement. In addition, one TEP member somewhat disagreed, and one TEP member moderately disagreed. The TEP member who moderately disagreed did not provide a reason. The reason for the other TEP member’s disagreement can no longer be accessed due to software restrictions.

**External Empiric Validity**

As noted above in section 2b1.2, none of the measures that we identified meet the criteria for a comparator measure that could be used for external validation so no quantitative comparisons to other measures were conducted.

Finally, we note that the measure has been in public reporting since December 2017, and stakeholders have not raised concerns to CMS about its validity. Furthermore, as described above, we have seen improvement in performance on this measure from HOPDs.

We therefore present face validity results for this measure as meeting the requirements for validity, in addition to providing feedback from measured entities and public reporting results.

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

The measure as specified has sufficient face validity, based on TEP agreement (86%) that the measure can be used to distinguish between higher and lower-performing facilities, and its acceptability to providers currently measured.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b2. EXCLUSIONS ANALYSIS**

**NA**  **no exclusions — *skip to section*** [***2b3***](#section2b4)

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

We examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion. We used Dataset #2 (January 1, 2016-December 31, 2018) for this analysis.

**2b2.2. What were the statistical results from testing exclusions**? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Applying our inclusion criteria (procedures with a qualifying colonoscopy procedure code performed for patients aged ≥65 years enrolled in Medicare Parts A & B FFS in the 12 months prior to the procedure, performed at ASCs or HOPDs) to the Medicare claims data from January 1, 2016 through December 31, 2018 (Dataset #2) resulted in an initial cohort of 2,484,741 procedures in HOPDs and 2,703,335 procedures in ASCs.

We then applied the exclusion criteria listed in the tables below to HOPDs (Table 2A) and ASCs (Table 2B). (Note that the excluded procedure groups are not mutually exclusive; see the Intent to Submit/Measure Submission Form, Sections S.8 and S.9, for full list of exclusions and codes.)

**Table 2A. HOPD Colonoscopy Measure Exclusions – 2016-2018 performance period**

|  |  |  |
| --- | --- | --- |
| **Exclusions** | **Number of Procedures** | **% of All Included Procedures** |
| [All included procedures] | 2,484,741 |  |
| Procedures for patients who lack continuous enrollment in Medicare FFS Parts A & B in the 7 days after the procedure | 1,397 | 0.06 |
| Colonoscopies that occur concurrently with high-risk upper GI endoscopies | 31,431 | 1.27 |
| Procedures followed by a subsequent outpatient colonoscopy within 7 days | 3,680 | 0.15 |
| Procedures for patients with a history or current diagnosis of IBD | 84,966 | 3.42 |
| Procedures for patients with a history or current diagnosis of diverticulitis | 98,192 | 3.95 |
| Colonoscopies that occur on the same day and at the same hospital as an ED visit that is billed on a different claim than the index colonoscopy, unless the ED visit has a diagnosis indicative of a complication of care | 4,277 | 0.17 |
| Colonoscopies that are billed on the same hospital claim as an emergency department (ED) visit and occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care | 1,502 | 0.06 |
| Colonoscopies that are billed on the same hospital outpatient claim and that occur after the ED visit | 6,938 | 0.28 |
| Colonoscopies that are billed on the same hospital outpatient claim as an observation stay | 11,015 | 0.44 |
| [Final Cohort] | 2,258,661 | 90.9 |

Counts may be duplicated across exclusions.

**Table 2B. ASC Colonoscopy Measure Exclusions – 2016-2018 performance period**

|  |  |  |
| --- | --- | --- |
| **ASC Exclusions** | **Number of Procedures** | **% of All Included Procedures** |
| [All included procedures] | 2,703,335 |  |
| Procedures for patients who lack continuous enrollment in Medicare FFS Parts A & B in the 7 days after the procedure | 1,429 | 0.05 |
| Colonoscopies that occur concurrently with high-risk upper GI endoscopies | 15,051 | 0.56 |
| Procedures followed by a subsequent outpatient colonoscopy within 7 days | 5,274 | 0.20 |
| Procedures for patients with a history or current diagnosis of IBD | 81,243 | 3.00 |
| Procedures for patients with a history or current diagnosis of diverticulitis | 79,846 | 2.95 |
| [Final Cohort] | 2,524,898 | 93.4 |

Counts may be duplicated across exclusions.

**2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis.*  *Note:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

We determined the exclusion criteria after extensive literature review, harmonization with similar measures and discussion with the working group and TEP members. The goal was to be as inclusive as possible while creating a clinically coherent cohort; the measure population had to be sufficiently similar in terms of the procedure and outcome risk profile to ensure that risk adjustment can be performed adequately. We therefore excluded: (1) high-risk procedures and patient groups for which risk adjustment would not be adequate and (2) procedures and patient groups for which the outcome of hospital visits was a less appropriate indicator of quality. These exclusions prevent an unfair distortion of performance results. The rationales for individual exclusions are detailed in the Measure Submission Form/Intent to Submit form, Section S.8 and S.9. After exclusions were applied, the measure captured the majority (91-93%) of all qualifying colonoscopies. The exclusions are very narrowly targeted and necessary for the measure’s validity.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**  
***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section*** [***2b4***](#section2b5)***.***

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

**No risk adjustment or stratification**

**Statistical risk model with** 15 **risk factors (23 model parameters)**

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

**Other,** Click here to enter description

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

We fitted a hierarchical generalized linear model (HGLM), which accounts for the clustering of observations within hospitals. 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 hospital by estimating a hospital-specific effect,  which we assume follows a normal distribution with mean and variance , the between-hospital variance component. The following equations *define* the HGLM:

(1)

*where*

*i=1,…, I; j=1,…,*

Where *Yij* denotes the outcome (equal to 1 if patient has one or more qualifying hospital visit within 7 days of facility outpatient colonoscopy, 0 otherwise) for the *j-th* patient who had an outpatient colonoscopy at the *i-th* facility; is a set of *p* patient-specific covariates derived from the data; and *I* denotes the total number of facilities and is the number of outpatient colonoscopies performed at facility *i*. The [facility-specific intercept](#g_facilityspecificintercept), or effect, of the *i-th* facility*,* , defined above, comprises *,* the adjusted average intercept over all facilities in the sample, and *,* the facility-specific intercept deviation from **. A point estimate of , greater or less than 0, determines whether facility performance is worse or better compared to the adjusted average outcome.

Modeling is performed separately for HOPDs and ASCs.

**Risk Variables**

The risk-adjustment model has 16 variables (age categories, age categorized x arrhythmia interaction, twelve comorbidity variables, and two surgical variables). With the exception of concomitant endoscopy and polypectomy during procedure, which we define using individual CPT® codes, we define comorbidity variables using v22 CMS Condition Categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9 and ICD-10 diagnosis codes maintained by CMS.

aSee Tab 5, “Colonos\_risk\_factor\_CCs” and Tab 6, “Colonos\_Risk\_Factor\_CPT” in the attached Data Dictionary for the list of CC and CPT codes used to define the colonoscopy model risk variables.

Model Variablesa:

1. Age Categorized (years 65-69; 70-74; 75-79; 80-84; 85+)
2. Concomitant Endoscopy
3. Polypectomy during Procedure
4. Congestive Heart Failure (CC 85)
5. Ischemic Heart Disease (CC 86-89)
6. Stroke/Transient Ischemic Attack (CC 99, CC 101)
7. Chronic Lung Disease (CC 111-113)
8. Metastatic Cancer (CC 8-11)
9. Liver Disease (CC 27-32)
10. Iron Deficiency Anemia (CC 49)
11. Disorders of Fluid, Electrolyte, Acid-Base (CC 24)
12. Pneumonia (CC 114-116)
13. Psychiatric Disorders (CC 57-59, 61-63)
14. Drug and Alcohol Abuse/Dependence (CC 54-56)
15. Arrhythmia (CC 96-97)
16. Age Categorized x Arrhythmia Interactions

**Table 3A. HOPDs: Adjusted ORs and 95% CIs for the Colonoscopy Logistic Regression Model (Dataset #2; January 1, 2016-December 31, 2018)**

| **Variable (CC)** | **Odds Ratios (95% CI)** |
| --- | --- |
| Concomitant Endoscopy | 1.31 (1.28-1.34) |
| Polypectomy during Procedure | 1.26 (1.24-1.29) |
| Congestive Heart Failure (CC 85) | 1.31 (1.28-1.35) |
| Ischemic Heart Disease (CC 86-89) | 1.29 (1.26-1.32) |
| Stroke/Transient Ischemic Attack (TIA) (CC 99-101) | 1.18 (1.15-1.22) |
| Chronic Lung Disease (CC 111-113) | 1.27 (1.24-1.30) |
| Metastatic Cancer (CC 8-11) | 1.07 (1.04-1.10) |
| Liver Disease (CC 27-32) | 1.24 (1.2-1.28) |
| Iron Deficiency Anemia (CC 49) | 1.30 (1.27-1.33) |
| Disorders of Fluid, Electrolyte, Acid Base (CC 24) | 1.42 (1.38-1.46) |
| Pneumonia (CC 114-116) | 1.19 (1.15-1.23) |
| Psychiatric Disorders (CC 57-59, 61-63) | 1.36 (1.33-1.39) |
| Drug and Alcohol Abuse/Dependence (CC 54-56) | 1.22 (1.18-1.26) |
| Age by Arrhythmia Interaction | - |
| Among those without Arrhythmia (CC 96-97) | - |
| Age 70-74 v. Age 65-69 | 1.05 (1.02-1.09) |
| Age 75-79 v. Age 65-69 | 1.24 (1.2-1.29) |
| Age 80-84 v. Age 65-69 | 1.51 (1.44-1.58) |
| Age 85+ v. Age 65-69 | 2.12 (1.99-2.26) |
| Among those with Arrhythmia (CC 96-97) | - |
| Age 70-74 v. Age 65-69 | 0.98 (0.93-1.03) |
| Age 75-79 v. Age 65-69 | 1.10 (1.04-1.15) |
| Age 80-84 v. Age 65-69 | 1.27 (1.2-1.35) |
| Age 85+ v. Age 65-69 | 1.63 (1.52-1.74) |

Notes: Results based on January 1, 2016 -December 31, 2018, performance period data. Risk-factor definitions in this table are based on the v22 CC definitions. OR=Odds ratio CI=Confidence interval

**Table 3B. ASCs: Adjusted ORs and 95% CIs for the Colonoscopy Logistic Regression Model (Dataset #2; January 1, 2016-December 31, 2018)**

| **Variable (CC)** | **Odds Ratios (95% CI)** |
| --- | --- |
| Concomitant Endoscopy | 1.32 (1.28-1.35) |
| Polypectomy during Procedure | 1.32 (1.29-1.35) |
| Congestive Heart Failure (CC 85) | 1.28 (1.23-1.33) |
| Ischemic Heart Disease (CC 86-89) | 1.21 (1.17-1.24) |
| Stroke/Transient Ischemic Attack (TIA)  (CC 99-101) | 1.18 (1.14-1.22) |
| Chronic Lung Disease (CC 111-113) | 1.3 (1.26-1.33) |
| Metastatic Cancer (CC 8-11) | 1.15 (1.11-1.19) |
| Liver Disease (CC 27-32) | 1.28 (1.23-1.32) |
| Iron Deficiency Anemia (CC 49) | 1.23 (1.2-1.26) |
| Disorders of Fluid, Electrolyte, Acid Base (CC 24) | 1.41 (1.36-1.46) |
| Pneumonia (CC 114-116) | 1.22 (1.16-1.27) |
| Psychiatric Disorders (CC 57-59, 61-63) | 1.39 (1.35-1.43) |
| Drug and Alcohol Abuse/Dependence (CC 54-56) | 1.26 (1.21-1.31) |
| Age by Arrhythmia Interaction | - |
| Among those without Arrhythmia (CC 96-97) | - |
| Age 70-74 v. Age 65-69 | 1.11 (1.08-1.15) |
| Age 75-79 v. Age 65-69 | 1.26 (1.22-1.31) |
| Age 80-84 v. Age 65-69 | 1.6 (1.52-1.68) |
| Age 85+ v. Age 65-69 | 2.11 (1.95-2.29) |
| Among those with Arrhythmia (CC 96-97) | - |
| Age 70-74 v. Age 65-69 | 0.97 (0.91-1.03) |
| Age 75-79 v. Age 65-69 | 1.12 (1.05-1.19) |
| Age 80-84 v. Age 65-69 | 1.35 (1.25-1.45) |
| Age 85+ v. Age 65-69 | 1.65 (1.5-1.82) |

Results based on January 1, 2016 -December 31, 2018, performance period data. Risk-factor definitions in this table are based on the v22 CC definitions. OR=Odds ratio, CI=Confidence interval

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

Not applicable. This measure is risk-adjusted.

**2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk** (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any “ordering” of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

Description of Risk Adjustment Method

We use a two-level hierarchical logistic regression model to estimate risk-standardized hospital visit rates (RSHVRs). This approach accounts for the clustering of patients within facilities and variation in sample size. Our approach to risk adjustment is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines [1,2].

The risk-standardization model has 15 patient-level variables (age, concomitant upper GI endoscopy, polypectomy, and 12 comorbidity variables) and one interaction variable. We define comorbidity variables using v22 CCs. Maps showing the assignment of ICD-9 codes and ICD-10 codes to CCs can be found at: <https://www.qualitynet.org/outpatient/measures/colonoscopy/resources>.

Certain CCs are considered possible complications of care and are not risk-adjusted for if they only occur during the procedure. This is because only comorbidities that convey information about the patient at the time of the procedure or in the 12 months prior, and not complications that arose during the colonoscopy procedure, are included in the risk adjustment. See attached Data Dictionary, Tab 7 “Colonos\_CoC\_CCs” for CCs that are considered possible complications of care and are not risk-adjusted for if they only occur at the procedure.

Selection of Risk-Adjustment Variables during Measure Development

Candidate risk-adjustment variables were patient-level risk adjustors that are expected to be predictive of hospital visits following colonoscopy, based on prior literature, clinical judgment, and empirical analysis. We limited our initial selection of candidate variables for inclusion in our preliminary colonoscopy-specific risk-adjustment model to variables with a strong clinical rationale for inclusion as identified in the literature and through clinical expert input. These variables include age, sex, indicators of comorbidity and disease severity, and two procedural factors associated with an increased risk of adverse outcomes following colonoscopy (concomitant upper GI endoscopy and polypectomy during the procedure).

Variable Selection

To select the final variables to include in the risk-adjustment model, using Dataset #1, we fitted a logistic regression model to predict the outcome with the candidate variable set. To develop a parsimonious model, we then removed non-significant variables from the initial model using a stepwise purposeful selection method described by Hosmer and Lemeshow [3]. Our goal was to minimize the number of variables in the model while preserving model performance (as measured by the c-statistic). During this process, the least significant variable in the model was removed one at a time until only statistically significant (p<0.05, assessed using a likelihood ratio test) variables remained in the model. Interaction terms between variables were tested and were only retained in the model if significant at a level of p<0.01. The higher threshold for statistical significance ensured that only interactions that have a higher likelihood of being true interactions were included.

More detail about risk adjustment variable selection, including a list of candidate risk adjustment variables, can be found in the “Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy Measure Technical Report,” 2015: <https://www.qualitynet.org/files/5d0d37ae764be766b010196e?filename=ClnscpyMsr_TechReport.pdf>.

**Social Risk Factors for Disparities Analyses**

We selected variables representing social risk factors based on a review of literature, conceptual pathways, and feasibility. In section 1.8, we describe the variables available in Medicare claims data that we considered and analyzed, based on this review. Below, we describe the pathways by which social risk factors may influence risk of the outcome.

**Causal Pathways for Social Risk Variable Selection**

Our conceptualization of the pathways by which patients’ social risk factors affect the outcome was informed by the literature [4-6] 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) [7-9].

**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. There were no studies that addressed colonoscopy specifically, therefore we did not find any studies that suggested that variation in patients’ social risk factors affected variation in colonoscopy outcome risk across facilities. A recent update of this original search, examining only studies published since 2016, did not identify any additional studies.

**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 a colonoscopy, 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 hospital visit risk is that patients may not receive equivalent care within a facility [4,7]. However, as noted above, studies of colonoscopy in the HOPD and ASC 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 [5,6]. While analogous data for patients undergoing colonoscopies at HOPDs and ASCs is lacking, a similar pattern may exist, leading to higher (worse) outcome rates for patients with social risk factors.

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 colonoscopy provider 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 their approach to preparation for the procedure, 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 procedures 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 or unstable housing may have a worse general health status and may present for their procedure with greater severity of underlying illness [7]. This causal pathway should be largely accounted for by current clinical risk-adjustment.

As indicated in Section 1.8, the social risk variables that we examined are:

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

**ICD-9 to ICD-10 Conversion**

Statement of Intent

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

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

[ ] The intent of the measure has changed.

**Process of Conversion**

ICD-10 codes were initially identified using General Equivalence Mapping (GEM) software. For the initial conversion to ICD-10, we reviewed the 2016 ICD-10 coding system in detail and enlisted the help of clinicians to select and evaluate which of the ICD-10 codes that mapped to the ICD-9 codes were appropriate for use in this measure. Upon updating the codes, we tested the performance of the measure’s risk model, and impact on risk-standardized hospital visit ratios at the facility level in the most recent measurement years of data available. We then solicited input from clinical and measure experts to confirm the clinical appropriateness of the changes to the specifications given the updates to the ICD-10 codes. In addition, changes to ICD-10 codes are routinely monitored for their potential impact on this measure, and updates are made accordingly on an annual basis (most recently in 2019).

Citations

1. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: An American Heart Association scientific statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council endorsed by the American College of Cardiology Foundation. Circulation. 2006; 113(3):456-462.

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

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

4. 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.

5. 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.

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

7. 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.

8. 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.

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

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

**Published literature**

**Internal data analysis**

**Other (please describe)**

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

As mentioned above, we iteratively removed non-significant variables from the initial model using a step-wise purposeful selection approach until only statistically significant (p<0.05, assessed using a likelihood ratio test) variables remained in the model. Interaction terms between variables were tested and were only retained in the model if significant at a level of p<0.01.

The following variables were selected as the final risk adjustment variables, updated to include v22 CCs (aSee Tab 5, “Colonos\_risk\_factor\_CCs” and Tab 6, “Colonos\_Risk\_Factor\_CPT” in the attached Data Dictionary for the list of CC and CPT codes used to define the colonoscopy model risk variables):

Age Categorized (65-69; 70-74; 75-79; 80-84; 85+)

Concomitant Endoscopy

Polypectomy during Procedure

Congestive Heart Failure (CC 85)

Ischemic Heart Disease (CC 86-89)

Stroke/Transient Ischemic Attack (CC 99-101)

Chronic Lung Disease (CC 111-113)

Metastatic Cancer (CC 8-11)

Liver Disease (CC 27-32)

Iron Deficiency Anemia (CC 49)

Disorders of Fluid, Electrolyte, Acid-Base (CC 24)

Pneumonia (CC 114-116)

Psychiatric Disorders (CC 57-59, 61-63)

Drug and Alcohol Abuse/Dependence (CC 54-56)

Arrhythmia (CC 96-97)

Age Categorized x Arrhythmia Interaction

**2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors** *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Methods

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

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

All analyses were performed with data from January 1, 2016-December 31, 2018 (Dataset #2). **Analysis #1.** **Distribution of social risk factors across measured entities**: To assess the extent to which any effects of social risk factors may differentially influence the scores of a subset of providers, we examined how the proportion of patients with each social risk factor varied across HOPDs and ASCs.

The prevalence of social risk factors varied across measured entities as shown in Table 4A and 4B, below. The distribution was skewed; among the HOPDs in the top quartile of the distribution, the proportion of patients with social risk factors ranged from >10.23% to 100% for the DE variable, and >27.27% to 100% for the low AHRQ SES Index variable. For ASCs in the top quartile of the distribution, the proportion of patients with social risk factors ranged from >5.60%-100% for the DE variable, and >17.20%-100% for the low AHRQ SES Index variable. We therefore also analyze this group of facilities separately in Analyses #6 and #7 (see page 30-34).

**Table 4A: HOPDs: Percent and count of patients with social risk factors, per facility**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Social risk variable** | **Min (%)** | **Min (N)** | **Median (%)** | **Median (N)** | **Max**  **(%)** | **Max (N)** | **Interquartile range (%)** | **Interquartile range (N)** |
| **DE (Yes)** | 0% | 0 | 5.47% | 17 | 100% | 949 | 2.65% - 10.23% | 5-41 |
| **AHRQ SES Index (lowest quartile)** | 0% | 0 | 13.00% | 36 | 100% | 1581 | 4.64% - 27.27% | 9-96 |

**Table 4B: ASCs: Percent of patients with social risk factors, per facility**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Social risk variable** | **Min (%)** | **Min (N)** | **Median (%)** | **Median (N)** | **Max**  **(%)** | **Max (N)** | **Interquartile range (%)** | **Interquartile range (N)** |
| **DE (Yes)** | 0% | 0 | 2.30% | 16 | 100% | 1272 | 0.95% - 5.60% | 4-46 |
| **AHRQ SES Index (lowest quartile)** | 0% | 0 | 8.35% | 50 | 100% | 2367 | 3.72% - 17.20% | 12-149 |

**Analysis #2. Patient-level observed hospital visit rates for patients with social risk factors**  
To evaluate the association of these risk factors with the outcome, we first quantified the overall observed hospital visit rate for each social risk factor group (dual-eligible: yes vs. no, AHRQ SES Index: lowest quartile of SES Index vs. all others) for HOPDs (Table 5A) and ASCs (Table 5B).

**For HOPDs**, the outcome rate for patients with dual-eligible (DE) status and low AHRQ SES was higher than the outcome rate for patients who do not have the social risk factor (Table 5A: DE: 3.02% vs. 1.55%, p<0.0001); AHRQ SES: 2.10% vs. 1.57%, p<0.0001). The outcome rate for all patients was 1.64%.

**For ASCs**, the difference in the observed outcome rate for patients with the social risk factors is less marked than for HOPDs (Table 5B: DE: 1.97% vs. 1.19%, p<0.0001; AHRQ SES: 1.59% vs. 1.18, p<0.0001). The outcome rate for all patients was 1.22%.

**Table 5A: HOPDs: Observed hospital visit rates for patients with, and without social risk factors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Observed rate in patients with** **the social risk factor** | **Observed rate in patients without the social risk factor** | **p-value (patients with vs. without the social risk factor)** | **Observed rate (all patients)** |
| **DE (Yes vs No)** | 3.02% | 1.55% | p<0.0001 | 1.64% |
| **AHRQ SES (lowest quartile vs. all others)** | 2.10% | 1.57% | p<0.0001 |

**Table 5B: ASCs: Observed hospital visit rates for patients with, and without social risk factors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Observed rate in patients with the social risk factor** | **Observed rate in patients without the social risk factor** | **p-value**  **(patients vs. without the social risk factor)** | **Observed rate (all patients)** |
| **DE (Yes vs No)** | 1.97% | 1.19% | p<0.0001 | 1.22% |
| **AHRQ SES (lowest quartile vs. all others)** | 1.59% | 1.18% | p<0.0001 |

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

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

**For HOPDs**, in the bivariate models, both social risk factors have an odds ratio greater than one, indicating patients with the social risk factor have an increased risk of the outcome (see Table 6A). When we include these variables in a multivariate model that includes all of the final risk model variables, the odds ratios for both the dual eligible and AHRQ SES variables in the multivariate model were lower than the odds ratio for the bivariate association (Table 6A; DE: OR 1.98 vs. 1.43; AHRQ SES: OR 1.34 vs. 1.2). This indicates that some of the relationship between hospital visits and social risk is accounted for by the final risk model variables, including clinical comorbidities. However, after the addition of the final model variables, odds ratios for both social risk factors remain significantly above 1.

**Table 6A: HOPDs: Odds ratios for DE and AHRQ SES social risk factors in a bivariate vs. multivariate model**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Bivariate** |  |  | **Multivariate** |  |  |
|  | **Odds ratio** | **95% CI** | **P-value** | **Odds ratio** | **95% CI** | **P-value** |
| **DE (Yes vs No)** | 1.98 | 2.05 - 1.92 | <0.0001 | 1.43 | 1.48 - 1.39 | <0.0001 |
| **AHRQ SES (lowest quartile vs. all others)** | 1.35 | 1.38 - 1.31 | <0.0001 | 1.20 | 1.23 - 1.16 | <0.0001 |

**For ASCs**,in the bivariate models both social risk factors have an odds ratio greater than one, indicating patients with the social risk factor have an increased risk of the outcome (see Table 6B). When we included these variables in a multivariate model that includes all of the final risk model variables, the odds ratios for both the dual eligible and AHRQ SES variables in the multivariate model were lower than the odds ratio for the bivariate association (Table 6B; DE: OR 1.67 vs. 1.27**;** AHRQ SES: OR 1.35 vs. 1.21). This indicates that some of the relationship between hospital visits and social risk is accounted for by the final risk model variables, including clinical comorbidities. However, after the addition of the final model variables, odds ratios for both social risk factors remain significantly above 1.

**Table 6B: ASCs: Odds ratios for DE and AHRQ SES social risk factors in a bivariate vs. multivariate model**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Bivariate** |  |  | **Multivariate** |  |  |
|  | **Odds ratio** | **95% CI** | **P-value** | **Odds ratio** | **95% CI** | **P-value** |
| **DE (Yes vs No)** | 1.67 | 1.75 - 1.59 | <0.0001 | 1.27 | 1.33 - 1.21 | <0.0001 |
| **AHRQ SES (lowest quartile vs. all others)** | 1.35 | 1.39 - 1.31 | <0.0001 | 1.21 | 1.25 - 1.18 | <0.0001 |

**Analysis #4:** To understand the effect of each risk factor in the performance and predictive ability of each the risk adjustment model, we compared the c-statistic with and without the addition of each of the social risk factors.

**For HOPDs**, the results shown below in Table 7A indicate that entering these (dual eligible, and low AHRQ SES index) variables into the risk-adjustment model does not meaningfully improve model performance.

**Table 7A: HOPDs: Comparing C-statistics for risk adjustment models with and without social risk factors**

|  |  |  |
| --- | --- | --- |
|  | **HOPDs: C-statistic (model with social risk factor)** | **HOPDs: C-statistic (model without social risk factor)** |
| **DE** | 0.687 | 0.684 |
| **AHRQ SES Index** | 0.685 | 0.684 |

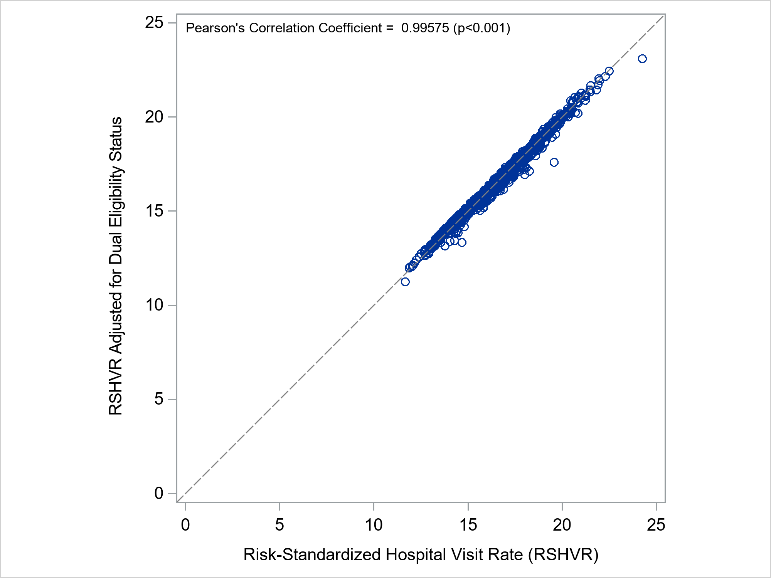
**For ASCs**, similarly, the results shown below in Table 7B indicate that entering these (dual eligible, and low AHRQ SES Index) variables into the risk-adjustment model does not improve model performance (C-statistics change minimally).

**Table 7B: ASCs: Comparing C-statistics for risk adjustment models with and without the social risk factor**

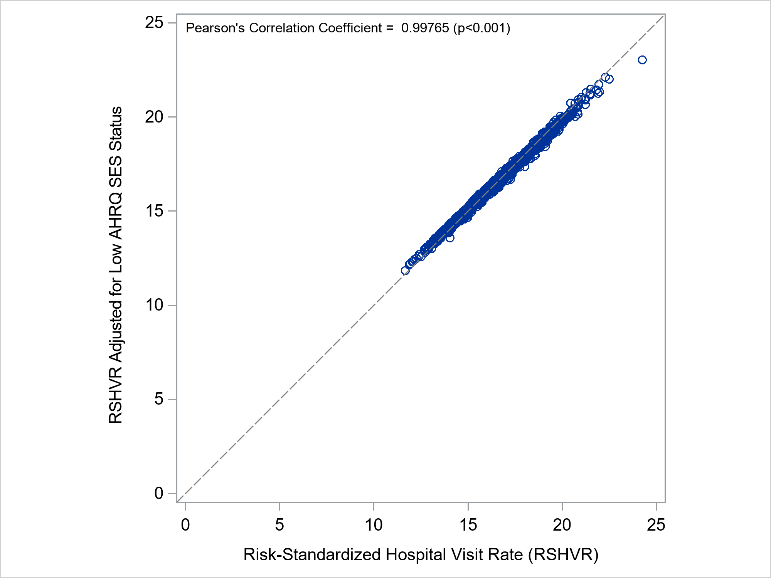
|  |  |  |
| --- | --- | --- |
|  | **ASCs: C-statistic (model with social risk factor)** | **ASCs: C-statistic (model without social risk factor)** |
| **DE** | 0.654 | 0.653 |
| **AHRQ SES** | 0.654 | 0.653 |

**Analysis #5: Impact of social risk factors on measure scores**To evaluate how social risk factors affect the measure score of individual facilities, we compared RSHVRs calculated for each facility with and without each social risk factor included in the model. For these analyses we calculated Pearson correlation coefficients for the paired scores. We also show scatter plots for these same analyses. We limited these analyses to facilities with at least 30 cases, which is the public reporting cut-off; only facilities that have at least 30 cases over a 3-year performance period have a publicly-reported RSHVR (discussed earlier in section 2a2.2).

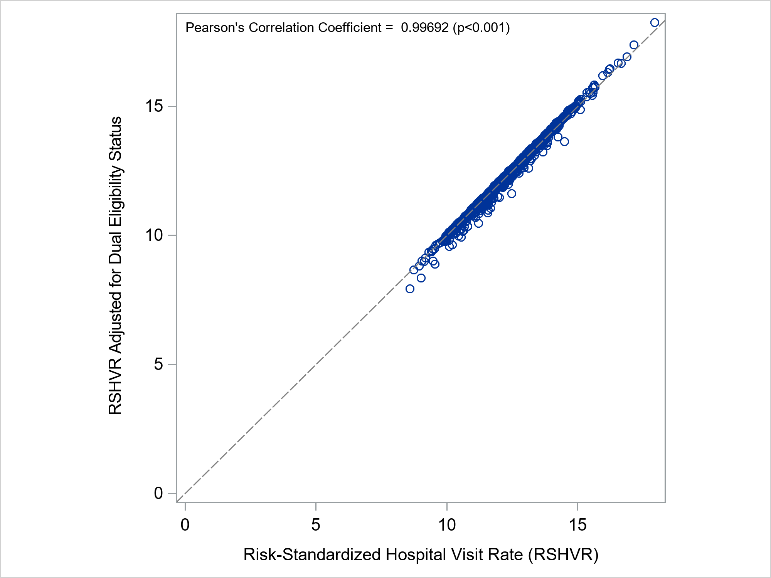
**For HOPDs** (Figures 1A and 1B), the results show that entering either of these variables into the risk-adjustment model did not substantially change hospital-level measure scores (RSHVRs). Correlation coefficients between RSHVRs with and without adjustment for these factors were near 1 (0.996 for dual-eligible, 0.998 for low SES patients). This indicates that including the DE and low AHRQ SES Index social risk factors in the risk model resulted in limited differences in facilities’ measure scores after accounting for other factors (age, comorbidities) included in the risk model.

**Figure 1A: HOPDs: Correlation of measure scores (RHSVRs) calculated with and without social risk factor adjustment for DE status (for facilities with at least 30 cases)**  


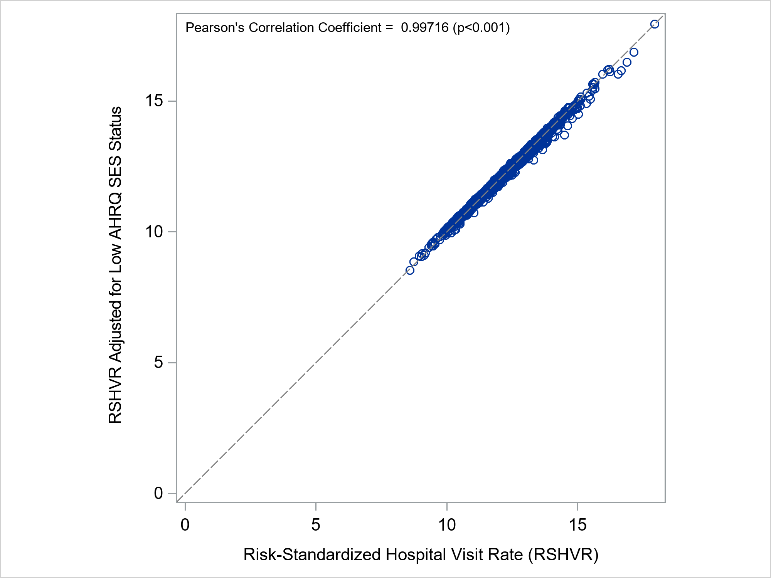
**Figure 1B. HOPDs: Correlation of measure scores (RHSVRs) calculated with and without social risk factor adjustment for low AHRQ SES (for facilities with at least 30 cases).**



**For ASCs** (Figure 2A and 2B), the results similarly show that entering either of these variables into the risk-adjustment model did not substantially change facility-level measure scores (RSHVRs). Correlation coefficients between RSHVRs with and without adjustment for these factors were near 1 (0.997 for dual-eligible, 0.997 for low SES patients). This indicates that including the DE and low AHRQ SES Index social risk factors in the risk model resulted in limited differences in facilities’ measure scores after accounting for other factors (age, comorbidities) included in the risk model.

**Figure 2A: ASCs: Correlation of measure scores (RHSVRs) calculated with and without social risk factor adjustment for DE status (for facilities with at least 30 cases)**

**Figure 2B: ASCs: Correlation of measure scores (RHSVRs) calculated with and without social risk factor adjustment for low AHRQ SES (for facilities with at least 30 cases).**



**Analysis #6: Comparison of RSHVRs between facilities with highest and lowest proportion of patients with social risk factors**  
  
Distributions of the measure score for facilities with a low proportion of patients with social risk factors (1st quartile) and high proportion of patients with social risk factors (4th quartile) by each social risk factor are shown in Table 8A for HOPDs and Table 8B for ASCs. The results showed higher measure scores for the 4th quartile (facilities with higher proportions of patients with the social risk factors) compared to the 1st quartile, but the distributions largely overlapped. **For HOPDs**, the median RSHVR varied minimally across quartiles of the proportion of patients with social risk factors (1st vs 4th quartiles) for both variables (DE: 16.2 vs 16.5; Low AHRQ SES: 16.2 vs. 16.6) (Table 8A). **For ASCs**, the median also varied minimally across quartiles (1st vs. 4th quartiles) for both variables (DE: 12.1 vs 12.3; Low AHRQ SES: 12.0 vs. 12.3).

**Table 8A: HOPDs: Comparison of measure scores (RHSVR) across the distribution, between 1st and 4th quartile of the proportion of patients with the social risk factor (DE and Low AHRQ SES)** (for facilities with at least 30 cases)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Dual eligible** | | **Low AHRQ SES** | |
| **1st Quartile**  **(<=2.94%)** | **4th Quartile**  **(>9.89%)** | **1st Quartile**  **(<=5.38%)** | **4th Quartile (>26.47%)** |
|
| **Number of HOPDs** | 894 | 895 | 896 | 894 |
| **Number of patients** | 768,473 | 336,342 | 659,707 | 307,490 |
| **Maximum RSHVR\*** | 21.52 | 24.27 | 20.86 | 24.27 |
| **90th** | 17.83 | 18.20 | 17.81 | 18.31 |
| **75th** | 16.98 | 17.29 | 16.93 | 17.35 |
| **Median** | 16.17 | 16.53 | 16.19 | 16.56 |
| **25th** | 15.42 | 15.89 | 15.50 | 15.95 |
| **10th** | 14.39 | 15.30 | 14.40 | 15.47 |

\*RSHVRs are per 1,000 colonoscopies

**Table 8B: ASCs: Comparison of measure scores (RHSVR) across the distribution, between 1st and 4th quartile of the proportion of patients with the social risk factor (DE and Low AHRQ SES)** (for facilities with at least 30 cases)

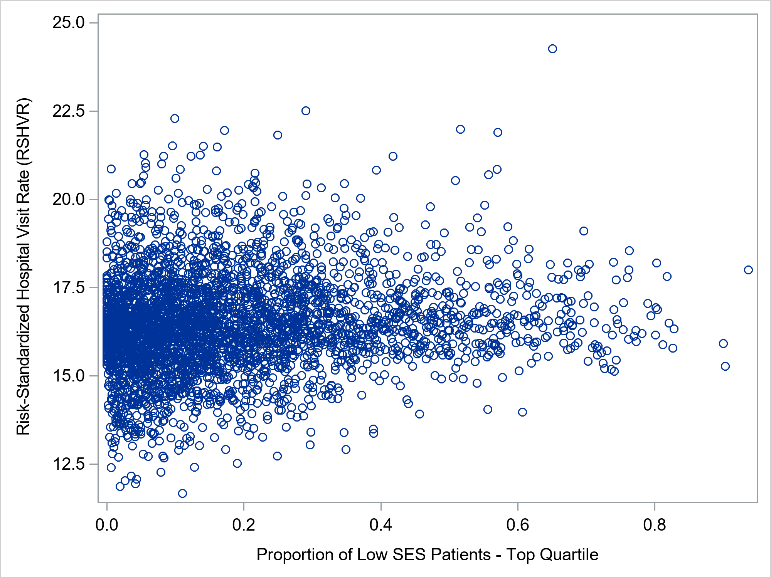
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Dual eligible** | | **Low SES** | |
| **1st Quartile**  **(<=1.09%)** | **4th Quartile**  **(>5.35%)** | **1st Quartile**  **(<=3.96%)** | **4th Quartile (>16.84%)** |
|
| **Number of ASCs** | 518 | 519 | 519 | 518 |
| **Number of patients** | 70,7563 | 393,510 | 665,512 | 488,590 |
| **Maximum RSHVR\*** | 16.02 | 17.15 | 16.20 | 17.15 |
| **90th** | 13.26 | 13.64 | 13.33 | 13.76 |
| **75th** | 12.68 | 12.86 | 12.59 | 13.04 |
| **Median** | 12.08 | 12.26 | 12.03 | 12.34 |
| **25th** | 11.58 | 11.76 | 11.45 | 11.79 |
| **10th** | 10.99 | 11.16 | 10.79 | 11.09 |
| **Minimum RSHVR** | 9.05 | 8.59 | 8.94 | 8.60 |

\*RSHVRs are per 1,000 colonoscopies

**Analysis #7: Relationship between RSHVR and percent of patients with social risk factors in facilities in the highest quartile for proportion of patients with the social risk factor.**    
  
Finally, for the quartile of facilities with the highest proportion of patients with social risk factors, we plotted the relationship between the proportion of a facilities’ patients with each risk factor (x-axis) and the ASC risk-standardized hospital visit rates (RSHVRs) (y-axis) in a scatter plot for the measure, and calculated the strength of the relationship between the facility-level measure score and the facility’s proportion of patients with social risk factors using the unweighted Spearman correlation coefficient (Figures 3 and 4, below).  For HOPDs and ASCs there was a weak positive correlation between the proportion of patients at the facility with DE, and proportion of patients with low SES status, and the measure score.

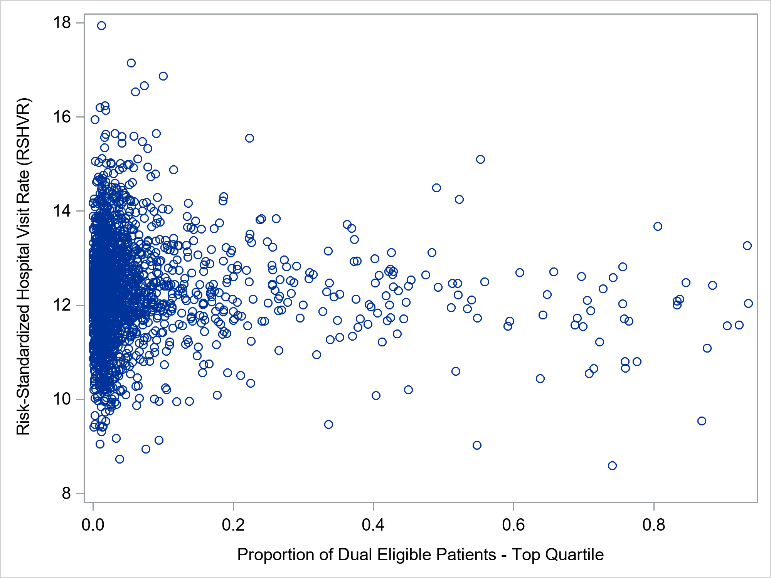
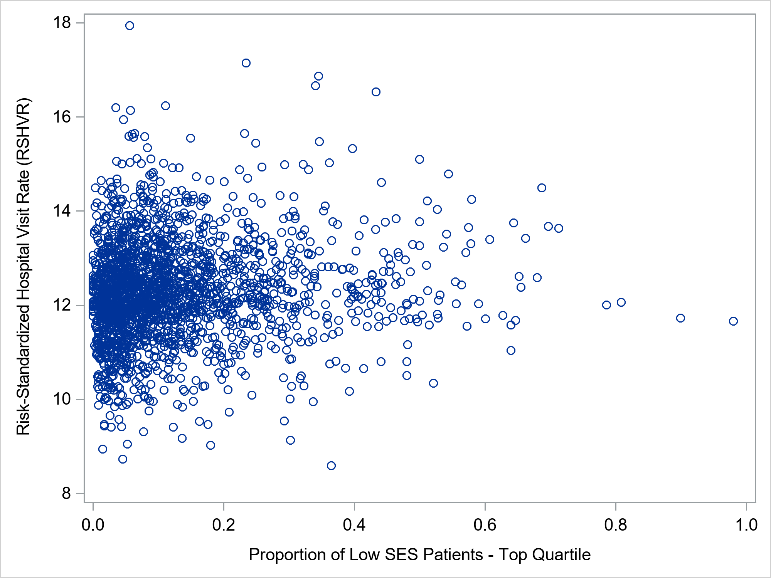
**Figure 3A and 3B: HOPDs: Relationship between the proportion of patients with dual-eligible status (A) and low AHRQ SES (B) and the risk-standardized hospital visit rates (RSHVRs)** (in facilities in the highest quartile for the proportion patients with the social risk factor; facilities with at least 30 cases).  
  
**Figure 3A: Dual Eligible**  **Figure 3B: Low AHRQ SES**

Spearman correlation coefficient: 0.126 Spearman correlation coefficient: 0.140



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

Spearman correlation coefficient: 0.057 Spearman correlation coefficient: 0.124



**Conclusion: Social Risk Factors**   
The analyses above show that DE patients and patients identified as low-SES using the AHRQ SES Index are at increased risk of post-colonoscopy hospital visits within seven days, even after adjusting for other risk factors in a multivariable model. However, the scores estimated for facilities with and without either social risk factor are highly correlated. Importantly, there is no meaningful or systematic increase in measure scores for facilities with the highest proportion of patients with social risk factors. Further, the absolute increase in the risk of a hospital visit for patients with either of the two social risk factors is low, given that the outcome rate for the measure in both settings is less than 2% (1.6% for HOPDs and 1.2% for ASCs), and the increase in risk as estimated by the odds ratios in multivariable models ranges from 1.2 to 1.4.

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

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

Given the testing results, CMS decided that on balance, the benefits of a measure that can illuminate the potential disparities for beneficiaries with the two social risk factors outweigh the concerns of fairness or unintended consequences of not adjusting for these. CMS therefore has decided not to adjust this measure for either DE or the AHRQ SES Index. CMS, however, is testing approaches to stratifying this measure by social risk factors under the IMPACT Act and will continue to assess the issue in measure reevaluation.

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

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.  
***If stratified, skip to*** [***2b3.9***](#question2b49)

We computed two summary statistics to assess model performance: the area under the receiver operating characteristic (ROC) curve (c-statistic) and the predictive ability.

A c-statistic of 1.0 indicates perfect prediction, implying patients’ outcomes can be predicted completely by their risk factors, and physicians and facilities play no role in patients’ outcomes. The c-statistic is an indicator of the model’s discriminant ability or ability to correctly classify those who did and did not have an unplanned hospital visit within 7 days of the colonoscopy. Potential values range from 0.5, meaning no better than chance, to 1.0, meaning perfect discrimination. Dataset #2 was used for this analysis.

To test model predictive ability, we calculated observed hospital visit rates in the lowest and highest deciles on the basis of predicted hospital visit probabilities. Dataset #2 was used for this analysis.

In addition, during the development of the original model, we calculated over-fitting indices in the Validation Sample. Over-fitting refers to the phenomenon in which a model describes the relationship between predictive variables and outcome well in the development datasets but fails to provide valid predictions in new patients. Estimated calibration values of γ0 far from 0 and estimated values of γ1 far from 1 provide evidence of over-fitting. Dataset #1 was used for this analysis.

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

**Table 9A. HOPDs: Colonoscopy Generalized Linear Model (Logistic Regression) Performance (January 1, 2016-December 31, 2018) (Dataset #2)**

| **Characteristic** | **Result** |
| --- | --- |
| c-statistic | 0.684 |
| Predictive ability, %  (lowest decile – highest decile) | 0.70-4.75 |

**Table 9B. ASCs: Colonoscopy Generalized Linear Model (Logistic Regression) Performance (January 1, 2016-December 31, 2018) (Dataset #2)**

|  |  |
| --- | --- |
| **Characteristic** | **Result** |
| c-statistic | 0.653 |
| Predictive Ability, %  (lowest decile - highest decile) | 0.59-3.11 |

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

Across risk deciles, using 2017 performance data (Dataset #1) the observed rates were accurately predicted (see calibration plots in 2b.3.8, below). Please note that while the model is recalibrated yearly, coefficients remain similar.

In addition, the results from original model/measure development are:

2010 Medicare 20% FFS Development Sample (Dataset #1a):

Calibration: (0,1)

2010 Medicare 20% FFS Validation Sample (Dataset #1b) results:

Calibration: (-0.03, 0.99)

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

**Figure 5A: HOPDs: Plot of observed vs. expected values for risk deciles (2017 performance period – Dataset #1) for HOPDs**

**Figure 5B: ASCs: Plot of observed vs. expected values for risk deciles (2017 performance period – Dataset #1) for**

**2b3.9. Results of Risk Stratification Analysis**:

Not applicable. This measure is not risk stratified.

**2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?** (i*.e., what do the results mean and what are the norms for the test conducted*)

**Discrimination Statistics**

The c-statistic of 0.684 for HOPDs, and 0.653 for ASCs, respectively, indicates good model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

**Calibration Statistics**

Over-fitting (Calibration γ0, γ1)

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

**Risk Decile Plots**

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

**2b3.11.** **Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

In Tables 10A for HOPDs, and 10B for ASCs, we include information on the consistency of data elements used in risk adjustment, showing the frequencies for all variables included in the final model. According to the results presented below, frequencies of the risk variables were similar across the time periods, indicating good variable consistency.

Tables 10A and 10B present the risk factor frequencies for HOPDs and ASCs individually for 2016, 2017, and 2018. The risk factor frequencies are very consistent with the original NQF endorsed measure and indicate that risk factor frequencies are stable over time.

**Table 10A: HOPDs: Risk Variable Frequencies: 2016, 2017, and 2018**

| **Risk Variable (CC)** | **01/2016-12/2016** | **01/2017-12/2017** | **01/2018-12/2018** |
| --- | --- | --- | --- |
| Concomitant Endoscopy | 18.89% | 19.06% | 19.55% |
| Polypectomy during Procedure | 37.94% | 39.07% | 40.59% |
| Congestive Heart Failure (CC 85) | 9.67% | 9.86% | 10.16% |
| Ischemic Heart Disease (CC 86-89) | 22.52% | 22.46% | 22.63% |
| Stroke/Transient Ischemic Attack (TIA) (CC 99-101) | 9.17% | 8.93% | 8.92% |
| Chronic Lung Disease (CC 111-113) | 18.50% | 18.81% | 18.98% |
| Metastatic Cancer (CC 8-11) | 9.89% | 9.83% | 9.82% |
| Liver Disease (CC 27-32) | 7.77% | 8.18% | 8.49% |
| Iron Deficiency Anemia (CC 49) | 24.36% | 24.24% | 24.34% |
| Disorders of Fluid, Electrolyte, Acid Base (CC 24) | 10.47% | 10.70% | 10.94% |
| Pneumonia (CC 114-116) | 5.05% | 5.23% | 5.32% |
| Psychiatric Disorders (CC 57-59, 61-63) | 17.68% | 18.58% | 19.47% |
| Drug and Alcohol Abuse/Dependence (CC 54-56) | 7.06% | 7.39% | 7.92% |
| Arrhythmia (CC 96-97) | 20.12% | 20.37% | 20.94% |
| Age 65-69 | 34.97% | 34.33% | 33.63% |
| Age 70-74 | 32.31% | 33.85% | 34.44% |
| Age 75-79 | 20.55% | 20.35% | 20.76% |
| Age 80-84 | 8.81% | 8.40% | 8.27% |
| Age 85+ | 3.36% | 3.07% | 2.91% |

With the exception of concomitant endoscopy and polypectomy during procedure, which we define using individual CPT® codes, we define comorbidity variables using CMS Condition Categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9 and ICD-10 diagnosis codes. Risk-factor definitions in this table are based on the v22 CC definitions, which can be found in the attached Data Dictionary in Tab 5 and Tab 6.

**Table 10B. ASCs: Risk Variable Frequencies, 2016, 2017, 2018**

| **Variable (CC)** | **01/2016-12/2016** | **01/2017-12/2017** | **01/2018-12/2018** |
| --- | --- | --- | --- |
| Concomitant Endoscopy | 17.50% | 17.04% | 17.32% |
| Polypectomy during Procedure | 38.36% | 39.26% | 40.61% |
| Congestive Heart Failure (CC 85) | 5.70% | 5.59% | 5.58% |
| Ischemic Heart Disease (CC 86-89) | 19.08% | 18.83% | 18.68% |
| Stroke/Transient Ischemic Attack (TIA) (CC 99-101) | 8.45% | 8.07% | 7.95% |
| Chronic Lung Disease (CC 111-113) | 13.87% | 13.83% | 13.77% |
| Metastatic Cancer (CC 8-11) | 8.17% | 7.92% | 7.86% |
| Liver Disease (CC 27-32) | 6.69% | 6.96% | 7.09% |
| Iron Deficiency Anemia (CC 49) | 20.84% | 20.68% | 20.51% |
| Disorders of Fluid, Electrolyte, Acid Base (CC 24) | 7.58% | 7.53% | 7.61% |
| Pneumonia (CC 114-116) | 3.42% | 3.53% | 3.55% |
| Psychiatric Disorders (CC 57-59, 61-63) | 13.76% | 14.40% | 15.52% |
| Drug and Alcohol Abuse/Dependence (CC 54-56) | 4.82% | 4.94% | 5.11% |
| Arrhythmia (CC 96-97) | 15.51% | 15.39% | 15.52% |
| Age 65-69 | 36.02% | 35.08% | 34.29% |
| Age 70-74 | 33.86% | 35.67% | 36.20% |
| Age 75-79 | 20.44% | 20.27% | 20.81% |
| Age 80-84 | 7.57% | 7.12% | 7.01% |
| Age 85+ | 2.11% | 1.86% | 1.69% |

With the exception of concomitant endoscopy and polypectomy during procedure, which we define using individual CPT® codes, we define comorbidity variables using CMS Condition Categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9 and ICD-10 diagnosis codes. Risk-factor definitions in this table are based on the v22 CC definitions, which can be found in the attached Data Dictionary in Tab 5 and Tab 6.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

**2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified** (*describe the steps―do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The measure score is a facility-level risk-standardized hospital visit rate (RSHVR). The RSHVR is calculated as the ratio of the predicted to the expected number of unplanned hospital visits among a facility’s qualifying colonoscopy procedures, multiplied by the national observed rate of unplanned hospital visits. For each facility, the numerator of the ratio is the number of hospital visits predicted for the facility’s procedures, accounting for its observed rate and patient case mix. The denominator is the number of hospital visits expected nationally for the facility’s case mix. To calculate a facility’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 patient demographics, patient comorbidities, and a random facility-specific intercept. A ratio greater than one indicates that the facility’s patients have more visits than expected, compared to an average facility with similar case mix. A ratio less than one indicates that the facility’s patients have fewer post-surgical visits than expected, compared to an average facility with similar case mix. More details on the measure score calculation can be found in the measure technical report: <https://www.qualitynet.org/outpatient/measures/colonoscopy/methodology>

We characterize the degree of variation by:

1) Providing the median odds ratio (MOR) [1]. The MOR represents the median increase in odds of a hospital visit if a procedure on a single patient was performed at a higher-risk facility compared to a lower-risk facility. It is calculated by taking all possible combinations of facilities, always comparing the higher risk facility to the lower risk facility. The MOR is interpreted as a traditional odds ratio would be.

2) Reporting the distribution of the RSHVR.

3) Reporting measure outliers. We use re-sampling and simulation techniques (bootstrapping) to derive an interval estimate to determine if a facility is performing better than, worse than, or no different from its expected rate. A facility is considered better than expected if its entire confidence interval falls below the expected rate, and considered worse if the entire confidence interval falls above the expected rate. It is considered no different if the confidence interval overlaps the expected rate. Full details of the bootstrapping procedure can be found in the measure technical report: <https://www.qualitynet.org/outpatient/measures/colonoscopy/methodology>.

All analyses were performed using Dataset #2.

Citations

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

**2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities?** (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

All analyses below are based on Medicare FFS data from the 2016-2018 performance period (three years of performance data) (Dataset #2).

**HOPDs**

The median odds ratio was 1.19.

The risk-standardized measure scores (RSHVRs) for 4,034 HOPDs estimated using Medicare FFS data (2016-2018 performance period) had a median value of 16.4 hospital visits per 1,000 colonoscopies. The values ranged from 11.7 to 24.3. The percentiles of the distribution are shown in Table 11. Figure 6 shows a histogram of the distribution.

**Table 11. HOPDs: Distribution of Risk-Standardized Hospital Visit Rates**

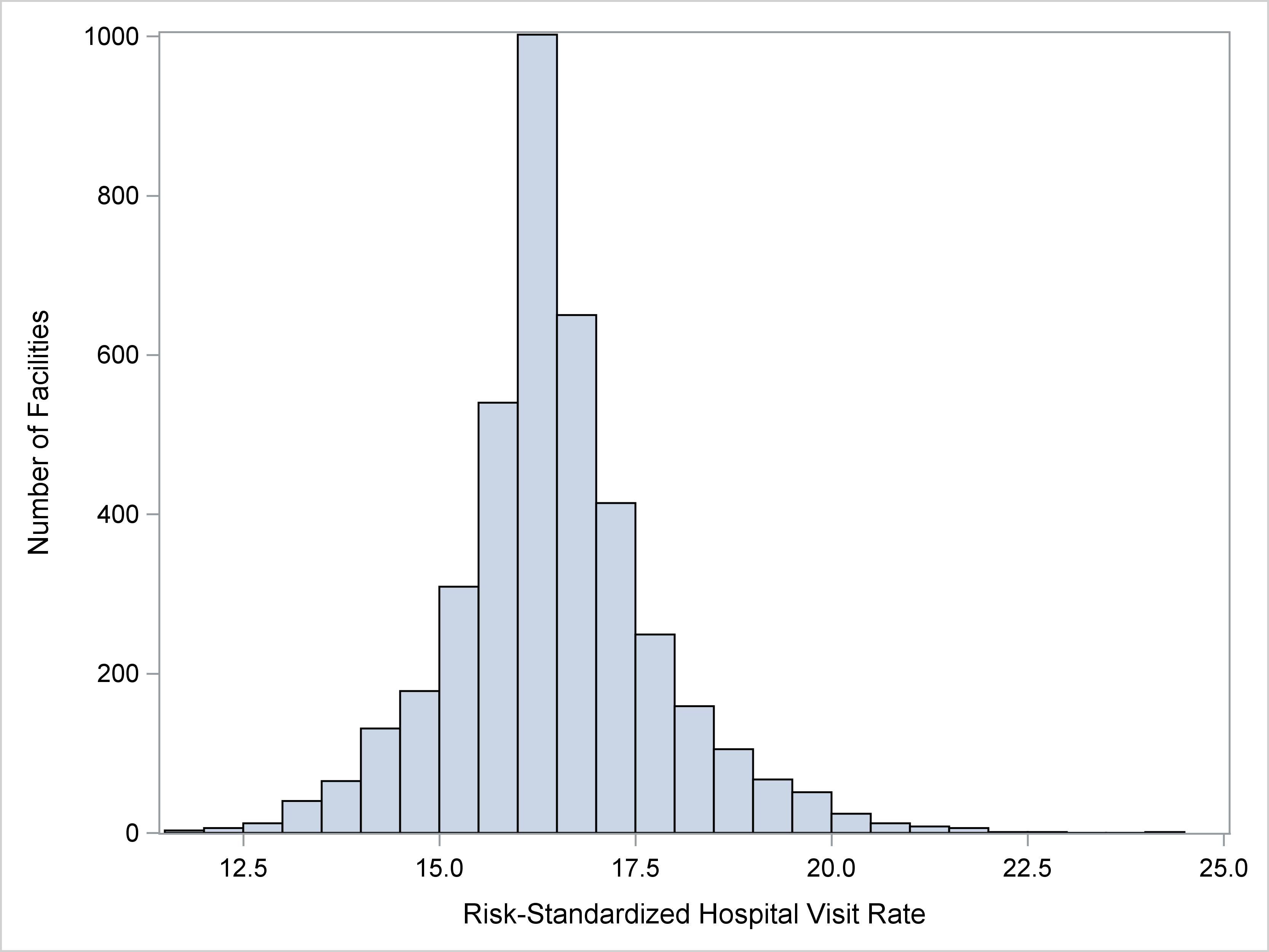
|  |  |
| --- | --- |
| **Characteristic** | **Value** |
| Number of facilities | 4034 |
| Mean RSHVR\* (SD) | 16.47 (1.32) |
| Range (min – max) | 11.67 - 24.27 |
| 10th percentile | 14.92 |
| 25th percentile | 15.76 |
| 50th percentile (median) | 16.38 |
| 75th percentile | 17.10 |
| 90th percentile | 18.10 |

Results based on January 1, 2016 -December 31, 2018, performance period data.

SD=standard deviation

\*RSHVRs are per 1,000 colonoscopies

**Figure 6. HOPDs: Distribution of Risk-Standardized Hospital Visit Rates**



Results based on January 1, 2016 -December 31, 2018, performance period data

**ASCs**

The median odds ratio was 1.18.

The RSHVRs for ASCs estimated using Medicare FFS data (2016-2018 performance period) had a median value of 12.23 hospital visits per 1,000 colonoscopies. The values ranged from 8.59 to 17.94. The percentiles of the distribution are shown in Table 12. Figure 7 shows a histogram of the distribution.

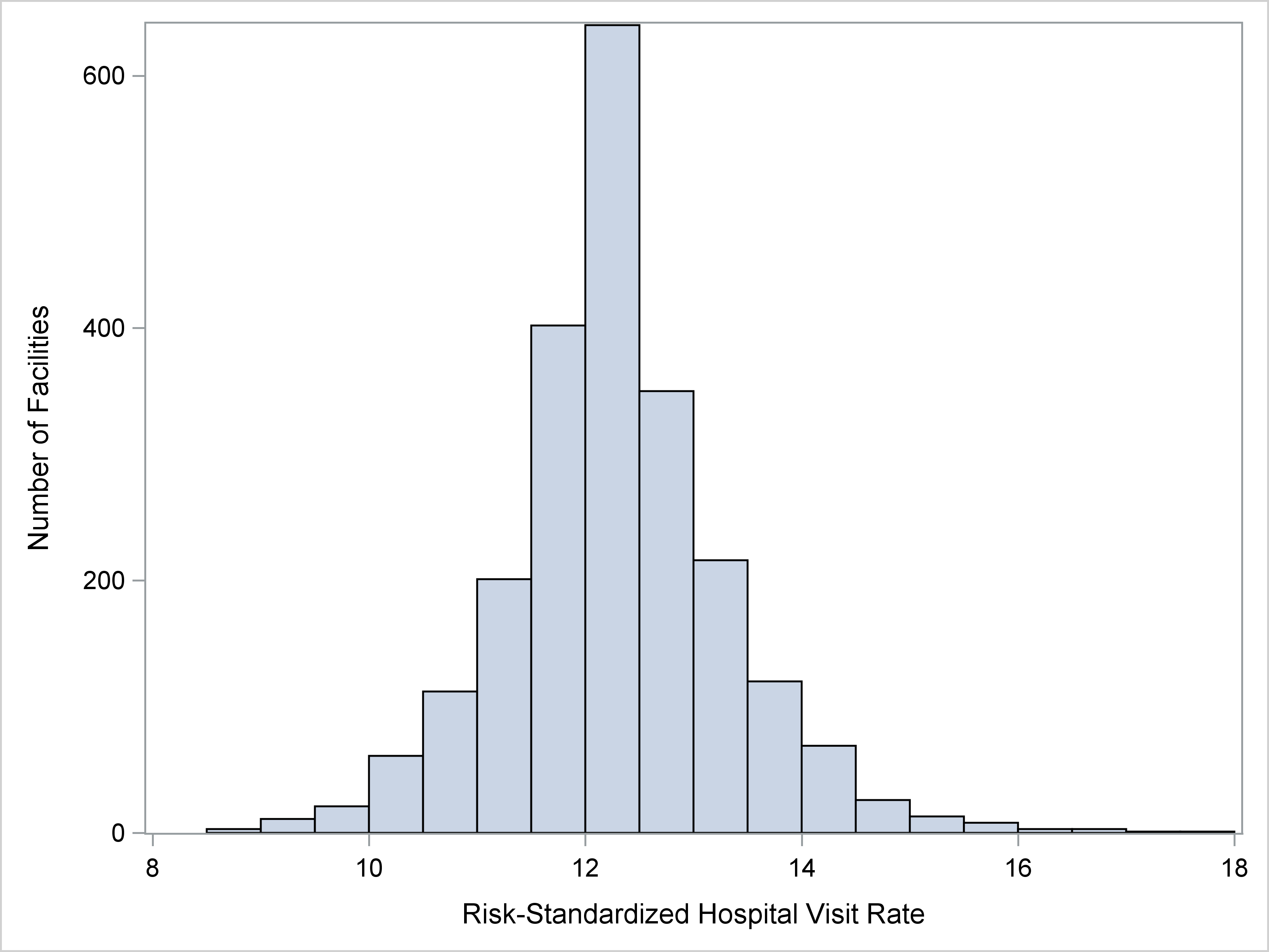
**Table 12. ASCs: Distribution of Risk-Standardized Hospital Visit Rates**

|  |  |
| --- | --- |
| **Characteristic** | **Value** |
| Number of facilities | 2261 |
| Mean RSHVR\* (SD) | 12.29 (1.03) |
| Range (min – max) | 8.59 - 17.94 |
| 10th percentile | 11.07 |
| 25th percentile | 11.75 |
| 50th percentile (median) | 12.23 |
| 75th percentile | 12.82 |
| 90th percentile | 13.57 |

Results based on January 1, 2016 -December 31, 2018, performance period data. SD=standard deviation

\*RSHVRs are per 1,000 colonoscopies

**Figure 7. ASCs: Distribution of Risk-Standardized Hospital Visit Rates**



Outliers

Applying the approach to identifying outliers described above, we found that of 4,034 HOPD facilities in the study cohort, 11 performed “Better than the National Rate,” 3,562 performed “No Different than the National Rate,” and 10 performed “Worse than the National Rate.” 451 were classified as “Number of Cases Too Small” (fewer than 30) to reliably tell how well the hospital is performing.

Of 2,261 ASC facilities in the study cohort, 15 performed “Better than the National Rate,” 2,042 performed “No Different than the National Rate,” and 16 performed “Worse than the National Rate.” 188 were classified as “Number of Cases Too Small” (fewer than 30) to reliably tell how well the ASC is performing.

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

The median odds ratios (MORs) suggest a meaningful increase in the risk of a hospital visit if a procedure was performed at a higher-risk facility compared to a lower-risk facility. Both MORs indicate that the impact of quality on the outcome rate is substantial at both HOPDs and ASCs.

* For HOPDs, a value of 1.19 indicates that a patient has a 19% increase in the odds of a hospital visit if the same procedure was performed at higher-risk HOPD compared to a lower-risk HOPD.
* For ASCs, a MOR of 1.18 indicates that a patient has a 18% increase in the odds of a hospital visit if the same procedure was performed at higher-risk ASC compared to a lower-risk ASC.

The distribution of measure scores also indicates that there is substantial variation in performance among both HOPDs and ASCs.

* **Among HOPDs**, the median RSHVR is 16.4 hospital visits per 1,000 colonoscopies, which indicates that patients undergoing colonoscopy at a facility performing at the median are expected to have an ED visit, observation stay, or admission to the hospital within 7 days 1.64% of the time.
  + The 10th and 90th percentiles (14.9 and 18.1 hospital visits per 1,000 colonoscopies, respectively) represent meaningful deviations from the median: a facility performing at the 10th percentile is performing about 9% better than an average performer, and a facility performing at the 90th percentile is performing about 11% worse than an average performer.
  + Furthermore, the best performing facilities (11.7 hospital visits per 1,000 colonoscopies) are performing 29% better than the median performer, while the worst (24.3 hospital visits per 1,000 colonoscopies) are performing 48% worse than the median performer.
* **Among ASCs**, the median RSHVR is 12.2 hospital visits per 1,000 colonoscopies, which indicates that patients undergoing colonoscopy at a facility performing at the median are expected to have an ED visit, observation stay, or admission to the hospital within 7 days 1.22% of the time.
  + The 10th and 90th percentiles (11.1 hospital visits and 13.6 hospital visits per 1,000 colonoscopies, respectively) represent meaningful deviations from the median: a facility performing at the 10th percentile is performing 9.5% better than a median performer, while a facility performing at the 90th percentile is performing nearly 11% worse than a median performer.
  + The best performing ASCs (8.6 hospital visits per 1,000 colonoscopies) are performing 35% better than a median performer, while the worst performing ASCs (17.9 hospital visits per 1,000 colonoscopies) are performing 47% worse than a median performer.

This variation in performance shows a clear quality gap, as some facilities can achieve substantially lower rates than the median performer, while other facilities are performing worse than the median performer. It is important to note that here the median performer refers to a facility with the same case mix performing at the median.

Finally, we identified relatively few outliers, which is expected given the measure’s low outcome rate and conservative 95% CIs. This, however, does not diminish the importance of the measure; we observed substantial variance in both observed and risk-adjusted rates among facilities.  Identifying those facilities that are outliers with a very high degree of confidence using the 95% CI can be informative to consumers and facilities.

In summary, this measure provides transparent data to facilities, allowing them to see their rates and reasons for return to the hospital. This invaluable data can be used to reduce negative patient outcomes and provide better quality. Overall, our results suggest that there is substantial need to reduce the variation in rates across HOPDs and ASCs, and that this improvement goal is achievable.

**\_\_\_\_\_\_\_\_\_\_\_**

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

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

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

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

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

Not applicable.

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

Not applicable

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

Not applicable.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

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

We did not perform an analysis of missing data for the measure because it is based on a 100% sample of paid, final action claims submitted by facilities for payment. To ensure complete claims, we allow at least 3 months of time between accessing the data and the end of the performance period.

**2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data?** (*e.g.,**results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Not applicable.

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

Not applicable.