NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

Purple text represents the responses from measure developers. Red text denotes developer information has changed since the last measure evaluation review. Some content in the document is from Measure Developers.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2539

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

Measure Steward: Centers for Medicare & Medicaid Services

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

Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about facility-level 7-day, risk-standardized hospital visit rates following outpatient colonoscopy.

Colonoscopy is a common and costly procedure performed at outpatient facilities and is frequently performed among relatively healthy patients to screen for colorectal cancer (CRC). Between January 1, 2016 and December 31, 2018, there were 2,258,661 colonoscopies performed in non-federal acute care hospital outpatient departments (HOPDs) and 2,524,898 performed in ambulatory surgical centers (ASCs). Given the widespread use of colonoscopy, understanding and minimizing procedure-related adverse events is a high priority. These adverse events, such as abdominal pain, bleeding, and intestinal perforation, can result in unanticipated hospital visits post procedure, and as outlined in the evidence attachment, a majority (68% in one study) of the reasons for emergency department visits following outpatient colonoscopy are due to the colonoscopy. Furthermore, physicians performing colonoscopies are often unaware that patients seek acute care at hospitals following the procedure and thus underestimate such events. This risk-standardized quality measure addresses this information gap and promotes quality improvement by providing feedback to facilities and physicians, as well as transparency for patients on the rates of and variation across facilities in unplanned hospital visits after colonoscopy.

Numerator Statement: Unplanned hospital visits within 7 days of a qualifying colonoscopy

Denominator Statement: Colonoscopies performed at hospital outpatient departments (HOPDs) and ambulatory surgical centers (ASCs) for Medicare FFS patients aged 65 years and older

Denominator Exclusions: We established the following exclusion criteria after reviewing the literature, examining existing measures, discussing alternatives with the working group and technical expert panel (TEP) members, reviewing feedback from the national dry run held in July 2015, and public reporting in 2018 and 2019, and annual re-evaluation of the measure in 2017, 2018, and 2019. The goal was to be as inclusive as possible; we excluded only those high-risk procedures and patient groups for which risk adjustment would not be adequate or for which hospital visits were not typically a quality signal. The exclusions, based on clinical rationales, prevent unfair distortion of performance results.

1) Colonoscopies for patients who lack continuous enrollment in Medicare FFS Parts A and B in the 7 days after the procedure.

Rationale: We exclude these patients to ensure full data availability for outcome assessment.

2) Colonoscopies that occur concurrently with high-risk upper gastrointestinal (GI) endoscopy procedures.

Rationale: Patients undergoing concurrent high-risk upper GI endoscopy procedures, such as upper GI endoscopy procedures for the control of bleeding or treatment of esophageal varices, and have a higher risk profile than typical colonoscopy patients. Therefore, these patients have a disproportionally higher risk for the outcome.

3) Colonoscopies for patients with a history of inflammatory bowel disease (IBD) or diagnosis of IBD at time of index colonoscopy or on the subsequent hospital visit outcome claim.

Rationale: We exclude these patients because:

- IBD is a chronic condition; patients with IBD undergo colonoscopy both for surveillance due to increased cancer risk and for evaluation of acute symptoms. IBD is likely to be coded as the primary diagnosis prompting the procedure irrespective of whether the patients are undergoing a screening procedure or a diagnostic procedure in the setting of an acute exacerbation of IBD. Therefore, we may not be able to adequately risk adjust for these patients, as we cannot identify relatively well versus acutely unwell patients among visits coded as IBD.
- Our aim is to capture hospital visits which reflect the quality of care. Admissions for acutely ill IBD patients who are evaluated with an outpatient colonoscopy and are subsequently admitted for medical treatment of an IBD flare do not reflect the quality of the colonoscopy. In our 2010 Medicare 20% FFS Full Development Sample (see the 2014 Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy Measure Technical Report posted at https://www.qualitynet.org/files/5d0d37ae764be766b010196e?filename=ClnscpyMsr_TechReport.pdf for full description of the dataset), more than one-third of IBD patients admitted to the hospital with colonoscopy had a discharge diagnosis of IBD, indicating their admission was for medical treatment of their IBD. We therefore excluded this group so that providers who treat a disproportionate number of IBD patients will not be disadvantaged in the measure.
- A post-index diagnosis of IBD, which represents a very small fraction of cases (less than 0.5% of the cohort) in the measure population, indicates that the condition was likely present at the time of the index colonoscopy but not coded.

4) Colonoscopies for patients with a history of diverticulitis or diagnosis of diverticulitis at time of index colonoscopy or on the subsequent hospital visit outcome claim.

Rationale: We exclude these patients because:

- It is unclear what the health status is of patients coded with a history or current diagnosis of diverticulitis, making it difficult to fully risk adjust for patients' health. Colonoscopies performed on patients with a history or current diagnosis of diverticulitis are likely to be coded as diverticulitis as the primary diagnosis irrespective of whether the patients are undergoing a screening procedure or a diagnostic procedure (i.e., are acutely unwell with active disease). Furthermore, the codes for diverticulitis and diverticulosis may not be consistently used; patients with diverticulosis may be erroneously coded as diverticulitis. Therefore, we

may not be able to adequately risk adjust as we cannot identify relatively well versus acutely unwell patients among visits coded as diverticulitis.

- Admissions for acutely ill patients with a history or current diagnosis of diverticulitis who are evaluated with an outpatient colonoscopy and are subsequently admitted for medical treatment of do not reflect the quality of the colonoscopy. In our 2010 Medicare 20% FFS Full Development Sample (see the Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy Measure Technical Report posted on the web page provided in data field S.1) more than one-quarter of patients with a history or current diagnosis of diverticulitis admitted to the hospital post colonoscopy had a discharge diagnosis of diverticulitis, indicating they were admitted for medical treatment of the condition. These admissions are likely unrelated to the quality of the colonoscopy. We therefore excluded this group so that providers who treat a disproportionate number of diverticulitis patients will not be disadvantaged in the measure.
- A post-index diagnosis of diverticulitis, which represents a very small fraction of cases (less than 0.5% of the cohort) in the measure population, indicates that the condition was likely present at the time of the index colonoscopy but not coded.

5) Colonoscopies followed by a subsequent outpatient colonoscopy procedure within 7 days.

Rationale: In these situations, the two colonoscopies are considered part of a single episode of care, for which the subsequent colonoscopy is considered the index procedure.

In addition, for colonoscopies performed at HOPDs, we exclude:

6) Colonoscopies that occur on the same day and at the same hospital as an emergency department (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.

Rationale: It is unclear whether the colonoscopy or ED visit occurred first. If the ED visit is coded with a diagnosis indicative of a complication of care, the measure assumes the ED visit occurred after the colonoscopy procedure and is counted in the measure. It is unlikely that a patient would experience an ED visit for an acute diagnosis at 1 facility and then travel to another facility for a routine colonoscopy on the same day. Accordingly, ED visits billed on the same day as a colonoscopy but at a different facility are included because they likely represent a routine procedure followed by a complication of care.

7) Colonoscopies that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.

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

8) Colonoscopies that are billed on the same hospital outpatient claim and that occur after the ED visit.

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

9) Colonoscopies that are billed on the same hospital outpatient claim as an observation stay.

Rationale: In these situations, it is not possible to use claims data to determine whether the colonoscopy was the cause of, subsequent to, or during the observation stay

Measure Type: Outcome

Data Source: Claims, Other

Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 23, 2014 **Most Recent Endorsement Date:** Dec 23, 2014

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

Summary of prior review in 2014

- The developer notes that gastrointestinal (GI) complications from colonoscopy are common and range from severe to mild – namely, GI bleeding, abdominal pain, distention, nausea, vomiting, and other non-GI complicaitons (e.g., hypotension, hypoxia). The developer notes that post-procedural infections can also occur.
- The developer cited studies in which the overall post-colonscopy hospital use, as measured as hospitalizations and ED visits, ranged from 1-2% within 30 days.
- The developer stated that providers are often unaware of complications for which patients return to the hospital, suggesting the need for measurement to drive quality improvement.
- Previously, the Standing Committee noted that colonoscopy is the most common procedure performed in the outpatient setting or ambulatory surgical centers (ASCs).
- The Committee also agreed with the evidence in support of the rationale. They noted that most patients return to the hospital with potentially preventable complications (e.g., abdominal pain, bleeding, perforation, aspiration because of the anesthesia).

Changes to evidence from last review

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

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

Updates:

- The developer states that many of the reasons for post-procedural hospital visits are related to a colonoscopy. The developer cites a 2018 study that found that of patients who experienced an ED visit

within 7 days of an outpatient colonoscopy 68% of the reasons for the ED visit were due to the colonoscopy.

- The developer cites studies that reported emergency department (ED) visit rates within 7 days of a colonoscopy of 0.76% and an average 7-day hospital visit rate (defined as an ED visit, observation stay, or inpatient hospitalization) of 1.63%.
- The developer states that provider- and facility-level factors can affect the outcome of complications and hospital visits related to a colonoscopy. The developer cites three studies the found that low provider volume and Fellow involvement in the procedure were significantly associated with a higher risk of an ED visit in one study, and low procedure volume was associated with a higher risk of infection in another study. Additionally, one study found the choice of sedation may influence complication rates – specifically, the use of anesthesia resulted in an increased risk of aspiration pneumonia.

Question for the Committee:

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

Guidance from the Evidence Algorithm

BOX 1: Measure an outcome (Yes) → BOX 2: Empirical evidence to support the relationship to a at least one structure or process (Yes) → PASS

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided score data for Hospital Outpatient Departments (HOPDs)
  - Distribution of measure scores for the colonoscopy measure for 2020 public reporting.
  - Data Source: Medicare FFS claims (Part A and B), January 1, 2016-December 31, 2018
  - Note: Sample includes all hospital outpatient departments with results
  - Distribution (percentiles) of the risk-standardized hospital visit rates (RSHVRs) per 1000 colonoscopies, all facilities (n = 4034)
    - Percentile//7-Day RSHVR
    - Min//11.67
    - P10//14.92
    - P25//15.76
    - P50//16.38
    - P75//17.10
    - P90//18.10
    - Max//24.27
    - Mean (SD)//16.47 (1.32)
- The developer provided score data for Ambulatory Surgical Centers (ASCs)
  - Data Source: Medicare FFS claims (Part A and B), January 1, 2016-December 31, 2018
  - Distribution (percentiles) of the risk-standardized hospital visit rates per 1000 colonoscopies, all facilities (n = 2,261)
  - Risk-standardized hospital visit rates per 1000 colonoscopies:
    - Percentile//7-Day RSHVR
    - Min//8.59

- P10//11.07
- P25//11.75
- P50//12.23
- P75//12.82
- P90//13.57
- Max//17.94
- Mean (SD)//12.29 (1.03)

Disparities

- The developer provides data demonstrating the measure's ability to identify performance gaps based on dual-eligibility and the AHRQ SES Index variables. However, the distribution of performance largely overlaps.

Questions for the Committee:

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

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

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

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- **New data supports the evidence base for this measure.**
- **Pass**
- **The evidence appears to relate directly to the outcome being measured.**
- **There are supporting data and actionable items to improve performance; I am not aware of other data sets since 2018**
- **meets**
- **Yes**
- **yes. There is a at least one intervention that can be made**
- **Evidence directly supports measure**
- **Evidence provided. No additional comments**
- **Yes**

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

- **Developer demonstrates quality problems and opportunities to improve.**
- **opportunity for improvement**
- **Performance data was provided and demonstrates opporunities to improve care to a degree that does warrant a national performance measure.**
- **Performance data were provided and a national measure is warranted.**
- **There is a high gap, developer accounted for disparities by stating equal overlap**
- **Yes, although lower in past few years**
- **Whether in ASCs or HOPDs, there is a significant performance gap**
- **continued opportunity to improve**
- **The distribution of hospital visit rates among HOPDs declined for 2019 reporting compared to 2018 reporting. This decline may reflect 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**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators:

- David Nerenz, PhD, Co-chair
- Sean O'Brien, PhD
- Lacy Fabian, PhD
- Marybeth Farquhar, PhD, MSN, RN
- Joseph Kunisch, PhD, RN-BC, CPHQ
- Sam Simon, PhD
- Alex Sox-Harris, PhD, MS
- Eric Weinhandl, PhD, MS
- Paul Kurlansky, MD

Methods Panel Review (Combined)

Scientific Methods Panel Votes

- Reliability: H-4; M-3; L-1; I-0 (Pass)
- Validity: H-1; M-4; L-1; I-2 (Pass)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

- In their preliminary analyses, the SMP passed the measure on reliability; however, consensus was not reached on validity.
- The SMP primarily raised concern with the developer's rationale for not providing empirical analyses of the validity testing for maintenance review. The developer provided a detailed written and verbal response to these concerns which facilitated a discussion with the panel on the other types of validity testing that could have been conducted other than what was described by the developers, and the feasibility of those testing approaches.
- With respect to empiric validity testing, the developer stated that they have previously examined the top reasons for return to the emergency department within 7 days of an outpatient colonoscopy, using an earlier version of the colonoscopy measure. The study found that 68% of ED visits were related to the colonoscopy.
- The developer's response can be found within the Standing Committee folder on Sharepoint.
- Ultimately, the panel voted to pass the measure on validity and this measure will be considered by the Admissions and Readmissions Standing Committee for the Spring 2020 cycle.

Reliability

- **Method(s) of reliability testing:**
  - Conducted at facility-level
  - Measure Score reliability testing was conducted at the data source and level of analysis indicated
  - The methods used were appropriate – the measure developer used the signal-to-noise method suggested by Adams et al. to assess reliability.
- **Reliability testing results:**
  - HOPDs – Hospital Outpatient Departments
    1. 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").
  - ASCs – Ambulatory Surgical Centers
    1. 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").

Validity

- **Method(s) of validity testing:**
  - Only face validity was conducted for this maintenance measure
    - The developers state that none of the existing measures are an appropriate comparator for validity testing.
- **Validity testing results:**
  - Face validity results indicated 71% of TEP members indicated at least moderate agreement that the is valid and 86% of TEP members indicated somewhat, moderately, or strongly agree.
  - None of the measures that the developer identified meet the criteria for a comparator measure that could be used for external validation. Therefore, only face validity was conducted
  - The developers claim that none of the existing measures are a fair comparator for validity testing. Members of the SMP recommended the developer consider the measure: "Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at ASCs (ASC General Surgery)" for facilities that have adequate volumes of target procedures.

    - The developer provided a response by stating that "many ASCs specialize in a single procedure" and that "few ASCs performing colonoscopies are the same facilities that would be measured in the ASC General Surgery measure."
- **Exclusions:**
    - No concerncs with exclusions indicated
- **Risk adjustment Summary:** Method – Statistical Modeling
    - Two-level hierarchical logistic regression model was used to estimate risk-standardized hospital visit rates (RSHVRs)
        - The c-stat of the adjustment model is modest -- 0.684 for HOPDs
        - The c-stat of the adjustment model is modest -- 0.654 for ASCs
    - The developer has chosen to leave both dual status and area SES out of the adjustment model, arguing that including them may mask disparities in care.
    - The developer points out that the results with and without the two variables are very highly correlated (.99), so that inclusion or exclusion of the two variables would not contribute to the performance of the risk adjustment model.
    - Lastly, the developer states that there was "no meaningful or systematic increase in measure scores for facilities with the highest proportion of patients with social risk factors."

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Is the rationale for not conducting empirical validity testing agreeable?*
- *Do you agree with the developer's decision, based on their analysis, to not include SES factors (e.g., race) and dual-eligible status in their risk-adjustment model?*
- *The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

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

Specifications precise unambiguous and complete (Box 1)→ Empirical reliability testing conducted (Box 2)→ Testing conducted at computed measure score level (Box 4)→ Method described and appropriate (Box 5) → Level of certainty or confidence that measure scores are reliable (Box 6) →MODERATE (rationale that reliability improves as the sample sizes increase, medium and small facilities have lower reliability estimates)

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

Specifications consistent with evidence (Box 1)→Potential threats to validity assessed (Box 2) →Empirical validity testing of measure as specified (Box 3) →Face validity was systematic (Box 4) → Level of certainty or confidence that measure score is a valid indicator of quality (Box 5) →Moderate

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

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

- **No concerns.**
- **None**
- **The algorithm logic and data elements appear to be clear.**
- **I think all data elements are defined.**
- **none**
- **none**
- **Currently in use**
- **No concerns**
- **Modifications have been made in the measure to align with specifications of similar measures and improve the accuracy of the algorithm**
- **The measure might give more reliable results for larger facilities.**

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

- **No concerns.**
- **none**
- **None.**
- **None**
- **moderate**
- **none**
- **three yrs data--facility level reliability scores with excellent agreement**
- **no concerns**
- **No**
- **Variation of results between small and large facilities.**

2b1. Validity -Testing: Do you have any concerns with the testing results?

- **No concerns.**
- **none**
- **None.**

- **I have no concerns, although I under the SMP did have concerns**
- **no**
- **none**
- **only face validity. SMP raised concerns re: absence of external empiric testing, esp for a measure already in use. Appreciate the developers' response and see their point**
- **no concerns**
- **hospital visits should be restricted to those that are directly related to the colonoscopy and represent complications (e.g., hemorrhage, perforation, abd pain, and complications of anesthesia). This would enhance the meaningfulness and actionable nature of the measure from the provider side.**
- **No**

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- **None identified.**
- **no**
- **Analyses indicate that this measure identifies meaningful differences about quality.**
- **No concerns**
- **pass**
- **No missing data that I can see**
- **did not seem to be issues with missing data**
- **no concerns**
- **No missing data**
- **2b6**

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure?2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- **Risk adjustment is logical though SES are not in the model.**

- **yes**
- **Exclusions are consistent with the evidence.**
- **None**
- **meets**
- **looks appropriate**
- **no issues re: exclusions. RE: social risk variables, I appreciate the presentation of results with and without inclusion (ie not much difference)**
- **Acceptable results**
- **yes**
- **2b2**

Criterion 3. Feasibility

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

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

- The developer states that all data elements are in defined fields in electronic claims.
- The developer also states that there have been no difficulties regarding data collection, availability of data, missing data, etc.

Questions for the Committee:

- None

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- **No concerns.**
- **none**
- **No current concerns regarding feasibility.**
- **I have no conerns**
- **none**
- **no concerns**
- **in use already, electronic data sources--highly feasible**
- **no concerns**
- **Claims based measure**
- **Yes**

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

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

- Ambulatory Surgical Center Quality Reporting (ASCQR) Program CMS
    - The ASCQR Program is a national pay-for-reporting, quality data program finalized by CMS under which ASCs report quality of care data for standardized measures to receive the full annual update to their ASC annual payment rate. Measured entities include all ambulatory surgical centers with eligible colonoscopies.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer states that in 2015, a "dry run" was performed, in which a preliminary analysis of data were reported to facilities so that they could review their measure results and ask questions about the measure and the methodology.
- The developer states that a variety of question topics were received with respect to the specific cases in facilities' data (40%), followed by requests for assistance accessing the FSR on the QualityNet website (23%), questions about the dry run process or the national provider calls (16%), and general methods questions (15%).
- The developer states that a number of situations were identified that suggested the need to make minor refinements to the measure methodology to ensure: (a) the algorithm for processing claims data accurately identifies cases for inclusion in the measure; and (b) the planned admission algorithm captures additional planned hospital visits.
- For 2019 public reporting, the developer states that measure results were reported to 3791 HOPDs and 1327 ASCs. For 2020 public reporting, measure results were reported to 4190 HOPDs and 1097 ASCs; reports were downloaded by 2915 HOPDS (69.6%) and 326 ASCs (29.7%).

Additional Feedback:

- The developer states that no other feedback was received from other users.

***Questions for the Committee**:*

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

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

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

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

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results:

- The developer states that hospital visit rates among HOPDs declined for 2019 reporting compared to 2018 (from 16.4 per 1000 cases in 2018 reporting to 14.8 per 1000 cases in 2019 reporting).
- The developer also states that the distribution of risk-standardized rates also declined for HOPDs; the interquartile range of rates for 2019 reporting lie completely below the 2018 interquartile range. However, the developer reports that hospital visit rates did not decline between 2019 and 2020 public

reporting. The developer states that "this can be attributed to a change in the measure's specifications that result in the use of three years of performance data that overlap with 2018 and 2019 performance periods."

- For ASCs, the developer states that there was a small decline in the hospital visit rates across the three public reporting years (2018, 2019, 2020).

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

Unexpected findings (positive or negative) during implementation

- The developer reports that no unexpected findings (benefites or harms) were encountered during implementation, including unintended impacts on patients.

Potential harms

- The developer reports that no unexpected findings (benefites or harms) were encountered during implementation, including unintended impacts on patients.

Additional Feedback:

- The developer reports that no additional feedback was received

***Questions for the Committee**:*

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

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

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

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- **Measure is being used in HOQR and ASCQR CMS Programs. Feedback has been obtained through a dry run and public reporting. Changes have been incorporated into the measure. .**
- **yes**
- **Feedback appears to have been incorporated into the measure as currently proposed.**
- **No concerns**
- **meets**
- **already in place for public reporting and accountability**
- **in use for quality/accountability programs, some feedback re: accuracy of claims and ensuring that planned hospital admissions were appropriately captured**
- **Currently used in programs**
- **Measure is publically reported**
- **Yes**

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

- **Some decline in hospital visit rates noted 2018 to 2019. No harms identified.**
- **none**
- **Yes, the performance results can be used to further the goal of high-quality, efficient healthcare.**
- **No concerns**
- **meets**
- **none**
- **has been improvement in hospitalizations/visits. No harms reported**
- **benefits outweigh potential harm**
- **Benefits outweight harm**
- **4b2**

Criterion 5: Related and Competing Measures

Related or competing measures

- 0658 : Appropriate Follow-Up Interval for Normal Colonoscopy in Average Risk Patients
- 2687 : Hospital Visits after Hospital Outpatient Surgery
- 3357 : Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers
- 3510 : Screening/Surveillance Colonoscopy

Harmonization

- The developer states that NQF 0034 focuses on colonoscopy screening in patients aged 50-75, therefore the targeted population overlaps with the CMS colonoscopy measure and reflects overall screening guidelines. The CMS colonoscopy outcome measure's purpose is to measure outcomes from colonoscopy procedures in Medicare-aged patients.
- The developer states that NQF 3510 has the same target population (Medicare beneficiaries) and would capture the physician-controlled costs related to hospital visits identified in the CMS colonoscopy measure. The timeframe for the two measures differs (7 days for the outcome measure vs. 14 days for the cost measure), and the level of measurement differs (facility-level for the outcome measure, and clinician or group level for the cost measure).
- The developer notes that NQF 3357 and NQF 2687 have the same outcome as CMS's colonoscopy measure presented in this re-endorsement application; an unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission. Additionally, the developer states that the patient cohort has no overlap with the colonoscopy measure, because they include patients undergoing surgical procedures, not colonoscopy.

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

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

- **There are related measures and they appear to be harmonized.**
- **I am not aware of any competing measures.**
- **none**
- **related, but not competing.**
- **As mentioned in the brief**
- **multiple related measures, current measure focuses on Medicare population rather than the > 50 population**
- **Harmonized**
- **None**
- **Not Sure**

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June 12, 2020

- **There have been no public comments or support/non-support choices as of this date.**

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2539

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

Type of measure:

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

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

Data Source:

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

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

☒☐ **Enrollment Data** ☐☒ **Other:** (Enrollment **– Panel Member #8; database files – Panel Member #5**)

Level of Analysis:

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

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

☐ **Integrated Delivery System** ☒☐ **Other: Ambulatory Surgical Centers; Hospital Outpatient Facilities (Panel Member #5) Ambulatory Surgical Center** (**Panel Member #4**)

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

Panel Member #1 None

Panel Member #3 I have no specific concerns. The measure specification is exceptionally clear.

Panel Member #4 No concerns

Panel Member #6 No concerns.

Panel Member #9 Specification in MIF is not clear, however measure is already implemented in ASC and hospital quality reporting programs so this may be less of an issue.

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure ☒ Yes ☐ No**

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

☐ Yes ☒☐ No

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

Submission Document: Testing Attachment, Section 2a2.2

Panel Member #1 median (range) signal to noise ratio for HOPDs and ASCs stratified by volume

Panel Member #2 The methods used were appropriate –the measure developer used the method suggested by Adams to assess reliability.

Panel Member #3 IUR is estimated.

Panel Member #4 Signal-to-noise reliability score was appropriate method

Panel Member #6 Reliability tested with signal to -noise method. No issues.

Panel Member #7 Adams formula of signal to noise

Panel Member #8 Signal to noise approach – appropriate.

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

Submission document: Testing attachment, section 2a2.3

Panel #1 generally good. The 30 procedure cutoff might be too low for the HOPDs given the lower bound of the IQR (0.59). It would be useful to see the values for the bottom quartile.

Panel Member #2 The reliability values obtained for HOPDs and ASCs are in the .7-.8 range. These are acceptable.

Panel Member #4 Results demonstrated sufficient reliability

Panel Member #6 Results show substantial reliability.

Submission document: Testing attachment, section 2a2.3

Panel Member #7 Reliability 0.744 for HOPD and 0.864 for ASC

The estimated IUR is equal to 0.744 in hospital outpatient departments and 0.864 in ambulatory surgical centers.

Panel Member #8 Results for both HOPD and ASCs were acceptable, particularly for facilities with >30 procedures.

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

Submission document: Testing attachment, section 2a2.2

☒☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

Panel Member #1 Good methodology and results. Some concerns about the lower quartile of reliability (distribution not provided) and (relatedly) if the 30 procedure cutoff if high enough. Also, I'm not sure the Landis and Koch (1977) citation for classifying observer agreement is relevant to this context. In general, I would call 0.70 adaquate and certainly <.50 inadaquate. See Adam's 2009.

Panel Member #2 The measure developer claims "high" reliability based on the Landis and Koch labels developed 40 years ago in the context of the kappa statistic for inter-rater agreement. More recent work by Adams, Zaslavsky, and others suggests that, in the context of provider profiling and financial incentive programs, reliabilities in the .7 range may be the minimum necessary to avoid significant misclassification problems. The reliability rates reported here are acceptable, but not so high as to allow great confidence in the measure's ability to avoid misclassification. Depending on the use context, misclassification may still be a problem with reliabilities at .9 or higher.

Panel Member #3 I have no specific concerns.

Panel Member #4 No concerns.

Panel Member #5

a. Signal to noise

b. 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].

c. 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].

Panel Member #7 Metric definitions well defined and reliability appropriately tested with reasonably high signal to noise ratio

Panel Member #8 Results were acceptable for HOPD with >=30 procedures (r=.78). ASCs had stronger reliability results (r = >0.85). Developer should note that Landis & Koch interpretation is not applicable to STN results.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Panel Member #1 None

Panel Member #3 The exclusions are not only clinically reasonable, but also quite small in sample size.

Panel Member #4 No concerns.

Panel Member #6 No concerns.

Submission document: Testing attachment, section 2b2.

Panel Member #7 Actually metric sponsors appear to have taken considerable care to exclude situations which might bias results by indication

Panel Member #8 None

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

Submission document: Testing attachment, section 2b4.

Panel Member #2 As noted above, the measure can only reliably identify extreme high or low outliers. It cannot identify meaningful differences in performance within the large main body of the distribution of scores.

Panel Member #3 I have no specific concerns.

Panel Member #4 No concerns.

Panel Member #6 No concerns.

Panel Member #7 Not many sites were identified as statistically significant outliers

Panel Member #8 The IQR for both HODPs and ASCs is very small (1.3 hospital visits per 1,000 and 1.0 hospital visits per 1,000, respectively); performance distribution is very constrained making it difficult to identify meaningful differences in performance.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

Panel Member #1 NA

Panel Member #2 N/A

Panel Member #3 This item is not applicable.

Panel Member #4 No concerns.

Panel Member #6 No concerns.

Submission document: Testing attachment, section 2b5.

Panel Member #7 Not applicable

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

Submission document: Testing attachment, section 2b6.

Panel Member #1 NA

Panel Member #2 No significant concerns

Panel Member #3 This item is not applicable.

Panel Member #6 Developer did not do an analysis because the measure is based on 100% sample of paid, final action claims submitted by facilities for payment.

Panel Member #7 Sponsors claim 10% availability of data with no analysis performed regarding missing data—this seems unlikely, although given the nature of claims data, aside from the potential problem of underreporting of comordities, there are likely very few missing data

Panel Member #8 None.

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒☐ Not applicable

16c. **Social risk adjustment:**

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

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

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

Panel Member #7 Sponsors demonstrate a very high correlation of the model prediction with or without social risk factors; however, they fail to present an analysis of exactly how classification might change for specific sites—high correlation of prediction and equivalent c-statistic for the model as a whole does not mean that there would not be some reclassification. Net reclassification index would be preferable.

Panel Member #7 Not addressed by the sponsors were the absence of sex and race in the model, each of which might theoretically—albeit for potentially unknown reasons—have an association with the outcome even after adjusting for other variables including social risk factors

16d.**Risk adjustment summary:**

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

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

Panel Member #7 Because the model is designed to predict an event subsequent to the procedure, rather than the risk of the procedure per se, inclusion of procedure-related factors such as polypectomy or associated endoscopy are not unreasonable

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

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☒☐ Yes ☐☒☐

Panel Member #7 No c-statistics of 0.687 and 0.654 (for hospital outpatient and ambulatory surgical centers respectively) are not great (<0.7); calibration curves do not demonstrate confidence intervals and are therefore less than ideal. One wonders what might have been the c-statistic with the inclusion of SES, sex and race

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

Panel Member #7 See my response to 16.c.3 above

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

Panel Member #1 Note that the section on the conceptual link between social risk factors and outcomes, as well as the methods to assess whether to include these factors in the risk adjustment model, might be something of an example for others to follow. My only concern is that they only considered 2 factors (dual eligible and AHRQ SES). It would have been very interesting to consider race/ethnicity in these analyses.

Panel Member #2 The developer presents a detailed and strong case for the influence of dual-elgibile status and area SES on post-colonoscopy hospitalization rates, and in fact finds a significant effect for both variables in both univariate and multivariate analyses. Still, the developer (CMS, actually, in this case) has chosen to leave both dual status and area SES out of the adjustment model, arguing that including them may "mask disparities In care". The developer points out that the results with and without the two variables are very highly correlated (.99), so that inclusion or exclusion of the two variables would make little diference in practice. The analysis of social risk factors is very thoughtfully and carefully done. I would have preferred that they include the two social risk factors based on the fully body of evidence presented. The very high correlation between the scores with and without the factors can be used on either side of the decision to include– if it makes little difference, there is no harm in including the two factors, and there is probably not much harm created by excluding the two factors.

Panel Member #3 The risk adjustment strategy is logical.

Panel Member #4 Good analysis and submittors noted the weak correlation using the AHRQ SES Index. Submittors also noted that CMS will not include the risk adjustment for reporting programs.

Panel Member #5 Ideally, continue to aim to select variables for the model based on theory first and not just statistical significance.

Panel Member #6 Extensive analysis

Panel Member #7 Rationale for exclusion of the SES highly debateable**;** rationale for exclusion of sex and race non-existant

Panel Member #8 Some concerns about lack of social risk adjustment. Overall, the C-statistic is .68 for HOPD, 0.65 for ASC. The value for both is concerning, especially for ASC – particularly without a confidence interval.

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

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

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

☒ Face validity

☒☐ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

Panel Member #1 The method for establishing face validity was OK. The developers claim that none of the existing measures are a fair comparator for validity testing. I think they should have attempted some analyses on the "Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at ASCs (ASC General Surgery)" for facilities that have adequate volumes of target procedures. Also, they could have done some validity testing on the outcome – what proportion of the numerator hospitalizations are related to the colonoscopy?

Panel Member #3 The measure was reviewed by a 14-member technical expert panel.

Panel Member #6 Extensive review and analysis of similar measures. Used a TEP to establish validity.

Panel Member #7 Survey of technical expert panel. Although good-faith effort appears to have been made to find other measures with which to perform empiric testing. They also note that the metric has already withstood test of time and appears to have been accepted. Interestingly, since initiating the metric, there does appear to have been a meaningful decrease in HOPD incidence of post-colonscopy hospital admissions within 7 days.

Panel Member #8 Approach to face validity was ok, however, the question is whether this was the only way the developer could assess face validity.

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1 Face validity testing yielded support for the measure. Unforunately, the reasons for the dissenting view were not reported. I think they should have done some empirical testing on the measure.

Panel Member #2 The face validity results are acceptable; the empirical validity testing results are generally quite weak, but in the predicted directions for the most part.

Panel Member #3 Twelve (12) of fourteen (14), or 86%, of panel members somewhat, moderately, or strongly agreed that the measure is valid.

Panel Member #4 Robust face validity with external reviews, TEPs and public comments

Panel Member #6 Appropriate method used and results.

Panel Member #7 14 of the 17 TEP members completed the survey—however they give the percentages among the respondants—if they gave the percentages for the total, validation would have been somewhat less compelling:

Moderately disagree 1/14 vs 1/17

Somewhat disagree 1/14 vs 1/17

Somewhat agree 2/14 vs 2/17

Moderately agree 8/14 vs 8/17

Strongly agree 2/14 vs 2/17

They conclude that 12/14 (86%) agree. Somewhat agree is hardly a strong support for face validity. This would leave 10/17 who agree, which, although less compelling, is still 59%

Panel Member #8 10 of 14 voting members of the TEP indicated at least moderate agreement that the measure was valid. This is reasonable evidence of face validity, though not particularly strong.

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

Submission document: Testing attachment, section 2b1.

☒☐ **Yes**

☐☒ **No**

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

Panel Member #6 Used face validity method.

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

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

☒☐ **Yes**

☒☐ **No**

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

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

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

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

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

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

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

27. **Panel Member #1** I think score and item-level validity testing should have been attempted.

Panel Member #2 As in the case of reliability, only score -level validity testing was done, so the moderate rating for score-level validity is the same as the moderate rating for overall validity. The moderate rating comes mainly from the face validity results; the empirical validity results are essentially non-existent. The measure developer argues that no complaints about measure validity have been made since the time that it has been in use in dry run reports. This isn't strong empirical support for validity.

Panel Member #3 Validity testing by way of expert approval is not bona fide validity testing, considering the complexity of claims data. No empirical validity testing was performed.

Panel Member #4 No concerns

Panel Member #5 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.

Rationale provided for not performing validity testing is that no other similar measures exist for comparison. There are other approaches to validity aside from comparison with existing measures.

Panel Member #7 Absence of race and sex in the final model as well as absence of net reclassification index with relatively moderate face validity

Panel Member #8 Several threats to validity include lack of social risk factors despite documented relationship with outcome, poor C-statistic indicating potential problems with the existing risk adjustment model, and constrained performance variation. Face validity results indicated 71% of TEP members indicated at least moderate agreement that the is valid.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

Panel Member #2 The measure has been shown to be reliable, but there is not much variation in scores through most of the performance distribution. There is very little difference between 10^{th} and 90^{th} percentiles, for example. The measure seems able to identify extreme high or low outliers, but not to distinguish performance between facilities in the broad middle of the distribution. NQF endorsement should reflect that limitation. The measure should not be used for other purposes based on an "NQF-endorsed" status.

Panel Member #8 Rationale for not using social risk factors, contention by developer that there is no way to empirically evaluate validity of the measure – i.e., no relevant benchmark for this measure (thereby forcing reliance upon face validity).

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

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

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF2539_colonoscopy_evidence_attachment_FINAL_040920.docx
1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.
Yes

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2539
Measure Title: Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: Click here to enter a date

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.

- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Outcome: All-cause, unplanned hospital visits within 7 days. We define a hospital visit as any emergency department (ED) visit, observation stay, or unplanned inpatient admission *Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)*

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

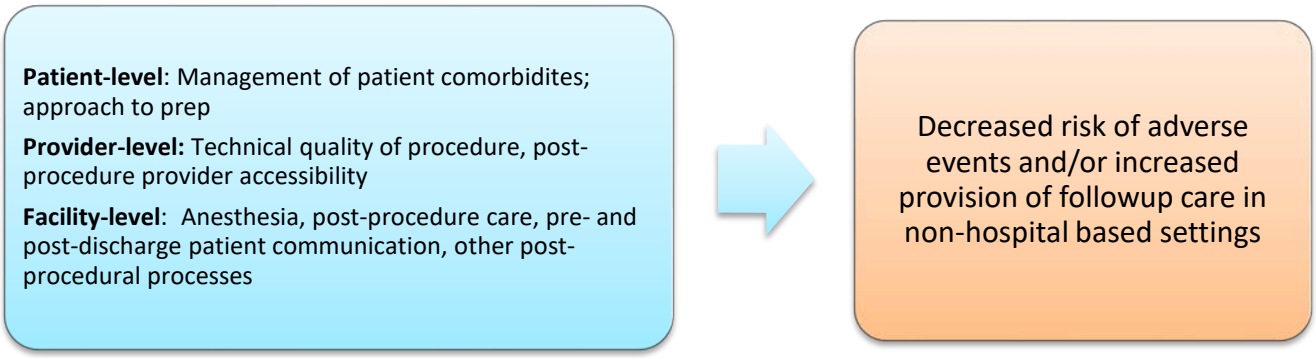
☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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

The conceptual model for colonoscopy quality, shown below, shows the pathway by which facilities can modify the outcome. For example, the model identifies that patient-level factors, such as comorbidities, increase the risk of unplanned hospitals visits [1]. Better management of the risk associated with these comorbidities may be a potential avenue for facilities to reduce unplanned hospital visits. Provider-level factors (technical quality of the procedure, post-procedure provider accessibility), and facility-level factors (such as the anesthesia, pre- and post-discharge patient communication, other post-procedural processes) may also contribute to the risk of unplanned hospital visits. Therefore, facilities may have opportunities to lower their unplanned hospital visit rates through quality-improvement efforts focused on patient, provider, and facility factors [1].



Citation:

1. Ranasinghe I, Parzynski CS, Searfoss R, Montague J, Lin Z, Allen J, Vender R, Bhat K, Ross JS, Bernheim S, Krumholz HM, Drye EE. Differences in Colonoscopy Quality Among Facilities: Development of a Post-Colonoscopy Risk-Standardized Rate of Unplanned Hospital Visits. Gastroenterology. Jan 2016;150(1):103-13.

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

Not applicable.

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

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

Patients may experience a range of potential adverse events after an outpatient colonoscopy, which could lead to unplanned hospital visits, including ED visits, observation stays, and unplanned inpatient admissions. This measure provides the opportunity to improve quality of care and to lower rates of adverse events leading to hospital visits after an outpatient colonoscopy.

Complications

Gastrointestinal complications from colonoscopy are common and range from severe to mild. Colonic perforation and gastrointestinal (GI) bleeding are relatively rare but severe adverse events reported after colonoscopy. A meta-analysis of 20 published studies of complications among patients aged ≥65 years in all care settings suggested these occur at a rate of 0.10% (95% confidence interval [CI] 0.09-1.50%) for colonic perforation and 0.63% (95% CI 0.57-0.70%) for GI bleeding [1]. Other GI complications after colonoscopy are considerably more common. Among surveyed patients, the reported frequency of complications ranges from, 20-34% [2,3]. These complications include abdominal pain, abdominal distension, nausea, vomiting, and other nonspecific symptoms.

Cardiovascular and pulmonary complications are the most frequent non-GI complications reported after colonoscopy. Pulmonary complications generally occur as a complication of the sedation given at the time of the procedure [4,5]. Excessive sedation may lead to hypoxia, hypotension, respiratory arrest, and aspiration pneumonia [4,5]. Cardiovascular complications may be attributed to many factors, including the effects of the anesthesia. The rates of cardiovascular and pulmonary complications reported in individual studies included in our review of the literature ranged from 0.012-1.94%. This range may reflect variation in definition of these events and differences in data sources used to capture these complications [6-9].
Post-procedural infections also occur following colonoscopy. For example, a 2018 study found rates of infection within 7 days of a screening colonoscopy performed by at an ASC to be 1.1 per 1000 colonoscopies [10]. Furthermore, the study authors found that the rates of infection varied widely by ASC, from 0 to 115 per 1000 colonoscopies.

Hospital visits following colonoscopy

The symptoms described above can result in the need for acute care. Overall, reported rates of post-procedure hospital use, as measured by inpatient admissions or a combination of admissions and ED visits, range from 1-2.4% within 30 days [3,6]. A more recent retrospectively review of 50,319 colonoscopies performed on 44,082 individuals (47% male, median age 59 years) reported an ED visit rate within 7 days of a colonoscopy of 0.76% [13], and a claims-based analysis found an average 7-day hospital visit rate (defined as an ED visit, observation stay, or inpatient hospitalization) of 1.63% [11]. The rate of hospitalization varies by type of complication; hospitalization rates were nearly 100% among patients who developed perforation and between 50.8% and 70.7% among patients who developed lower GI bleeding [12]. In contrast, hospitalizations among patients with an abdominal pain or nausea diagnosis were less common [12].

Studies have shown that many of the reasons for post-procedural hospital visits are related to the colonoscopy. For example, a 2018 single-center study examined the medical records (including medication information) of patients who experienced an emergency department (ED) visit within 7 days of an outpatient colonoscopy [13]. The study authors extracted patients' chief complaint from medical records, assigned the chief complaints as related or unrelated to the colonoscopy, and found that 68% of the reasons for the ED visit were due to the colonoscopy. The most common reasons for related ED visits were abdominal pain (38.2%), gastrointestinal bleeding (29.7%), cardiopulmonary disorders (12.7%), and nausea/vomiting (4.2%). In another study, the authors examined the most frequent diagnoses in claims data associated with an unplanned hospital visit within 7 days, which included hemorrhage (6.4% of all unplanned visits), accidental operative laceration (3.0%), abdominal pain (3.0%), GI hemorrhage (2.7%), chest pain (1.9%), and urinary tract infection (1.8%) [11]. (Please note the measure developer plans to update this analysis in ICD-10 data and have the results on hand for review by the Standing Committee in June.)

Pathways for improvement

Provider- and facility-level factors can affect the outcome of complications and hospital visits related to a colonoscopy. For example, provider-level factors such as low provider volume and fellow involvement in the procedure were significantly associated with a higher risk of an ED visit in one study [13], and low procedure volume was associated with a higher risk of infection in another study [10], suggesting facilities can influence the patients' outcome through these modifiable pathways. In addition, the choice of sedation may influence complication rates. For example, in a 2018 retrospective claims-based analysis of more than 3 million outpatient colonoscopies, researchers found that the use of anesthesia assistance (sedation with agents that result in deeper sedation, such as Propofol, rather than conscious sedation), resulted in increased risk of aspiration pneumonia (OR, 1.63; 95% CI, 1.11–2.37) [14].

Providers are often unaware of complications for which patients visit the hospital, leading to understated complication rates and suggesting the need for better measurement to drive quality improvement [15]. Both patients and providers can benefit from outcome measures that capture the full range of adverse experiences associated with outpatient colonoscopy and illuminate quality differences.

Public reporting

The Hospital Outpatient Quality Reporting Program (HOQR) provides CMS with data to help Medicare beneficiaries make more informed decisions about their healthcare. As of December 2017, this measure has been publicly available on Hospital Compare, and since July 2015, results have been available in the form of facility-specific quality reports that conduct outpatient colonoscopies. Thus, it is important to continue to make this information transparent to patients choosing among providers who offer this elective procedure.

Importantly, providing outcome rates to providers will make meaningful quality differences visible to clinicians, thus incentivizing improvement. The national rate of hospital visits per 1,000 colonoscopies among HOPDs declined from 16.4 in 2018 reporting (2017 data) to 14.8 in 2019 reporting (2018 data), and the distribution of risk-standardized rates also declined (the interquartile range of rates for 2019 is completely below the 2018 interquartile range). This decline may reflect 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. (Note that the 2020 national rate was 16.4, however this difference compared to 2019 can be attributed to a change in the measures' specifications that result in the use of three years of performance data that overlap with 2018 and 2019 performance periods.)

Citations

1. Day LW, Kwon A, Inadomi JM, Walter LC, Somsouk M. Adverse events in older patients undergoing colonoscopy: a systematic review and meta-analysis. *Gastrointest Endosc.* Oct 2011;74(4):885-896.

2. Baudet JS, Diaz-Bethencourt D, Aviles J, Aguirre-Jaime A. Minor adverse events of colonoscopy on ambulatory patients: the impact of moderate sedation. *Eur J Gastroenterol Hepatol.* Jun 2009;21(6):656-661.

3. Ko CW, Riffle S, Shapiro JA, et al. Incidence of minor complications and time lost from normal activities after screening or surveillance colonoscopy. *Gastrointest Endosc.* Apr 2007;65(4):648-656.

4. Ko CW, Dominitz JA. Complications of colonoscopy: magnitude and management. Gastrointestinal endoscopy clinics of North America. Oct 2010;20(4):659-671.

5. Committee ASoP, Fisher DA, Maple JT, et al. Complications of colonoscopy. *Gastrointest Endosc.* Oct 2011;74(4):745-752.

6. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med. Jun 16 2009;150(12):849-857, W152.

7. Crispin A, Birkner B, Munte A, Nusko G, Mansmann U. Process quality and incidence of acute complications in a series of more than 230,000 outpatient colonoscopies. *Endoscopy.* Dec 2009;41(12):1018-1025.

8. Radaelli F, Meucci G, Minoli G. Italian Association of Hospital G. Colonoscopy practice in Italy: a prospective survey on behalf of the Italian Association of Hospital Gastroenterologists. Nov 2008;40(11):897-904.

9. Singh H, Penfold RB, DeCoster C, et al. Colonoscopy and its complications across a Canadian regional health authority. *Dig Liver Dis.* Mar 2009;69(3):665-671.

10. Wang P, Xu T, Ngamruengphong S, Makary MA, Kalloo A, Hutfless S. Rates of infection after colonoscopy and osophagogastroduodenoscopy in ambulatory surgery centres in the USA. Gut. 2018;67(9):1626–1636.

11. Ranasinghe I, Parzynski CS, Searfoss R, Montague J, Lin Z, Allen J, Vender R, Bhat K, Ross JS, Bernheim S, Krumholz HM, Drye EE. Differences in colonoscopy quality among facilities: Development of a post-colonoscopy risk-standardized rate of unplanned hospital visits. *Gastroenterology*. Jan 2016;150(1):103-13.

12. Wang L, Mannalithara A, Singh G, Ladabaum U. Low rates of gastrointestinal and non-gastrointestinal complications for screening or surveillance colonoscopies in a population-based study. *Gastroenterology*. Oct 2018;154:540-555.

13. Grossberg LB, Vodonos A, Papamichael K, Novack V, Sawhney M, Leffler DA. Predictors of post-colonoscopy emergency department use. *Gastrointest Endosc*. Feb 2018;87(2):517-525.

14. Bielawska B, Hookey LC, Sutradhar R, et al. Anesthesia assistance in outpatient colonoscopy and risk of aspiration pneumonia, bowel perforation, and splenic injury. Gastroenterology. 2018;154(1):77–85.e3.

15. Leffler DA, Kheraj R, Garud S, Neeman N, Nathanson LA, Kelly CP, Sawhney M, Landon B, Doyle R, Rosenberg S, Aronson M. The incidence and cost of unexpected hospital use after scheduled outpatient endoscopy. *Arch Intern Med.* 2010;170:1752-1757.

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

Not applicable. This is an outcome measure.

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Not applicable. This is an outcome measure.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Not applicable. This is an outcome measure.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Not applicable. This is an outcome measure.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Not applicable. This is an outcome measure.
Provide all other grades and definitions from the evidence grading system	Not applicable. This is an outcome measure.

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

1a.4 OTHER SOURCE OF EVIDENCE

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

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

Not applicable. This is an outcome measure.

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

Not applicable. This is an outcome measure.

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

Not applicable. This is an outcome measure.

1b. Performance Gap

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

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

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

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

The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about facility-level 7-day, risk-standardized hospital visit rates following outpatient colonoscopy.

Colonoscopy is a common and costly procedure performed at outpatient facilities and is frequently performed among relatively healthy patients to screen for colorectal cancer (CRC). Between January 1, 2016 and December 31, 2018, there were 2,258,661 colonoscopies performed in non-federal acute care hospital outpatient departments (HOPDs) and 2,524,898 performed in ambulatory surgical centers (ASCs). Given the widespread use of colonoscopy, understanding and minimizing procedure-related adverse events is a high priority. These adverse events, such as abdominal pain, bleeding, and intestinal perforation, can result in unanticipated hospital visits post procedure, and as outlined in the evidence attachment, a majority (68% in one study) of the reasons for emergency department visits following outpatient colonoscopy are due to the colonoscopy. Furthermore, physicians performing colonoscopies are often unaware that patients seek acute care at hospitals following the procedure and thus underestimate such events. This risk-standardized quality measure addresses this information gap and promotes quality improvement by providing feedback to facilities and physicians, as well as transparency for patients on the rates of and variation across facilities in unplanned hospital visits after colonoscopy.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Below we describe the distribution of performance on the colonoscopy measure during public reporting, which reflects the measure as specified in this submission. The data below are for a three-year performance period. In the CY2019 final rule, CMS finalized the measure to include three years of performance data (83 FR 58818).

Hospital Outpatient Departments (HOPDs)

Distribution of measure scores for the colonoscopy measure for 2020 public reporting.

Data Source: Medicare FFS claims (Part A and B), January 1, 2016-December 31, 2018

Note: Sample includes all hospital outpatient departments with results

Distribution (percentiles) of the risk-standardized hospital visit rates (RSHVRs) per 1000 colonoscopies, all facilities (n = 4034)

Percentile//7-Day RSHVR

Min//11.67

P10//14.92

P25//15.76

P50//16.38

P75//17.10

P90//18.10

Max//24.27

Mean (SD)//16.47 (1.32)

Distribution (deciles) of the RSHVRs per 1000 colonoscopies, all facilities (HOPDs):

Decile//# facilities//Minimum RSHVR//Maximum RSHVR

1//403//11.67//14.92
2//403//14.92//15.57
3//404//15.57//15.95
4//403//15.95//16.21
5//404//16.21//16.38
6//403//16.38//16.60
7//404//16.60//16.90
8//403//16.90//17.35
9//404//17.35//18.10
10//403//18.10//24.27

Ambulatory Surgery Centers (ASCs)
Distribution of measure scores for the colonoscopy measure for 2020 public comment reporting
Data Source: Medicare FFS claims (Part A and B), January 1, 2016-December 31, 2018

Distribution (percentiles) of the risk-standardized hospital visit rates per 1000 colonoscopies, all facilities (n = 2,261)

Risk-standardized hospital visit rates per 1000 colonoscopies:

Percentile//7-Day RSHVR
Min//8.59
P10//11.07
P25//11.75
P50//12.23
P75//12.82
P90//13.57
Max//17.94
Mean (SD)//12.29 (1.03)

Distribution (deciles) of the RSHVRs per 1000 colonoscopies, all facilities:
Decile//# facilities//Minimum RSHVR//Maximum RSHVR

1//226//8.59//11.07
2//226//11.07//11.58
3//226//11.59//11.88
4//226//11.88//12.08
5//226//12.08//12.23
6//227//12.23//12.41
7//226//12.41//12.65
8//226//12.65//13.02

9//226//13.02//13.57
10//226//13.58//17.94

Change in performance over time:

Hospital Outpatient Departments

The distribution of hospital visit rates among HOPDs declined for 2019 reporting compared to 2018 reporting. This decline may reflect 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. In 2020, CMS started to use three years of data with data dates that overlap with 2018 and 2019 public reporting.

The national rate of hospital visits per 1,000 colonoscopies among HOPDs by year of public reporting was:

2018 public reporting, 2016 data (January 1, 2016-December 31, 2016): 16.4

2019 public reporting, 2017 data (January 1, 2017-December 31, 2017): 14.8

2020 public reporting, 2016-2020 data (January 1, 2016-December 31, 2018): 16.4

Ambulatory Surgery Centers:

The rate of hospital visits following colonoscopies among ASCs reporting declined slightly from 2018 to 2019. The national rate of hospital visits per 1,000 colonoscopies among ASCs by year of public reporting was:

2018 public reporting, 2016 data (January 1, 2016-December 31, 2016): 12.5

2019 public reporting, 2017 data (January 1, 2017-December 31, 2017): 12.3

2020 public reporting, 2016-2020 data (January 1, 2016-December 31, 2018): 12.2

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

Not applicable; we provide performance data above.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We provide an extensive analysis of disparities in the testing form, section 2b4.4b.

For all social risk factor analyses we used Medicare FFS claims from January 1, 2016-December 31, 2018.

Dual eligible variable:

Distribution of the measure score [hospital visit rates (RSHVRs)] between the first and fourth quartiles, by proportion of dual-eligible patients (for facilities with at least 30 patients):

HOPDs

Characteristic //Duals, 1st quartile (<=2.94%)//Non-Duals, 4th quartile (>9.89%)
Number of HOPDs//894//895
Number of patients//768,473//336,342
Maximum RSHVR*//21.52//24.27
90th //17.83//18.20
75th //16.98//17.29
Median//16.17//16.53
25th//15.42//15.89
10th//14.39//15.30

*RSHVRs are per 1,000 colonoscopies.

ASCs
Characteristic //Duals, 1st quartile (<=1.09%)//Non-Duals, 4th quartile (>5.35%)
Number of ASCs//518//519
Number of patients//70,7563//393,510
Maximum RSHVR*//16.02//17.15
90th//13.26//13.64
75th //12.68//12.86
Median//12.08//12.26
25th//11.58//11.76
10th//10.99//11.16
Minimum RSHVR//9.05//8.59

AHRQ SES variable:
Distribution of the measure score [hospital visit rates (RSHVRs)] between the first and fourth quartiles, for the proportion of patients with the low AHRQ SES variable (for facilities with at least 30 patients):

HOPDs
Characteristic//Low AHRQ SES, 1st quartile (<=5.38%)//Low AHRQ SES, 4th quartile (>26.47%)
Number of HOPDs//896//894
Number of patients//659,707//307,490
Maximum RSHVR*//20.86//24.27
90th//17.81//18.31
75th//16.93//17.35
Median//16.19//16.56
25th//15.50//15.95
10th//14.40//15.47
Minimum RSHVR//11.87//12.91

ASCs

Characteristic//Low AHRQ SES, 1st quartile (<=3.96%)//Low AHRQ SES, 4th quartile (>16.84%)

Number of ASCs//519//518

Number of patients//665512//488590

Maximum RSHVR//16.2//17.15

90th//13.33//13.76

75th//12.59//3.04

Median//12.03//12.34

25th//11.45//11.79

10th//10.79//11.09

Minimum RSHVR//8.94//8.59

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

Not applicable. Data on disparities are presented above

2. Reliability and Validity – Scientific Acceptability

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

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

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

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

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

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

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

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

Attachment **Attachment:** Colonoscopy_Measure_Data_Dictionary_v2019a.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
Not an instrument-based measure

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
Yes

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
Changes made in 2017

Changes made in 2017 are described in more detail in the following report:
https://www.qualitynet.org/files/5d0d38ed764be766b0102e71?filename=2017_Colonoscopy_AUS_Report.pdf

1. Applies to HOPDs only. Expansion of same outpatient claim ED visit exclusion to include colonoscopies matched to inpatient claims with ED visits, except for those with a primary diagnosis on the facility claim that is a complication of care as defined by four AHRQ CCS categories.
Rationale: The measure previously excluded colonoscopies that are billed on the same hospital outpatient claim as an ED visit and those 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, since it is not possible to determine the order of events in these situations. The 3-day rule cases are similar to those that are excluded using the two existing ED-related exclusions, except that the Part B colonoscopy claim was matched to an inpatient claim instead of an outpatient claim. During the dry run and throughout the 2016 CDR release cycle, facilities noted rare instances in which the measure was counting ED visit outcomes that occurred before the colonoscopy procedure. A review of the top facility diagnosis codes for these cases indicated that a portion of them are clear complications of care, while the rest are ambiguous in terms of indicating whether the colonoscopy happened before or after the ED visit. This targeted exclusion ensures that the measure will continue to include cases with ED visits for clear complications of care, but also minimizes the number of cases we include that began with an ED visit.

2. Applies to HOPDs only. Revised current ED-related exclusions to be consistent with the new exclusion described above, and to align with ED-related exclusions in Hospital Visits after Hospital Outpatient Surgery (OP-36), to only exclude colonoscopies on the same claim or on the same day and at the same facility as an ED visit, if the facility claim does not have a diagnosis that indicates a complication.
Rationale: While we cannot determine the order of events in these cases, we are keeping cases with facility diagnoses that indicate a complication of care, in order to ensure that the measure captures its intended outcome. This change aligned all colonoscopy ED-related exclusions for consistency within the measure and with Hospital Visits after Hospital Outpatient Surgery (OP-36).

3. Modification of the planned admission algorithm to align with appropriate changes signaled during ICD-10 code testing and review.
Rationale: First, the algorithm was aligned with version 4.0 (ICD-10) of CMS's Planned Readmission Algorithm (PRA) used in the hospital inpatient readmission measures and the 2017 ACO admission measures. Next, additional ICD-10-PCS and ICD-10-CM codes were removed or added, as appropriate to the colonoscopy measure, following review of new FY2017 codes and general equivalence mappings.

Changes made in 2018

Changes made in 2018 are described in more detail in the following report:
https://www.qualitynet.org/files/5d0d3704764be766b0100ec0?filename=2018_Colonoscopy_AnlUpdtRpt.pdf

1. Modification of the planned admission algorithm (PAA) to align with changes made to CMS's PRA version 4.0_2019.
Rationale: These changes improve the accuracy of the algorithm.

2. Applies to HOPDs only. Modification of the list of AHRQ CCS categories used to define complications of care for ED visit exclusions.
Rationale: The list of AHRQ CCS categories used to identify complications of care in the same claim/same day ED visit exclusions was modified and expanded to include an ICD-10 diagnosis code. The changes were made to improve the accuracy of the measure and ensure that it captures complications of care following low-risk colonoscopies.

In the CY2019 Final Rule (83 FR 58818), CMS extended the performance period for the colonoscopy measure to 3 years.

Changes made in 2019:
We provide this information in greater detail in than for the 2018 and 2017 changes because the updated report is not yet available on QualityNet. The updated report should be available to the public in January 2020 at the following URL: https://www.qualitynet.org/outpatient/measures/colonoscopy/methodology.

1.Modification of the PAA to align with changes made to CMS's Planned Readmission Algorithm version 4.0 2020.
For this update, we studied the 2019 versions of the AHRQ CCS for diagnoses and procedures, respectively, to determine how the newly implemented ICD-10 codes in the 2018 code set were categorized, and to examine any code shifts that may have occurred from the previous version of the AHRQ CCS to the most recent AHRQ CCS. Review of these versions of the AHRQ CCS was extensive, and included:
•Examination of seven AHRQ CCS diagnosis categories and 13 AHRQ CCS procedure categories to determine how the newly implemented ICD-10 codes should be incorporated into the Planned Readmission Algorithm specifications; and,
•Examination of one AHRQ CCS diagnosis category and eight AHRQ CCS procedure categories that shifted to investigate where code shifts may affect the specialty cohort definitions and Planned Readmission Algorithm.

We then solicited input from clinical and measure experts to confirm the clinical appropriateness of the AHRQ CCS categorization of the newly implemented ICD-10 codes and any changes warranted due to the code shifts that occurred. The experts also reviewed the newly implemented ICD-10 codes in the FY 2019 version of the ICD-10-CM/PCS to determine which, if any, should be added to the singular ICD-10 code lists that are also used in the algorithm (conditions that are not captured by AHRQ CCS categories). The intent was to maintain the clinical integrity of the algorithm.
Changes for potentially planned procedures included:
•The addition of four AHRQ CCS procedure categories (Procedure CCS 96, 118, 162, 163), which consisted of procedures that clinicians deemed potentially planned. Examples of these categories are "Other OR lower GI therapeutic procedures" (CCS 96) and "Other OR therapeutic procedures on joints" (CCS 162). We previously included subsets of ICD-10-PCS codes within CCS 96, 118, and 163 on the potentially planned procedures list.
•The addition of selected ICD-10-PCS codes within CCS group 112 ("Other OR therapeutic procedures of urinary tract").
•The removal of CCS 95 ("Other non-OR lower GI therapeutic procedures") and 174 ("Other non-OR therapeutic procedures on skin subcutaneous tissue fascia and breast") as a whole; we previously included a subset of codes on the potentially planned procedures list.
•An additional 14 CCS categories were previously specified for Colonoscopy, including CCS 70, 72, 73, 75, 76, 77, 90, 92, 93, 95, 96, 97, 98, and 194 in the 2018 reporting cycle (v4.0_2019 PAA). These codes were carried into the current v4.0_2020 PAA.
Changes in acute diagnoses included:
•An additional five ICD-10-CM codes were specified for colonoscopy within CCS106 ("Dysrhythmia") and CCS 155 ("Other gastrointestinal disorders") in the 2018 reporting cycle (v4.0_2019 PAA). These codes were carried into the current v4.0_2020 PAA.

The complete set of codes reflected in the v4.0_2020 Planned Readmission Algorithm adopted as the PAA for the colonoscopy measure are available in the data dictionary tables: tabs "Colonos PAA PA1 Always Plnnd Px", "Colonos PAA PA2 Always Plnnd Dx", "Colonos PAA PA3 Pot Plnnd Px" and "Colonos PAA PA4 Acute Dx".

Rationale: These changes align with the specifications of similar measures and improve the accuracy of the algorithm

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

With this update, we further refine the same-claim ED exclusion. Prior to this update, surgeries billed on the same claim as an ED visit were excluded from the measure, unless the claim had a diagnosis indicating a complication of care occurred. This update further refines this exclusion to exclude surgeries that occur on the same day and on the same claim as the surgery, unless there is a diagnosis of complication of care indicated on the claim. Additionally, we expand the exclusion criteria to exclude surgeries that are billed on the same hospital outpatient claim, but occur after the ED visit, regardless of whether complications of care are billed or not. Note that this update was applied prior to the release of 2020 reporting to be responsive to stakeholder feedback.

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

3. Update to exclusion for surgeries that are billed on the same hospital outpatient claim and that occur after the ED visit.
Rationale: In these situations, we assume that the surgery was subsequent to the ED visit and may not represent a routine surgery. Timing of the ED visits is determined using revenue center dates from the outpatient claim.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Unplanned hospital visits within 7 days of a qualifying colonoscopy.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Outcome Definition
The outcome for this measure is all-cause, unplanned hospital visits within 7 days of an outpatient colonoscopy. Hospital visits include ED visits, observation stays, and unplanned inpatient admissions. If more than one unplanned hospital visit occurs, only the first hospital visit within the outcome timeframe is counted in the outcome.

Identification of Planned Admissions
The measure outcome includes any inpatient admission within the first 7 days after the colonoscopy, unless that admission is deemed a "planned" admission as defined by the measure's PAA. The Centers for Medicare & Medicaid Services (CMS) seeks to count only unplanned admissions in the measure outcome, because variation in "planned" admissions does not reflect quality differences. We based the PAA on the CMS PRA Version 4.0_2019, which CMS created for its hospital-wide readmission measure. In brief, the algorithm identifies admissions that are typically planned and may occur after the patient's index event. The algorithm always considers a few specific, limited types of care planned (e.g., major organ transplant, rehabilitation, or maintenance chemotherapy). Otherwise, the algorithm defines a planned admission as a non-acute admission for a scheduled procedure (e.g., total hip replacement or cholecystectomy), and the algorithm never considers admissions for acute illness or for complications of care planned. For example, the algorithm considers hip replacement unplanned if hip fracture (an acute condition) is the discharge diagnosis, but planned if osteoarthritis (a non-acute condition) is the discharge diagnosis. The algorithm considers admissions that include potentially planned procedures with acute diagnoses or that might represent complications of a colonoscopy unplanned and thus counts these admissions in the measure outcome.

For more information about the PAA, please see the Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy Measure 2018 Measure Updates and Specifications Report posted on the web page provided in data field S.1. Also see sheets 'PAA PA1 always planned Px', 'PAA PA2 always planned Dx', 'PAA PA3 post planned Px', and 'PAA PA4 acute Dx' in the attached Data Dictionary for the most up-to-date sets of codes in the algorithm for

'always planned procedures' (PA1), 'always planned diagnoses' (PA2), 'potentially planned procedures' (PA3), and 'acute' diagnoses (PA4).

Definition of ED and Observation Stay
We defined ED visits and observation stays using one of the specified billing codes or revenue center codes identified in Medicare Part B Outpatient hospital claims. The codes that define ED visits and observation stays are in the attached Data Dictionary, sheet "Colons_Outcome_ED_Obs."

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Colonoscopies performed at hospital outpatient departments (HOPDs) and ambulatory surgical centers (ASCs) for Medicare FFS patients aged 65 years and older.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Target Population
The measure includes colonoscopies performed at HOPDs and ASCs. The measure calculates a facility-level score for all eligible facilities separately for HOPDs, and ASCs.

The target population is patients aged 65 years and older who have a colonoscopy, to screen for colorectal cancer, biopsy or remove pre-cancerous lesions, or evaluate non-emergent symptoms and signs of disease. We limited the measure cohort to patients who are 65 and older, enrolled in Medicare FFS, and have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of procedure since national data linking risk factors, procedures, and outcomes across care settings are only available for this group.

Eligible colonoscopies were identified using specified Current Procedural Terminology (CPT)/Healthcare Common Procedure Coding System (HCPCS) procedure codes in the Medicare Carrier (Part B Physician) Standard Analytical File (SAF). The CPT and HCPCS procedure codes that define the cohort are in the attached Data Dictionary, sheet "Colonos_Cohort."

We considered all colonoscopy codes during development of the measure cohort. We did not include in the measure colonoscopy CPT procedure codes that reflected fundamentally higher-risk or different procedures. Those procedures billed with a qualifying colonoscopy procedure code and a high-risk colonoscopy procedure code (see attached Data Dictionary, sheet "Colonos_Excll") were not included in the measure.

Colonoscopy is not possible among patients who have had a prior total colectomy. Any claim for a colonoscopy in a patient with a prior total colectomy is therefore likely to be a coding error. We perform an error check to ensure the measure does not include these patients with a total colectomy recorded in their prior medical history. The CPT and HCPCS procedure codes and International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and ICD-10-CM codes that define the total colectomy data reliability check are in the attached Data Dictionary, sheet "Colonos_Excl."

Capture of Colonoscopies Affected by the Medicare 3-Day Payment Window Policy:
Colonoscopies performed at HOPDs can be affected by the Medicare 3-day payment window policy. The policy states that outpatient services (including all diagnostic services such as colonoscopy) provided by a hospital or any Part B entity wholly owned or wholly operated by a hospital (such as an HOPD) in the three calendar days preceding the date of a beneficiary's inpatient admission are deemed to be related to the admission [1]. For outpatient colonoscopies affected, the facility claim (for the technical portion of the colonoscopy) is bundled with the inpatient claim, although the Medicare Part B physician claim for professional services rendered is still submitted. This policy has implications for the measure because it may lead to: (1) failure to completely capture outpatient colonoscopies performed at HOPDs; and (2) underreporting of outcomes for colonoscopies performed in the HOPD setting.

To ensure the capture of HOPD colonoscopies, we identify physician claims for colonoscopy in the HOPD setting from Medicare Part B claims, which had an inpatient admission within three days and lacked a corresponding HOPD facility

claim. We then attribute the colonoscopies identified as affected by this policy to the appropriate HOPD facility using the facility provider ID from the inpatient claim.

Citations

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

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

We established the following exclusion criteria after reviewing the literature, examining existing measures, discussing alternatives with the working group and technical expert panel (TEP) members, reviewing feedback from the national dry run held in July 2015, and public reporting in 2018 and 2019, and annual re-evaluation of the measure in 2017, 2018, and 2019. The goal was to be as inclusive as possible; we excluded only those high-risk procedures and patient groups for which risk adjustment would not be adequate or for which hospital visits were not typically a quality signal. The exclusions, based on clinical rationales, prevent unfair distortion of performance results.

1) Colonoscopies for patients who lack continuous enrollment in Medicare FFS Parts A and B in the 7 days after the procedure.
Rationale: We exclude these patients to ensure full data availability for outcome assessment.

2) Colonoscopies that occur concurrently with high-risk upper gastrointestinal (GI) endoscopy procedures.
Rationale: Patients undergoing concurrent high-risk upper GI endoscopy procedures, such as upper GI endoscopy procedures for the control of bleeding or treatment of esophageal varices, and have a higher risk profile than typical colonoscopy patients. Therefore, these patients have a disproportionally higher risk for the outcome.

3) Colonoscopies for patients with a history of inflammatory bowel disease (IBD) or diagnosis of IBD at time of index colonoscopy or on the subsequent hospital visit outcome claim.
Rationale: We exclude these patients because:

- IBD is a chronic condition; patients with IBD undergo colonoscopy both for surveillance due to increased cancer risk and for evaluation of acute symptoms. IBD is likely to be coded as the primary diagnosis prompting the procedure irrespective of whether the patients are undergoing a screening procedure or a diagnostic procedure in the setting of an acute exacerbation of IBD. Therefore, we may not be able to adequately risk adjust for these patients, as we cannot identify relatively well versus acutely unwell patients among visits coded as IBD.
- Our aim is to capture hospital visits which reflect the quality of care. Admissions for acutely ill IBD patients who are evaluated with an outpatient colonoscopy and are subsequently admitted for medical treatment of an IBD flare do not reflect the quality of the colonoscopy. In our 2010 Medicare 20% FFS Full Development Sample (see the 2014 Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy Measure Technical Report posted at https://www.qualitynet.org/files/5d0d37ae764be766b010196e?filename=ClnscpyMsr_TechReport.pdf for full description of the dataset), more than one-third of IBD patients admitted to the hospital with colonoscopy had a discharge diagnosis of IBD, indicating their admission was for medical treatment of their IBD. We therefore excluded this group so that providers who treat a disproportionate number of IBD patients will not be disadvantaged in the measure.
- A post-index diagnosis of IBD, which represents a very small fraction of cases (less than 0.5% of the cohort) in the measure population, indicates that the condition was likely present at the time of the index colonoscopy but not coded.

4) Colonoscopies for patients with a history of diverticulitis or diagnosis of diverticulitis at time of index colonoscopy or on the subsequent hospital visit outcome claim.
Rationale: We exclude these patients because:

- It is unclear what the health status is of patients coded with a history or current diagnosis of diverticulitis, making it difficult to fully risk adjust for patients' health. Colonoscopies performed on patients with a history or current diagnosis of diverticulitis are likely to be coded as diverticulitis as the primary diagnosis irrespective of whether the patients are undergoing a screening procedure or a diagnostic procedure (i.e., are acutely unwell with active disease). Furthermore, the codes for diverticulitis and diverticulosis may not be consistently used; patients with diverticulosis may be erroneously coded as diverticulitis. Therefore, we may not be able to adequately risk adjust as we cannot identify relatively well versus acutely unwell patients among visits coded as diverticulitis.

- Admissions for acutely ill patients with a history or current diagnosis of diverticulitis who are evaluated with an outpatient colonoscopy and are subsequently admitted for medical treatment of do not reflect the quality of the colonoscopy. In our 2010 Medicare 20% FFS Full Development Sample (see the Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy Measure Technical Report posted on the web page provided in data field S.1) more than one-quarter of patients with a history or current diagnosis of diverticulitis admitted to the hospital post colonoscopy had a discharge diagnosis of diverticulitis, indicating they were admitted for medical treatment of the condition. These admissions are likely unrelated to the quality of the colonoscopy. We therefore excluded this group so that providers who treat a disproportionate number of diverticulitis patients will not be disadvantaged in the measure.
- A post-index diagnosis of diverticulitis, which represents a very small fraction of cases (less than 0.5% of the cohort) in the measure population, indicates that the condition was likely present at the time of the index colonoscopy but not coded.

5) Colonoscopies followed by a subsequent outpatient colonoscopy procedure within 7 days.
Rationale: In these situations, the two colonoscopies are considered part of a single episode of care, for which the subsequent colonoscopy is considered the index procedure.

In addition, for colonoscopies performed at HOPDs, we exclude:

6) Colonoscopies that occur on the same day and at the same hospital as an emergency department (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.
Rationale: It is unclear whether the colonoscopy or ED visit occurred first. If the ED visit is coded with a diagnosis indicative of a complication of care, the measure assumes the ED visit occurred after the colonoscopy procedure and is counted in the measure. It is unlikely that a patient would experience an ED visit for an acute diagnosis at 1 facility and then travel to another facility for a routine colonoscopy on the same day. Accordingly, ED visits billed on the same day as a colonoscopy but at a different facility are included because they likely represent a routine procedure followed by a complication of care.

7) Colonoscopies that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.
Rationale: In these situations, it is not possible to use claims data to determine whether the colonoscopy was the cause of, subsequent to, or during the ED visit. However, if the ED visit is coded with a diagnosis for a complication, the assumption is that it occurred after the colonoscopy procedure.
8) Colonoscopies that are billed on the same hospital outpatient claim and that occur after the ED visit.
Rationale: In these situations, we assume that the colonoscopy was subsequent to the ED visit and may not represent a routine colonoscopy procedure. Timing of the ED visits is determined using revenue center dates from the outpatient claim.

9) Colonoscopies that are billed on the same hospital outpatient claim as an observation stay.
Rationale: In these situations, it is not possible to use claims data to determine whether the colonoscopy was the cause of, subsequent to, or during the observation stay.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
1) Colonoscopies for patients who lack continuous enrollment in Medicare FFS Parts A and B in the 7 days after the procedure.
Lack of continuous enrollment in Medicare FFS for 7 days after the procedure is determined by patient enrollment status in FFS Parts A and B using the Medicare Enrollment Database. The enrollment indicators must be appropriately marked for the month(s) which fall within 7 days of the procedure date.

2) Colonoscopies that occur concurrently with high-risk upper GI endoscopy procedures.
The list of the CPT codes for the upper GI endoscopy procedures identified as “high-risk” are in attached Data Dictionary, sheet “Colonos_Excl”

3) Colonoscopies for patients with a history of IBD or diagnosis of IBD at time of index colonoscopy or on the subsequent hospital visit outcome claim.
The ICD-9-CM and ICD-10-CM codes that define IBD are in the attached Data Dictionary, sheet "Colonos_Excl."

4) Colonoscopies for patients with a history of diverticulitis or diagnosis of diverticulitis at time of index colonoscopy or on the subsequent hospital visit outcome claim.
The ICD-9-CM and ICD-10-CM codes that define diverticulitis are in the attached Data Dictionary, sheet "Colonos_Excl."

5) Colonoscopies followed by a subsequent outpatient colonoscopy procedure within 7 days.
For cases in which a colonoscopy is followed by another colonoscopy within 7 days, the measure will use the subsequent colonoscopy as the index colonoscopy.

The following are in addition to those above, but only for HOPDs:

6) Colonoscopies that occur on the same day and at the same hospital as an ED visit that is billed on a separate claim than the index colonoscopy, unless the ED visit has a diagnosis indicative of a complication of care.
The billing and revenue center codes that define ED visits are in the attached Data Dictionary, sheet "Colonos_Outcome_ED_Obs." The same facility is defined as having the same CMS Certification Number (CCN). Complications of care codes (shown in tab "Colons_Excl_ED_CoC" include the following AHRQ CCS catgories: AHRQ CCS 257 – Other aftercare; AHRQ CCS 238 – Complications of surgical procedures or medical care; AHRQ CCS 2616 - Adverse effects of medical care; AHRQ CCS 2617 - Adverse effects of medical drugs; and ICD-10-CM G89.18 – Other acute postprocedural pain.

7) Colonoscopies that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care.
The billing and revenue center codes that define ED visits are in the attached Data Dictionary, sheet "Colonos_Outcome_ED_Obs." Complications of care codes (shown in tab "Colons_Excl_ED_CoC" include the following AHRQ CCS catgories: AHRQ CCS 257 – Other aftercare; AHRQ CCS 238 – Complications of surgical procedures or medical care; AHRQ CCS 2616 - Adverse effects of medical care; AHRQ CCS 2617 - Adverse effects of medical drugs; and ICD-10-CM G89.18 – Other acute postprocedural pain.

8) Colonoscopies that are billed on the same hospital outpatient claim and that occur after the ED visit.
The billing and revenue center codes that define ED visits are in the attached Data Dictionary, sheet ""Colonos_Outcome_ED_Obs."

9) Colonoscopies that are billed on the same hospital outpatient claim as an observation stay.
The billing and revenue center codes that define observation stays are in the attached Data Dictionary, sheet "Colonos_Outcome_ED_Obs."

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
N/A. This measure is not stratified.

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Statistical risk model
If other:

S.12. **Type of score:**
Rate/proportion
If other:

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

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

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

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

3. Identify and create a binary (0/1) flag for an unplanned hospital visit within 7 days of the colonoscopy described above in Section S.5.

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

5. Fit a hierarchical generalized linear model (HGLM) to produce a ratio of the number of "predicted" hospital visits to the number of "expected" hospital visits for each facility, given its case mix. The HGLM is adjusted for clinical risk factors that vary across patient populations, are unrelated to quality, and influence the outcome.

6. Multiply the ratio estimated in step 3 by the observed national 7-day hospital visit rate to obtain a risk-standardized hospital visit (RSHV) rate for each facility.

7. Use bootstrapping to construct a 95% confidence interval estimate for each facility's RSHV rate.

For more information about the measure methodology, please see the Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy 2018 Measure Updates and Specifications Report posted on the web page provided in data field S.1.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.
N/A

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
Specify calculation of response rates to be reported with performance measure results.
N/A

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

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.
Medicare administrative claims and enrollment data

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

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

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

If other:

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

2. Validity – See attached Measure Testing Submission Form
Colonoscopy_nqf_testing_attachment_V3_FINAL_010520.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
Yes

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
No

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

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

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.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]
AND
If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14],[15] and has demonstrated adequate discrimination and calibration;

OR
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) 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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*).

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

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

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

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

The number of patients 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	**Description of Dataset**	• **Use and Section in the Testing Attachment**

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 Number of facilities: 4034 Number of facilities with >=30 procedures: 3583 Mean age of patients: 72.6 % female: 53.4% ASCs: Number of procedures: 2,524,898 Number of facilities: 2261 Number of facilities with >=30 procedures: 2073 Mean age of patients: 72.3 % 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].

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

2. 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

Statistic	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. Patient One of Two
17. Patient Two of Two

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*

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
Exclusions	**Number of Procedures**	**% of All Included Procedures**
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]	2,258,6	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.

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, α_i which we assume follows a normal distribution with mean μ and variance τ^2 , the between-hospital variance component. The following equations *define* the HGLM:

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

$$where\ \alpha_i=\mu+\omega_i;\ \omega_i{\sim}N(0,\tau^2)$$

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

Where Y_{ij} 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; $Z_{ij}=\left(Z_{ij1},Z_{ij2},\ldots,Z_{ijp}\right)^T$ is a set of *p* patient-specific covariates derived from the data; and *I* denotes the total number of facilities and n_i is the number of outpatient colonoscopies performed at facility *i*. The facility-specific intercept, or effect, of the *i-th* facility, α_i, defined above, comprises μ, the adjusted average intercept over all facilities in the sample, and ω_i, the facility-specific intercept deviation from μ . A point estimate of ω_i, greater or less than 0, determines whether 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.

[a]See 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 Variables[a]:

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)

Variable (CC)	Odds Ratios (95% CI)
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)

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.

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 ([a]See 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

Social risk factor	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

Social risk factor	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

Model	***Bivariate***			***Multivariate***		
Social risk factor	**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

Model	***Bivariate***			***Multivariate***		
Social risk factor	**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

Social risk factor	**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

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).

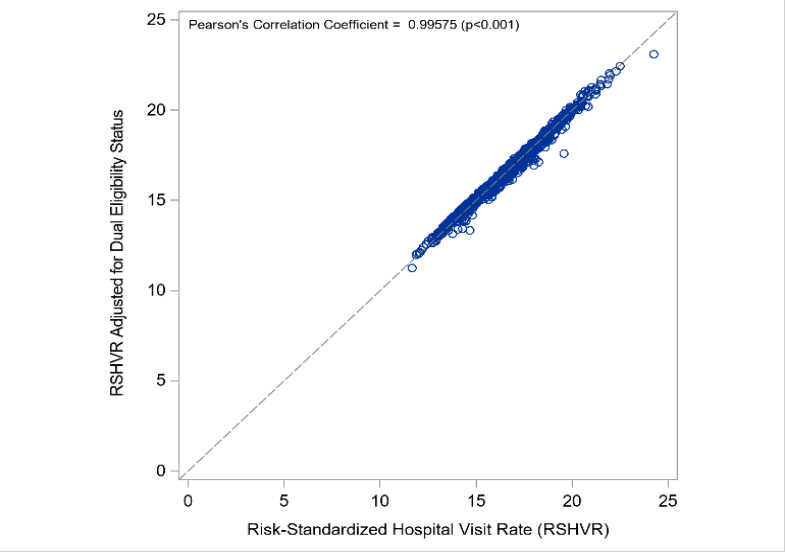
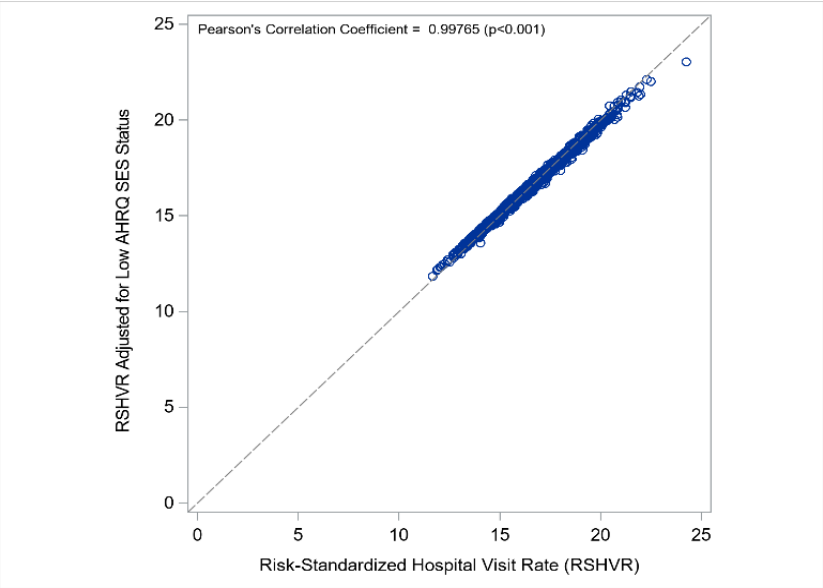


Figure 1B. HOPDs: Correlation of measure scores (RHSVRs) calculated with and without social risk factor adjustment for low AHRQ SES (for facilities with at least 30 cases).



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).

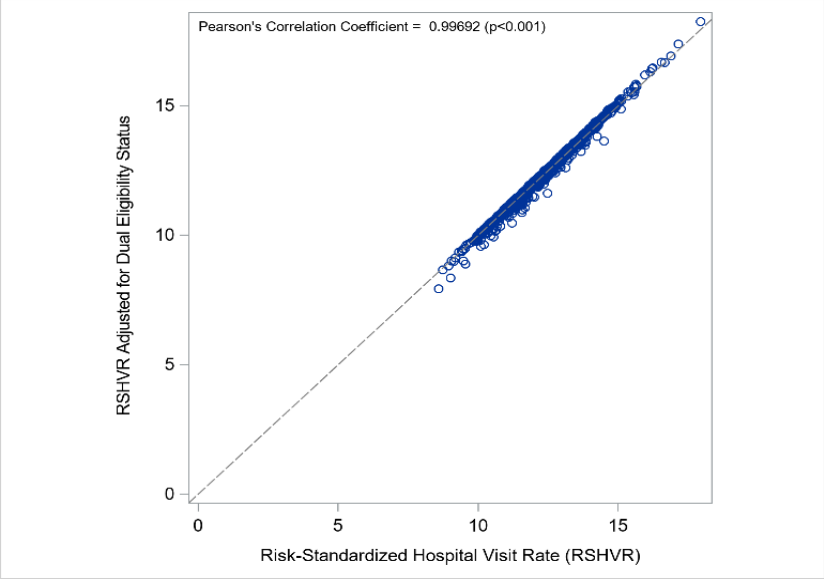
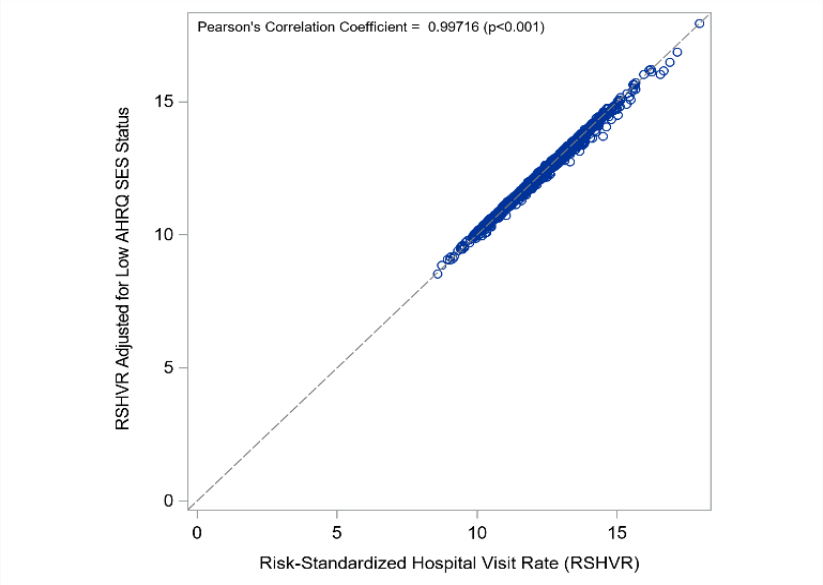


Figure 2B: ASCs: Correlation of measure scores (RHSVRs) calculated with and without social risk factor adjustment for low AHRQ SES (for facilities with at least 30 cases).



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)

Characteristic	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)

Characteristic	**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

Spearman correlation coefficient: 0.126

Figure 3B: Low AHRQ SES

Spearman correlation coefficient: 0.140

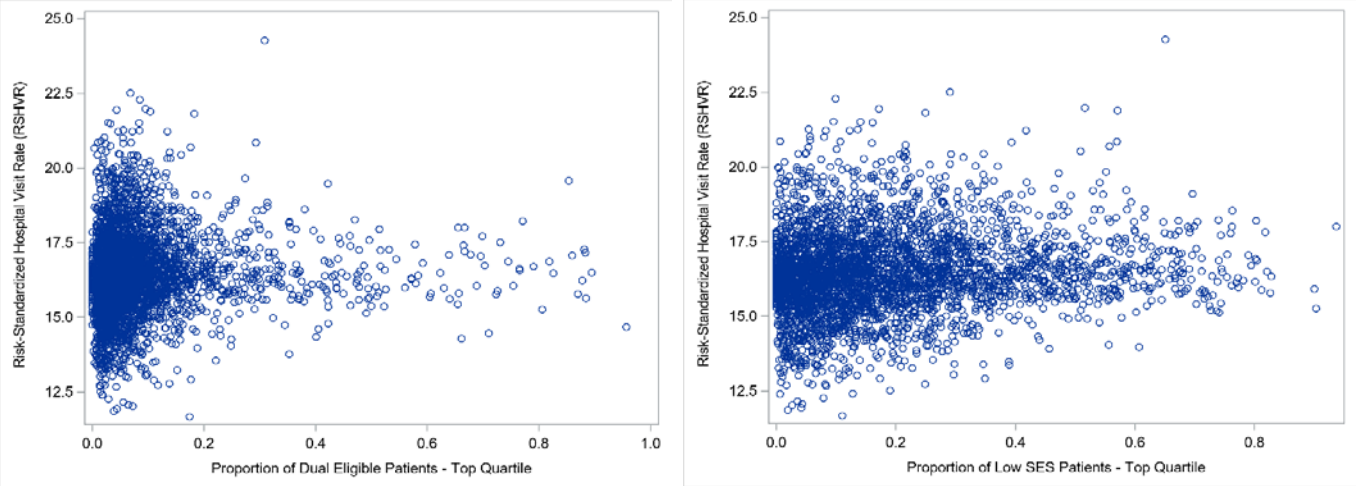


Figure 4A and 4B: ASCs: Relationship between the proportion of patients with dual-eligible status (A) and low AHRQ SES (B) and the risk-standardized hospital visit rates (RSHVRs) (in facilities in the highest quartile for the proportion patients with the social risk factor; facilities with at least 30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

Figure 4A: Dual Eligible

Spearman correlation coefficient: 0.057

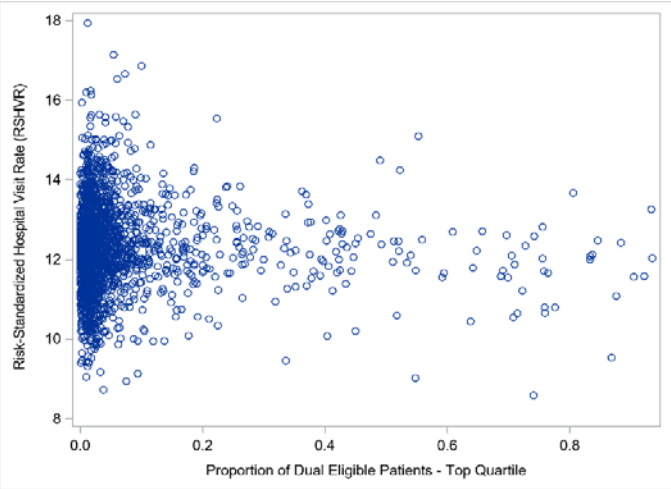


Figure 4B: Low AHRQ SES

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

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

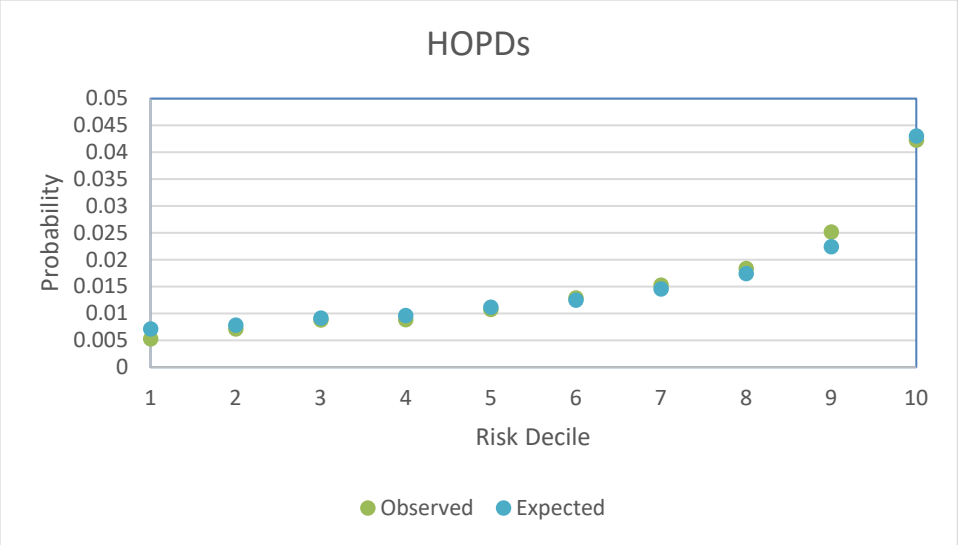
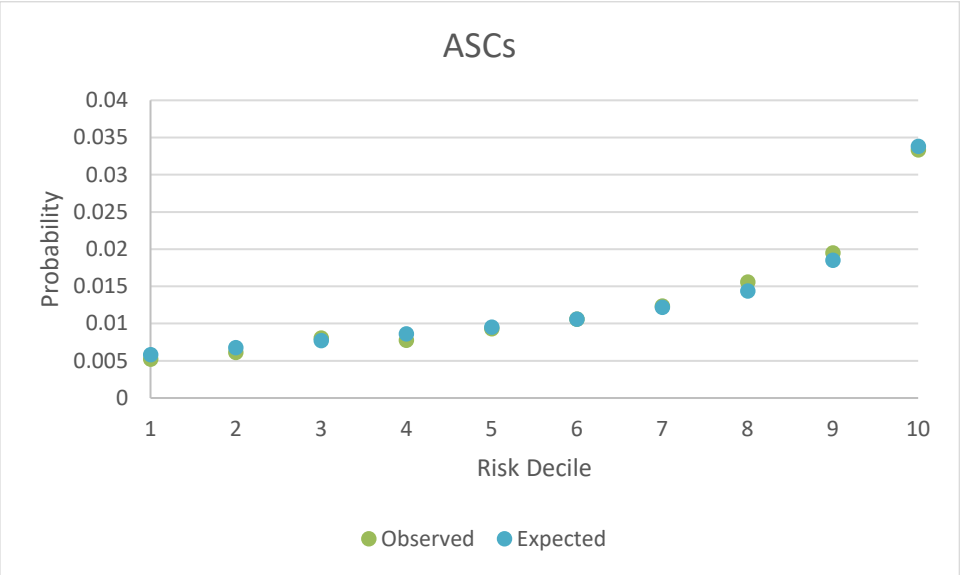


Figure 5B: ASCs: Plot of observed vs. expected values for risk deciles (2017 performance period – Dataset #1) for:



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%

Variable (CC)	01/2016-12/2016	01/2017-12/2017	01/2018-12/2018
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.

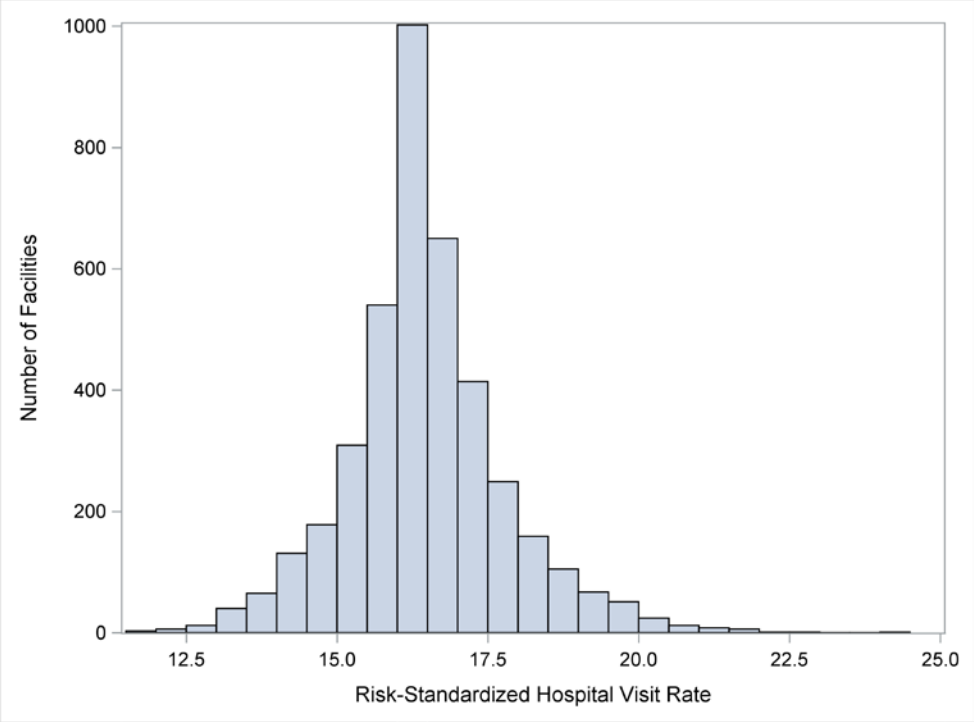
The risk-standardized measure scores (RSHVRs) for 4,034 HOPDs estimated using Medicare FFS data (2016-2018 performance period) had a median value of 16.4 hospital visits per 1,000 colonoscopies. The values ranged from 11.7 to 24.3. The percentiles of the distribution are shown in Table 11. Figure 6 shows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

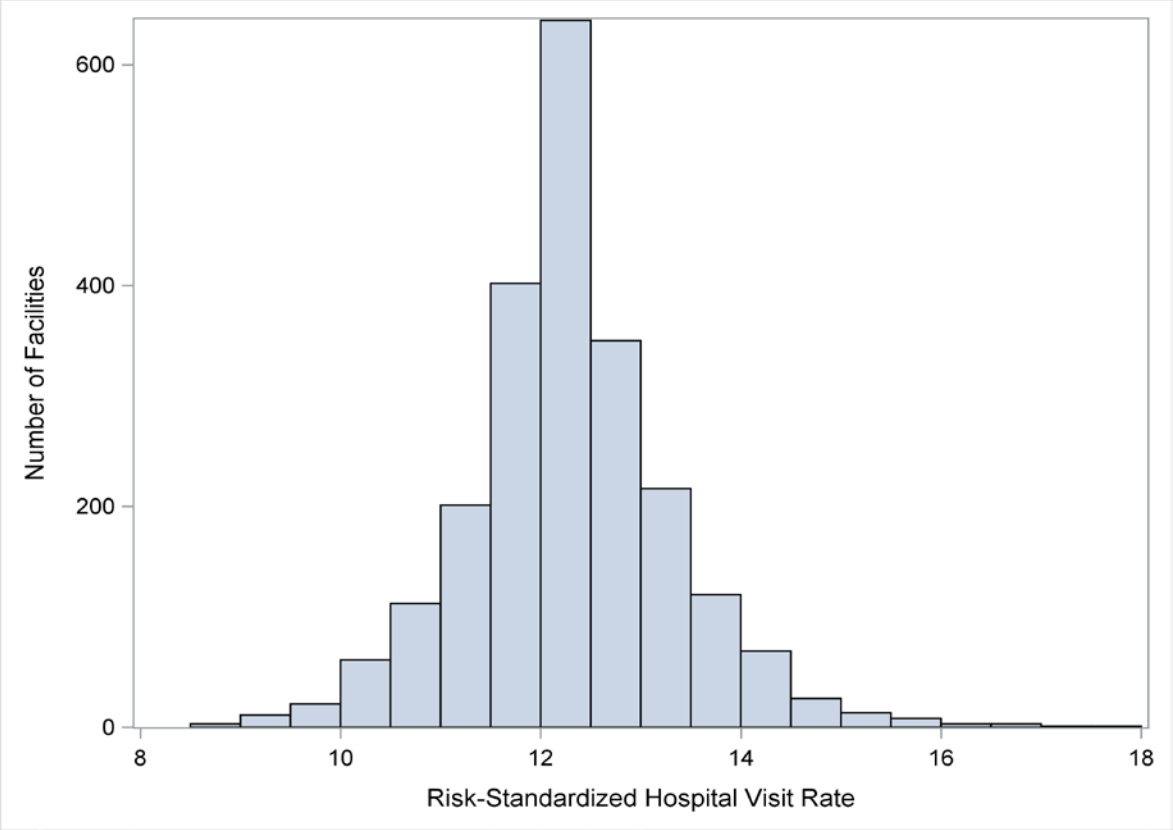
The RSHVRs for ASCs estimated using Medicare FFS data (2016-2018 performance period) had a median value of 12.23 hospital visits per 1,000 colonoscopies. The values ranged from 8.59 to 17.94. The percentiles of the distribution are shown in Table 12. Figure 7 shows a histogram of the distribution.

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.

3. Feasibility

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

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues. IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.**
This is a claims-based measure, data is generated during the course of billing. There have been no difficulties regarding data collection, availability of data, missing data, etc. Because completion of claims is required for hospital reimbursement, there is little missing data. The measures do not require any additional data collection and offer no data collection burden to facilities.

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

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees, licenses or other requirements needed to use this measure as specified.

3. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Not in use	Public Reporting Ambulatory Surgical Center Quality Reporting (ASCQR) https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/asc-quality-reporting/index.html Hospital Outpatient Quality Reporting (HOQR) https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitaloutpatientqualityreportingprogram.html Payment Program Pay for Reporting in ASCQR Pay for Reporting in HOQR https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/asc-quality-reporting/index.html https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitaloutpatientqualityreportingprogram.html

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Program Name, Sponsor: Hospital Outpatient Quality Reporting (HOQR) Program, CMS

Implemented by CMS for outpatient services, the Hospital OQR is a national pay for quality data reporting program mandated by the Tax Relief and Health Care Act of 2006. This act requires hospitals to submit data on measures on the quality of care furnished by hospitals in outpatient settings. The HOQR program provides hospitals with a financial incentive to report their quality of care measure data and CMS with data to help Medicare beneficiaries make more informed decisions about their health care. The measure includes all short-term acute care hospitals with eligible colonoscopies (excluding PPS-exempt cancer hospitals). For the final cohorts from January 1, 2016 – December 31, 2018, there were 2,258,661 colonoscopies performed in 4034 HOPDs, representing about 91% of all eligible colonoscopies.

Program Name, Sponsor: Ambulatory Surgical Center Quality Reporting (ASCQR) Program CMS

The ASCQR Program is a national pay-for-reporting, quality data program finalized by CMS under which ASCs report quality of care data for standardized measures to receive the full annual update to their ASC annual payment rate. Measured entities include all ambulatory surgical centers with eligible colonoscopies. For the final cohort from

January 1, 2016 – December 31, 2018 there were 2,524,898 procedures performed across 2,261 ASCs, representing 93.4% of all eligible colonoscopies.

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

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

4a2.1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
In July 2015, CMS held a dry run of the colonoscopy measure. The primary goals of the dry run were to educate HOPD and ASC facilities about the measure in advance of its use in public reporting, allow facilities to review data and results, provide facilities with the opportunity to ask questions about the measure, and to test the measure production process. All open facilities that had at least one qualifying colonoscopy for the measure during the performance period were provided with Facility-Specific Reports (FSRs) containing their measure results and detailed patient-level data. Additionally, claims detail reports (CDRs) were made available to facilities at three stages (September and December of 2017, and March of 2018) prior to the final measure calculation and public reporting of measure results. The CDRs provided facilities subject to the measure with information on their colonoscopy cases that would be included in the measure calculation for January 2019 public reporting. Facilities were also provided with information to help them understand the measure, interpret their data and measure results, and facilities could comment on or ask questions through an email Question & Answer (Q&A) inbox.

During the dry run, measure results were confidentially reported to 4,069 HOPDs and 1,160 ASCs with active QualityNet Secure Portal accounts. Of these, 2,955 (72.6%) HOPDs and 580 (50.0%) ASCs downloaded their reports.

For 2019 public reporting, measure results were reported to 3791 HOPDs and 1327 ASCs; reports were downloaded by 2480 HOPDs (65%) and 443 ASCs (33.4%). For 2020 public reporting, measure results were reported to 4190 HOPDs and 1097 ASCs; reports were downloaded by 2915 HOPDS (69.6%) and 326 ASCs (29.7%).

4a2.1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
A dry run is a preliminary analysis of data in which facilities may review their measure results and ask questions about and become familiar with the measure methodology. The 2015 dry run consisted of a 30-day period of confidential reporting (from July 1 to July 31, 2015) during which facilities had the opportunity to review their measure results and the data used in measure calculation. Two national provider calls were held to provide further information on the measure and answer questions.

In anticipation of public reporting in January 2019, CMS provided facilities with interim reports with their cases and outcomes for confidential review. Facilities were provided with three interim claims-detail reports, and in October 2018 they were provided with a full facility-specific report with their results and all cases for the 2017 performance period.

For public reporting, in January 2020, CMS provided facilities with their facility-specific reports with their results and all included cases for the performance period (January 1, 2016-December 31, 2018).

4a2.2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.

Stakeholder feedback was obtained during the dry run national provider calls and the dry run email Q&A period. Feedback continues to be gathered through email Q&A.

4a2.2.2. Summarize the feedback obtained from those being measured.
Before and during the dry run, CMS received 437 emails (478 including follow-up questions) via the colonoscopy measure email inbox. Facilities asked for assistance interpreting their patient-level data, asked questions about the measure methodology, and in a small number of cases, flagged findings in their report that seemed inconsistent with the methodology. The topics of the questions and comments raised during the national provider calls were similar to those received by email.

A wide variety of question topics were received in the measure inbox during the dry run period. The most common types of questions were inquiries about specific cases in facilities' data (40%), followed by requests for assistance accessing the FSR on the QualityNet website (23%), questions about the dry run process or the national provider calls (16%), and general methods questions (15%).

Facilities' careful review of patient data identified a number of situations that suggest the need to make minor refinements to the measure methodology to ensure: (a) the algorithm for processing claims data accurately identifies cases for inclusion in the measure; and (b) the planned admission algorithm captures additional planned hospital visits.

Specifically:

The feedback identified several types of patient cases that may not have been properly identified and classified by the measure algorithms in the claims data:

- The patient was in observation status before the colonoscopy was performed, but the measure counted the observation stay as an unplanned hospital visit following the colonoscopy. Hospitals identified situations in which a patient was placed into observation status before the colonoscopy, either to evaluate acute symptoms such as a GI bleeding or to complete the preparation for the procedure. These were situations in which the colonoscopy and observation stay were billed on the same outpatient claim.

- The hospital visit was planned but was still counted in the measure outcome. Stakeholders reported cases for which they considered the follow-up hospital visits to be planned, but those visits were counted in the measure outcome. These included situations where the admission was (1) for treatment to address an issue found during the colonoscopy (such as cancer); (2) a planned procedure (such as colectomy, ileostomy take-down, and rectoplexy) for which the colonoscopy was part of the pre-operative workup; and (3) a planned surgery unrelated to the colonoscopy (such as renal artery stent surgery).

- The colonoscopy was performed while the patient was a hospital inpatient. Stakeholders identified situations where a colonoscopy was performed after the patient was in inpatient status, but the case was included in the measure as an outpatient colonoscopy with a hospital admission outcome.

- The colonoscopy was performed after an ED visit on the same day. These situations involved an ED visit and a colonoscopy billed on separate claims on the same day. The colonoscopy was not excluded from the measure.

- The colonoscopy was performed at another facility or a different procedure was performed. In these situations, facilities notified CMS of instances in which a procedure attributed to their facility was performed at another facility, or that their records indicated that a procedure other than a qualifying colonoscopy was performed.

- There were cases that facilities felt should have been excluded, but were not. Facilities questioned why certain cases were not excluded for a history of diverticulitis or IBD or for a concurrent upper GI endoscopy.

Summary of Questions or Comments from Hospitals submitted through the Q & A process:
For the Colonoscopy measure inquiries received from hospitals since January 2018 have included the following:
1. Requests for clarification of how inclusion and exclusion criteria are applied; and
2. Requests for interpretation and clarification of results.
3. Questions about the characterization of specific procedures as planned or unplanned.

4a2.2.3. Summarize the feedback obtained from other users
We have not received feedback from other users.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.
Each year issues raised through the Q&A or in the literature related to this measure are considered by measure and clinical experts. Any issues that warrant additional analytic work due to potential changes in the measure specifications are addressed as a part of annual measure reevaluation. If small changes are indicated after additional analytic work is complete, those changes are usually incorporated into the measure in the next measurement period. If the changes are substantial CMS may propose the changes through rulemaking and adopt the changes only after CMS received public comment on the changes and finalizes those changes in the OPPS or other rule.

The current measure specifications submitted with this application reflect the information gathered during the dry run and was used for measure implementation for the calendar year 2020 payment determination for the HOQR and ASCQR programs. These updates are discussed in detail in section S.3.2 and include: 1) Modification of the PAA to align with changes made to CMS's Planned Readmission Algorithm version 4.0 2020, 2) Update to exclusions for surgeries: excluding surgeries that occur on the same day and on the same claim as the colonoscopy, unless there is a diagnosis of complication of care indicated on the claim, and excluding colonoscopies that are billed on the same hospital outpatient claim, but occur after the ED visit, regardless of whether complications of care are billed or not. For more information about these updates, please see the Facility 7-day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy 2019 (version 5.0) Measure Updates and Specifications Report posted on the web page provided in data field S.1.

Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
As mentioned above in section 1.b2, hospital visit rates among HOPDs declined for 2019 reporting compared to 2018 (from 16.4 per 1000 cases in 2018 reporting to 14.8 per 1000 cases in 2019 reporting). The distribution of risk-standardized rates also declined for HOPDs; the interquartile range of rates for 2019 reporting lie completely below the 2018 interquartile range. This decline may reflect quality improvement as there were no large specification changes to the measure for 2019 reporting that would impact rates, nor were there noticeable differences in patient mix. Hospital visit rates did not decline between 2019 and 2020 public reporting, but this is likely due to a change in the measure methodology, which now uses data from January 1, 2016 through December 31, 2018, and therefore the performance data between the two public reporting periods overlap.

There was a small decline in the hospital visit rates for ASCs across the three public reporting years (2018, 2019, 2020). Historically, CMS engagement with ASCs has been lower than with hospitals. For example, about 33% of ASCs (443 of 1327 facilities) downloaded a facility-specific report containing their performance data in October 2018, compared with about 65% of HOPDs (2480 of 3791 facilities).

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

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.
We have encountered no unexpected findings during implementation, including unintended impacts on patients.

4b2.2. **Please explain any unexpected benefits from implementation of this measure.**
We have identified no unexpected benefits.

3. Comparison of Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0658 : Appropriate Follow-Up Interval for Normal Colonoscopy in Average Risk Patients
2687 : Hospital Visits after Hospital Outpatient Surgery
3357 : Facility-Level 7-Day Hospital Visits after General Surgery Procedures Performed at Ambulatory Surgical Centers
3510 : Screening/Surveillance Colonoscopy

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

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We identified two colonoscopy-related measures that are currently endorsed by NQF. One (NQF 0658) is a process measure that identifies the percentage of patients aged 50 years to 75 years who received a screening colonoscopy and who had a recommended follow-up interval of at least 10 years for repeat colonoscopy documented in their colonoscopy report. The second measure (NQF 3510) is a cost measure. Both measures are process measures related to screening, and while both measures address colonoscopy, these measures differ from the CMS colonoscopy measure, which is an outcome measure. More information on each of the related colonoscopy measures is provided below. 1. NQF 0034: 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. NQF 0034 focuses on colonoscopy screening in patients aged 50-75, therefore the targeted population overlaps with the CMS colonoscopy measure and reflects overall screening guidelines. The CMS colonoscopy outcome measure's purpose is to measure outcomes from colonoscopy procedures in Medicare-aged patients. 2. NQF 3510: Screening/Surveillance Colonoscopy The Screening/Surveillance Colonoscopy cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure and includes costs of services that are clinically related to the attributed clinician's role in managing care for 14 days from the "trigger" of the episode. NQF 3510 has the same target population (Medicare beneficiaries) and would capture the physician-controlled costs related to hospital visits identified in the CMS colonoscopy measure. The timeframe for the two measures differs (7 days for the outcome measure vs. 14 days for the cost measure), and the level of measurement differs (facility-level for the outcome measure, and clinician or group level for the cost measure). We also identified two related NQF-endorsed outcome measures: 1. NQF 3357: Facility-Level 7-Day Hospital Visits after General

Surgery Procedures Performed at ASCs (ASC General Surgery), and 2. NQF 2687: Hospital Visits after Hospital Outpatient Surgery (HOPD Surgery). The outcome of both measures is the same as CMS's colonoscopy measure presented in this re-endorsement application; an unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission. Hence, these related measures target the same quality domains as the CMS colonoscopy measure. The patient cohort is also somewhat similar in that the related 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 surgical procedures, not colonoscopy. The CMS colonoscopy measure is a claims-based measure, therefore any differences in measure specifications create no burden to facilities as the measures are calculated from data produced during the billing process. In terms of interpretability, the CMS colonoscopy measure is an outcome measure, and therefore is conceptually distinct from the process measure and the cost measure; the cost measure also targets a different level of measurement (provider, not facility). The outcome for the CMS colonoscopy measure is harmonized with the related NQF-endorsed outcome measures for these settings (ASCs/HOPDs), as discussed in section 5a1.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable. There are no competing measures, only related measures.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Colonoscopy_Measure_Appendix_FINAL_02-21-14.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Nicole, Hewitt, nicole.hewitt@cms.hhs.gov, 410-786-7778-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Elizabeth, Drye, elizabeth.drye@yale.edu, 203-764-5700-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
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.

External Working Group members were:
John Allen, MD, MBA, Clinical Chief of Gastroenterology and Hepatology, Yale School of Medicine President Elect, American Gastroenterological Association

Ronald Vender, MD
Professor of Medicine (Digestive Diseases) and Associate Dean for Clinical Affairs, Yale School of Medicine Chief Medical Officer, Yale Medical Group Immediate Past President, American College of Gastroenterology

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.

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

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2014
Ad.3 **Month and Year of most recent revision:** 11, 2019
Ad.4 **What is your frequency for review/update of this measure?** Annual
Ad.5 **When is the next scheduled review/update for this measure?** 11, 2020

Ad.6 **Copyright statement:** Not applicable
Ad.7 **Disclaimers:** Not applicable

Ad.8 **Additional Information/Comments:** Not applicable