## CBE ID

3510

## Title

Screening/Surveillance Colonoscopy

## Project

Cost and Efficiency

## Endorsement Status

Endorsed

## Is Under Review

Yes

## Next Maintenance Cycle

Spring 2026

## Previous Endorsement Cycle

Spring 2019

## Initial Endorsement

Wed, 11/13/2019 - 05:46

## Steward

Centers for Medicare & Medicaid Services

## 1.0 New or Maintenance

Maintenance

## 1.1 Measure Structure

Single Measure

## 1.3 Electronic Clinical Quality Measure (eCQM)

No

## 1.6 Measure Description

The Screening/Surveillance Colonoscopy episode-based cost measure evaluates the risk-adjusted cost to Medicare for beneficiaries who receive this procedure. Performance is assessed at the clinician (Tax Identification Number-National Provider Identifier [TIN-NPI]) and group (Tax Identification Number [TIN]) levels. A measure score reflects the average risk-adjusted cost per episode across all episodes attributed to the clinician or group during the performance period. This measure includes costs of services clinically related to the attributed clinician's role in managing care from the day of the clinical event that triggers the episode through 14 days after the trigger. Beneficiaries eligible for this measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

## 1.6a Material Specification Change(s)

No

## 1.7 Measure Type

Cost/Resource Use

## 1.8 Level of Analysis

Clinician: Group/Practice, Clinician: Individual

## 1.9 Care Setting

Ambulatory Care: Office, Ambulatory Surgery Center, Hospital: Outpatient

## 1.10 Measure Rationale

Colorectal cancer (CRC) is among the most prevalent and preventable cancers in the United States.[1] Screening colonoscopy is the most common CRC screening procedure performed among U.S. adults.[2] The 2021 U.S. Preventive Services Task Force (USPSTF) colorectal cancer screening guidelines recommend screening for all adults aged 45–75 at average risk, reflecting a lowered starting age based on evidence of rising incidence in younger adults.[3] In performance year (PY) 2024, the measure attributed 977,369 episodes to 3,205 groups and 15,017 clinicians, one of the highest-volume procedural episode groups in the Merit-based Incentive Payment System (MIPS), covering approximately 0.3 percent of total Medicare Part A and Part B expenditures.[4]

Cost performance is a required component of the MIPS composite score under the Medicare Access and CHIP Reauthorization Act (MACRA). Episode-based cost measures attribute Medicare expenditures to clinicians for services clinically related to their role in managing a discrete clinical event, giving clinicians and groups actionable information about their resource use relative to peers. The Screening/Surveillance Colonoscopy episode-based cost measure was first implemented in MIPS in PY2019,[5] developed under MACRA section 101(f) based on input from the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee, which identified the measure's impact in terms of patient population size, clinician coverage, and the opportunity to incentivize cost-effective CRC screening care. Beginning with the 2025 performance year, the measure is also available within the Gastroenterology Care MIPS Value Pathway (MVP), which provides gastroenterologists, nurse practitioners, and physician assistants a specialty-specific reporting framework that includes this measure as a core cost component alongside the Total Per Capita Cost measure.[6]

Measuring cost for this episode is appropriate because meaningful cost variation within the 14-day episode window is attributable to clinician practice patterns rather than patient complexity alone. Bowel preparation quality is a clinician-influenced factor that directly determines whether a colonoscopy can be completed. Inadequate bowel preparation is associated with an increased likelihood of an incomplete procedure (one that cannot adequately visualize the entire colon), requiring a repeat examination within the episode window and adding directly to episode costs.[7] Variation in the use of anesthesiologist services during screening colonoscopy is a second clinician-attributable cost driver. American Society for Gastrointestinal Endoscopy (ASGE) guidelines on sedation and anesthesia in GI endoscopy state that anesthesia provider-

administered sedation is more expensive and does not appear to result in improved safety compared with endoscopist-directed sedation for routine endoscopic procedures in low-risk patients.[8] In the Medicare population, anesthesiologist involvement in screening colonoscopy increased from 11 percent of procedures in 2001 to 23 percent by 2006, with each anesthesiologist-assisted procedure associated with approximately 20 percent higher cost.[9] Nationally representative hospital data confirm this trend continued, with anesthesia assistance used in 58 percent of outpatient colonoscopies by 2015.[10]

Footnotes:

[1] Siegel RL, Wagle NS, Star J, Kratzer TB, Smith RA, Jemal A. Colorectal cancer statistics, 2026. *CA Cancer J Clin*. 2026;76(2):e70067. PMID: 41769777.

[2] He Y, Xu T, Fang J, et al. Trends in colorectal cancer screening in the United States, 2012 to 2020. *J Med Screen*. 2023;30(3):125-133. PMID: 37157812.

[3] Davidson KW, Barry MJ, Mangione CM, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA*. 2021;325(19):1965-1977. PMID: 34003218.

[4] Centers for Medicare & Medicaid Services. 2026 MIPS Summary: Cost Measures. CMS; 2026. https://www.cms.gov/files/document/2026-mips-summary-cost-measures.pdf.

[5] Centers for Medicare & Medicaid Services, HHS. Medicare program; revisions to payment policies under the physician fee schedule and other revisions to Part B for CY 2019; Medicare Shared Savings Program requirements; Quality Payment Program; and Medicaid Promoting Interoperability Program. Final rule. *Fed Regist*. 2018;83(226):59452-60303.

[6] Centers for Medicare & Medicaid Services. Gastroenterology Care MIPS Value Pathway (M1422). Quality Payment Program. https://qpp.cms.gov/mips/explore-mips-value-pathways/2025/M1422. Accessed April 30, 2026.

[7] Jacobson BC, Anderson JC, Burke CA, et al. Optimizing bowel preparation quality for colonoscopy: consensus recommendations by the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol*. 2025;120(4):738-764. PMID: 40035345.

[8] Early DS, Lightdale JR, Vargo JJ, et al. Guidelines for sedation and anesthesia in GI endoscopy. *Gastrointest Endosc*. 2018;87(2):327-337. PMID: 29306520.

[9] Khiani VS, Soulos P, Gancayco J, Gross CP. Anesthesiologist involvement in screening colonoscopy: temporal trends and cost implications in the Medicare population. *Clin Gastroenterol Hepatol*. 2012;10(1):58-64. PMID: 21782768.

[10] Krigel A, Chen L, Wright JD, Lebwohl B. Substantial increase in anesthesia assistance for outpatient colonoscopy and associated cost nationwide. *Clin Gastroenterol Hepatol*. 2019;17(12):2489-2496. PMID: 30625407.

## 1.11 Measure Webpage

https://qpp.cms.gov/docs/cost_specifications/2025-12-py2026-mif-ebcm-ss-clnscpy...

### 1.13 Data Dictionary

Attached

### 1.13a Attach Data Dictionary

2025-12-py2026-codes-list-ss-clnscpy.xlsx

### 1.14 Numerator

The cost measure numerator is the sum of the ratio of observed to expected payment-standardized cost to Medicare for all Screening/Surveillance Colonoscopy episodes attributed to a clinician or group. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

### 1.14a Numerator Details

The cost measure numerator is the sum of the ratio of observed to expected payment-standardized cost to Medicare for all Screening/Surveillance Colonoscopy episodes attributed to a clinician or group. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

Episodes are defined by Current Procedural Terminology (CPT) and Healthcare Common Procedure Coding System (HCPCS) colonoscopy procedure codes on Part B Physician/Supplier claims. An episode is triggered when a claim line with a trigger code is billed by a MIPS-eligible clinician, does not have a post-operative modifier code, and is the highest-cost trigger claim for that beneficiary on that date. The episode begins on the trigger date and extends through 14 days after the trigger. Full trigger code specifications are provided in the Measure Codes List (Triggers tab) and Measure Information Form (MIF) Section 4.1.

### 1.15 Denominator

The denominator is the total number of episodes from the Screening/Surveillance Colonoscopy episode group attributed to a clinician or group during the performance period. Episodes are constructed from Medicare Parts A and B fee-for-service claims for beneficiaries who undergo a screening or surveillance colonoscopy procedure that triggers an episode (see Field 1.14a for trigger logic and 1.18 Step 1 for episode construction).

### 1.15a Denominator Details

A clinician is attributed for an episode when they bill a trigger code on the day of the procedure as the main performing clinician. At the group level, the episode is attributed to the group associated with the trigger claim; at the clinician level, the episode is attributed to the clinician within that group who billed the trigger code. This is a performing physician attribution model based on who performed the colonoscopy. Full attribution specifications are provided in the codes list (Attribution tab) and MIF Section 2.0, Step 2.

### 1.15b Denominator Exclusions

Two types of denominator exclusions apply to this measure: measure-specific exclusions and standard episode-based cost measure (EBCM) exclusions, described in Field 1.15c.

## 1.15c Denominator Exclusions Details

Measure-specific exclusions remove episodes where: an endoscopic mucosal resection procedure is billed on the same day as the trigger; the episode is triggered in an inpatient hospital or emergency room setting; an upper gastrointestinal (GI) endoscopy is performed on the same day as the trigger procedure; or the patient has a diagnosis of inflammatory bowel disease, Crohn's disease, ulcerative colitis, microscopic colitis, or diverticular disease in the 120 days prior to the trigger. Full exclusion code specifications are provided in the codes list (Exclusions and Exclusions_Details tabs) and MIF Section 4.4.

Standard EBCM exclusions remove episodes where: the patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period; the patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window, or was enrolled in Medicare Part C for any portion; no main clinician is attributed the episode; the patient's date of birth is missing; the patient's death date occurred before the episode end date; or the episode trigger claim was not performed in an ambulatory/office-based care, outpatient hospital, or ambulatory surgery center (ASC) setting based on its place of service. Standard exclusion specifications are provided in MIF Section 4.4.

## 1.15d Age Group

Older Adults (65 years and older)

## 1.16 Type of Score

Ratio

## 1.17 Type of Cost Measure

Per episode

## 1.18 Calculation of Measure Score

The calculation of Screening/Surveillance Colonoscopy episode-based cost measure scores involves two overarching processes: episode construction (Steps 1–3) and measure calculation (Steps 4–6).

Step 1: Trigger and define episodes. Episodes are identified from Part B Physician/Supplier claims with a colonoscopy trigger code. An episode is triggered when the claim line is filed by a MIPS-eligible clinician who is the main performing clinician (no post-operative modifier present) and represents the highest-cost trigger claim for that beneficiary on that date. The episode window begins on the trigger date and extends through 14 days after the trigger. At the time of triggering, each episode is placed into one of three mutually exclusive care-setting sub-groups (ambulatory surgery center [ASC], office, and hospital outpatient department [HOPD]) based on the place of service on the trigger claim. After exclusions (Step 4), the denominator is the total number of

Screening/Surveillance Colonoscopy episodes attributed to a clinician or group. Full trigger code specifications are provided in the Triggers tab of the codes list and in MIF Section 4.1.

Step 2: Attribute episodes. An episode is attributed to the clinician who billed the qualifying trigger claim identified in Step 1 on the day of the procedure. At the group level, the episode is attributed to the group associated with that trigger claim. Full attribution logic and eligible clinician codes are provided in the Attribution tab of the codes list and in MIF Section 4.2.

Step 3: Assign costs and calculate standardized observed episode cost. Clinically related services occurring during the episode window are assigned to the episode using service assignment rules. Examples of clinically related services include anesthesia for the colonoscopy procedure, emergency department visits for post-procedure complications such as hemorrhage or perforation, and inpatient admissions for surgical management of complications. Costs are measured as payment-standardized Medicare allowed amounts, which include both Medicare trust fund payments and applicable beneficiary cost-sharing (deductibles and coinsurance). Payment standardization removes the effect of geographic payment adjustments (including hospital wage indexes, geographic price cost indices [GPCIs], and teaching hospital adjustments) to allow fair cost comparisons across providers in different markets regardless of where care is furnished. Standardized payment amounts are drawn from the Integrated Data Repository (IDR). The sum of payment-standardized allowed amounts across all assigned services constitutes the standardized observed episode cost. For more information on the CMS payment standardization methodology, see the "CMS Parts A and B Payment Standardization — Basics" and "CMS Parts A and B Payment Standardization — Detailed Methods" documents on the CMS Price (Payment) Standardization Overview page. Service assignment specifications are provided in the Service_Assignment tab of the codes list and in MIF Section 4.3.

Step 4: Exclude episodes. Before measure calculation, standard EBCM exclusions and measure-specific exclusions are applied. Standard exclusions remove episodes where: the patient had a primary payer other than Medicare during any part of the episode window or 120-day lookback period; the patient was not continuously enrolled in Medicare Parts A and B, or was enrolled in Part C, during the lookback plus episode window; no main clinician was attributed; the patient's date of birth is missing; the patient died before the episode ended; or the trigger claim was not filed from an ambulatory, office, outpatient hospital, or ASC setting. Measure-specific exclusions remove episodes involving endoscopic mucosal resection on the trigger date, inflammatory bowel disease or diverticular disease in the 120-day lookback, concurrent upper GI endoscopy on the trigger date, or an inpatient or emergency room trigger setting. Full exclusion code specifications are provided in the Exclusions and Exclusions_Details tabs of the codes list and in MIF Section 4.4.

Step 5: Risk adjustment and stratification. Risk adjustment is performed separately within each of the three care-setting sub-groups using an ordinary least squares (OLS) regression model with the standardized observed episode cost as the dependent variable. The model includes standard CMS HCC Version 24 variables (86 hierarchical condition category [HCC] codes mapped to thousands of International Classification of Diseases, 10th Revision, Clinical Modification [ICD-10-CM] diagnosis codes), interaction terms for comorbidity combinations, patient age category, disability status, end-stage renal disease (ESRD) status, and recent institutional long-term care use, plus eight measure-specific clinical comorbidity adjustors. Risk adjustors are defined from claims and administrative data in the 120-day period prior to the trigger date. After estimation, extreme low

expected costs are bottom-coded by assigning the value at the 0.5th percentile to all expected costs below that threshold, and the winsorized values are renormalized to the sub-group's average expected cost. Episodes with residuals below the 1st or above the 99th percentile of the residual distribution are then excluded as outliers, and expected costs are renormalized to the sub-group's average standardized observed cost after exclusions. Risk adjustment variable specifications are provided in the RA and RA_Details tabs of the codes list and in MIF Section 4.5.

Step 6: Calculate measure score. For each attributed episode, the ratio of standardized observed cost to risk-adjusted expected cost is calculated. These ratios are averaged across all episodes attributed to a clinician or group, combined across all three care-setting sub-groups. The average ratio is multiplied by the national average observed episode cost to generate a dollar figure representing risk-adjusted average episode cost. This dollar figure is the measure score. The numerator is the sum of observed-to-expected cost ratios across all attributed episodes; dividing by the denominator and multiplying by the national average observed episode cost yields the final measure score. A score below the national average indicates lower-than-expected costs; a score above indicates higher-than-expected costs. The measure score calculation is described in MIF Section 4.6.

The data collection time period is one MIPS performance year. The measure captures episodes ending within the calendar year performance period; because an episode may begin up to 14 days before its end date, episodes triggered in the final days of the prior calendar year may be included when their episode window ends within the performance year.

## 1.19 Measure Stratification Details

Every Screening/Surveillance Colonoscopy episode is placed into one of three mutually exclusive and exhaustive sub-groups at the time of episode construction, before exclusions and risk adjustment are applied. Sub-groups are designed to ensure clinical comparability by placing episodes into groups with a consistent care-setting cost structure, so that each episode's expected cost is estimated from episodes performed in the same setting. Because the resources required to perform a colonoscopy and the resulting episode costs differ systematically across care settings, grouping episodes by setting allows the risk adjustment model to capture setting-specific cost structure and ensures that each episode's expected cost is estimated from other episodes performed in the same care setting.

The stratification variable is the Place of Service (POS) code present on the Part B Physician/Supplier trigger claim at the time the colonoscopy was performed. The three sub-groups are: Ambulatory Surgery Center, Office, and Outpatient Department (which encompasses both on-campus and off-campus outpatient hospital settings). Each episode is assigned to exactly one sub-group based on this variable. Full POS code specifications, including codes and descriptors for each sub-group, are provided in the codes list (Sub_Groups and Sub_Groups_Details tabs).

Risk adjustment is performed separately within each of the three care-setting sub-groups. A separate ordinary least squares regression model is estimated within each sub-group, with payment-standardized episode cost as the dependent variable and patient-level risk adjustors as covariates. This approach allows the model to capture differences in expected cost structure across settings and ensures that expected costs are estimated from episodes performed in the

same care environment. Risk adjustors include CMS HCC Version 24 condition category indicators, standard beneficiary indicators (ESRD, long-term care institutional status, originally disabled), eight measure-specific clinical comorbidity indicators, HCC interaction terms, and a categorical age variable. Full risk adjustment variable specifications are provided in the codes list (RA and RA_Details tabs) and summarized in Section 5.4.

After risk adjustment is performed separately within each sub-group, each episode receives an expected cost value from its sub-group-specific model. The measure score for each clinician or group is the average observed-to-expected (O/E) ratio across all attributed episodes, combined across all three sub-groups. This O/E ratio is then multiplied by the national average observed episode cost to express the score as a dollar figure.

## 1.20 Types of Data Sources

Administrative Data, Claims Data

## 1.25 Data Source Details

Medicare Parts A and B claims data from the Common Working File (CWF) provide the primary source for identifying trigger events, assigned services, and episode costs. Standardized Medicare Parts A and B payment data from the IDR provide payment-standardized cost amounts, removing the effect of geographic payment adjustments and other payment differences to allow fair cost comparisons across providers and regions. The Common Medicare Environment (CME) provides beneficiary enrollment, demographic, and disability status data. The Long Term Care Minimum Data Set (LTC MDS) identifies beneficiaries with recent institutional long-term care stays, which is incorporated as a risk adjustment variable.

All data elements are contained in defined fields in electronic claims and administrative records. There is no additional data submission burden for clinicians beyond claims already submitted for Medicare reimbursement.

## 1.26 Minimum Sample Size

A minimum of 10 attributed episodes is required for a clinician or group to receive a measure score.

## 2.1 Attach Logic Model

3510-2.1-LogicModel-Spring2026.pdf

## 2.2 Evidence of Measure Importance

Colorectal cancer (CRC) is among the most prevalent and preventable cancers in the United States.[1] Screening colonoscopy is the most common CRC screening procedure performed among U.S. adults and represents a high-volume, high-spend area within the Medicare program.[2] In performance year (PY) 2024, 977,369 episodes were attributed to 3,205 groups and 15,017 clinicians, covering approximately 0.3 percent of total Medicare Part A and Part B expenditures.[3] The Medicare beneficiary population eligible for CRC screening is expected to

remain large, sustaining high colonoscopy volumes and the measure's ongoing relevance to program spending.

The desired health outcome for this measure is high-value colonoscopy: effective CRC screening delivered at appropriate cost. Several clinician-attributable factors contribute to episode cost variation, and the measure is designed to capture performance across these domains. Adequate bowel preparation is a critical quality determinant that affects the likelihood of a complete, effective colonoscopy; inadequate preparation increases procedural time, intraprocedural cleansing work, and the likelihood of an incomplete procedure that necessitates a repeat examination within the 14-day episode window, all of which add directly to episode costs.[4] Variation in anesthesiologist involvement during routine screening colonoscopy is an additional source of cost variation; for routine, average-risk patients, anesthesiologist-administered sedation adds meaningfully to procedure cost,[5] and ASGE sedation guidelines indicate that anesthesia provider-administered sedation does not appear to improve safety compared with endoscopist-directed sedation for low-risk patients.[6,7] Empirical validity testing using PY2024 data confirms that measure scores capture clinician-attributable variation in costs and clinical outcomes; full results are presented in Section 5.3.

Cost performance is a required component of the MIPS composite score under the Medicare Access and CHIP Reauthorization Act (MACRA). The Screening/Surveillance Colonoscopy episode-based cost measure was developed under MACRA section 101(f) based on input from the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee, which identified the measure's impact in terms of patient population size, clinician coverage, and the opportunity to incentivize cost-effective CRC screening care. The measure was first implemented in MIPS in PY2019.[3] Beginning with PY2025, the measure is also available within the Gastroenterology Care MIPS Value Pathway (MVP), providing gastroenterologists, nurse practitioners, and physician assistants a specialty-specific reporting framework with this measure as a core cost component.[8]

Footnotes:

[1] Siegel RL, Wagle NS, Star J, Kratzer TB, Smith RA, Jemal A. Colorectal cancer statistics, 2026. *CA Cancer J Clin*. 2026;76(2):e70067. PMID: 41769777.

[2] He Y, Xu T, Fang J, et al. Trends in colorectal cancer screening in the United States, 2012 to 2020. *J Med Screen*. 2023;30(3):125-133. PMID: 37157812.

[3] Centers for Medicare & Medicaid Services. 2026 MIPS Summary: Cost Measures. CMS; 2026. https://www.cms.gov/files/document/2026-mips-summary-cost-measures.pdf.

[4] Jacobson BC, Anderson JC, Burke CA, et al. Optimizing bowel preparation quality for colonoscopy: consensus recommendations by the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol*. 2025;120(4):738-764. PMID: 40035345.

[5] Krigel A, Chen L, Wright JD, Lebwohl B. Substantial increase in anesthesia assistance for outpatient colonoscopy and associated cost nationwide. *Clin Gastroenterol Hepatol*. 2019;17(12):2489-2496. PMID: 30625407.

[6] Early DS, Lightdale JR, Vargo JJ, et al. Guidelines for sedation and anesthesia in GI endoscopy. *Gastrointest Endosc*. 2018;87(2):327-337. PMID: 29306520.

[7] Khiani VS, Soulos P, Gancayco J, Gross CP. Anesthesiologist involvement in screening colonoscopy: temporal trends and cost implications in the Medicare population. *Clin Gastroenterol Hepatol*. 2012;10(1):58-64. PMID: 21782768.

[8] Centers for Medicare & Medicaid Services. Gastroenterology Care MIPS Value Pathway (M1422). Quality Payment Program. https://qpp.cms.gov/mips/explore-mips-value-pathways/2025/M1422.

## 2.4 Performance Gap

This measure observes a sustained performance gap across attributed clinicians and groups, reflecting variation in risk-adjusted episode costs after accounting for patient complexity and care setting. The score distribution presented below demonstrates that meaningful spread persists across the score range, indicating opportunity for cost-efficiency improvement remains.

In PY2024, 977,369 episodes from the Screening/Surveillance Colonoscopy episode group were attributed to clinicians and groups. Mean observed episode cost was $1,173 per episode, with a standard deviation of $389. Observed episode costs ranged from approximately $730 at the 10th percentile to $1,568 at the 90th percentile (a spread of approximately $838), and from $166 at the minimum to $2,943 at the maximum. The mean O/E ratio of 1.0 at the group level confirms that the risk adjustment model is well-calibrated in aggregate, while preserving meaningful variation at both the clinician and group levels.

At the 10-episode case minimum, 3,205 groups and 15,017 clinicians were eligible for performance scores. The full decile distribution of performance scores at both the group and clinician levels is presented in Table 1. Episode counts in Table 1 reflect episodes attributed to entities meeting the 10-episode case minimum (972,965 at the group level; 949,157 at the clinician level) and therefore differ from the total 977,369 attributed episode count above, which includes episodes at sub-minimum entities not eligible for scoring.

## Table 1. Performance Scores by Decile

**Mean Performance Score by Decile, Groups Meeting 10-Episode Case Minimum, 2024 Study Period**

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean Performance Score | 1.0 | 0.5 | 0.7 | 0.8 | 0.9 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.1 | 1.2 | 2.3 |
| Number of Entities | 3,205 | 1 | 320 | 321 | 320 | 321 | 320 | 321 | 321 | 320 | 321 | 320 | 1 |

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Persons / Encounters / Episodes | 972,965 | 10 | 17,016 | 24,645 | 42,222 | 69,965 | 89,317 | 138,359 | 171,857 | 205,262 | 146,084 | 68,251 | 80 |
| Mean Per Episode Cost | $1,174 | $529 | $842 | $969 | $1,036 | $1,083 | $1,117 | $1,149 | $1,178 | $1,212 | $1,254 | $1,341 | $2,682 |
| Mean Per Entity Cost | $1,123 | $529 | $856 | $962 | $1,037 | $1,082 | $1,116 | $1,147 | $1,178 | $1,211 | $1,255 | $1,382 | $2,682 |

**Mean Performance Score by Decile, Clinicians Meeting 10-Episode Case Minimum, 2024 Study Period**

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean Performance Score | 1.0 | 0.3 | 0.8 | 0.9 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.1 | 1.1 | 1.2 | 2.3 |
| Number of Entities | 15,017 | 1 | 1,501 | 1,502 | 1,502 | 1,502 | 1,501 | 1,502 | 1,502 | 1,502 | 1,502 | 1,501 | 1 |
| Number of Persons / Encounters / Episodes | 949,157 | 10 | 63,881 | 88,337 | 102,845 | 106,408 | 106,066 | 104,507 | 104,794 | 96,380 | 92,246 | 83,726 | 25 |
| Mean Per Episode Cost | $1,175 | $381 | $917 | $1,038 | $1,086 | $1,121 | $1,151 | $1,182 | $1,215 | $1,255 | $1,307 | $1,428 | $2,711 |
| Mean Per Entity Cost | $1,170 | $381 | $910 | $1,036 | $1,085 | $1,121 | $1,151 | $1,182 | $1,215 | $1,255 | $1,308 | $1,433 | $2,711 |

## 2.6 Meaningfulness to Target Population

CMS and the measure developer identify two target populations for this measure: Medicare beneficiaries who undergo screening or surveillance colonoscopy, and the clinicians and groups whose cost performance is assessed.

CMS convened a Person and Family Committee (PFC) to gather input from Medicare patients and caregivers on the meaningfulness of episode-based cost measures developed under MACRA.[1] Fifteen patients and caregivers were engaged through a series of structured discussions, including eight individual in-depth interviews.[1] PFC findings were shared with the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee during measure development to ensure patient perspectives informed episode design. PFC members identified five criteria for a cost episode to be meaningful and valuable for patient decision-making: it should affect a large number of Medicare beneficiaries; have the potential to improve or maintain quality of life; be associated with related quality measures; provide information usable for health care decisions; and have potential to reduce unnecessary Medicare spending.[1] The Screening/Surveillance Colonoscopy measure satisfies all five criteria. PFC members further

noted that non-emergent procedures, such as screening colonoscopy, offer patients and caregivers the greatest opportunity to evaluate cost performance information before making care decisions, making this measure especially relevant to the patient population it serves.[1] PFC members indicated that cost information is most meaningful when presented alongside quality data, and acknowledged limited patient familiarity with Medicare costs beyond out-of-pocket exposure;[1] both considerations reinforce the importance of pairing this measure with quality information in reporting contexts. The measure's focus on cost efficiency addresses variation that affects both Medicare program expenditures and patient cost exposure.

Clinicians and groups are the directly assessed population and have demonstrated sustained engagement with the measure since its initial field test. The 2017 national field test remains the only formal field testing program conducted for this measure.[2] In fall 2017, CMS and Acumen conducted that national field test during a 35-day comment period; 6,739 groups and 19,085 clinicians received confidential performance reports.[2] Of those, 1,364 groups and 10,628 clinicians accessed their reports during the feedback period, and approximately 1,000 participants attended two National Provider Calls.[2] Written feedback submitted during the comment period included 219 survey responses and 53 comment letters; participants submitted approximately 120 questions during the provider calls.[2] Twenty-one stakeholders provided colonoscopy-specific written feedback, including the American College of Gastroenterology, the American Gastroenterological Association, and a joint comment from those organizations together with the American Society for Gastrointestinal Endoscopy.[2] Colonoscopy-specific feedback was shared with the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee prior to December 2017 measure refinement webinars and was used to inform measure specification changes.[2] The proportion of recipients who actively engaged with their reports indicates that clinicians found the performance feedback meaningful for understanding their resource use relative to peers.

The NQF Cost and Efficiency Standing Committee reviewed the measure during the Spring 2019 review cycle and recommended it for endorsement; the NQF Consensus Standards Approval Committee approved the measure for endorsement on November 12, 2019.[3] Specialty society stakeholders contributed public comments during NQF review, and the Committee identified an opportunity for improvement to decrease costs associated with screening/surveillance colonoscopy.[3] The measure's risk adjustment approach, which includes CMS HCC Version 24 variables and measure-specific comorbidity adjustors, accounts for clinical patient complexity raised during measure development. During the 2022 Wave 1 Reevaluation public comment period, stakeholders continued to demonstrate active engagement, recommending alignment with the updated USPSTF guidelines, raising questions about cost variation across care settings, calling for greater public availability of performance data, and reaffirming the importance of adequate risk adjustment.[4] This feedback reflects ongoing clinician and stakeholder investment in measure quality and has informed continued maintenance of the measure specification.

Footnotes:

[1] Acumen LLC. Person and Family Committee Guiding Principles: MACRA Episode-Based Cost Measures Clinical Subcommittees. Acumen, LLC; April 2019.

[2] Centers for Medicare & Medicaid Services / Acumen LLC. Field Testing Feedback Summary Report for Eight MACRA Episode-Based Cost Measures. CMS; June 2018.

[3] National Quality Forum, Cost and Efficiency Standing Committee. Cost and Efficiency, Spring 2019 Review Cycle: CDP Report. NQF; February 21, 2020.

[4] Centers for Medicare & Medicaid Services / Acumen LLC. MACRA Episode-Based Cost Measures: Comprehensive Reevaluation — Public Comment Summary Report. CMS; August 2022.

## 3.1 Contributions Towards Closing Care Gaps

This domain is optional for the Spring 2026 cycle.

## 4.1a Data Structure and Availability

All data elements required for the measure are sourced from existing CMS administrative claims and beneficiary infrastructure (see Field 1.25 for data source detail). All elements are stored as structured, machine-readable fields, including diagnosis codes (ICD-10-CM), procedure codes (CPT, HCPCS, ICD-10-PCS), place-of-service codes, dates of service, payment amounts, beneficiary enrollment and demographic indicators, and institutional long-term-care assessments. The measure does not draw on free-text or unstructured documentation, and no manual chart abstraction is required.

Two categories of missing data are addressed through measure-specific exclusions: episodes attributed to beneficiaries with missing dates of birth, and episodes attributed to beneficiaries without continuous Medicare Parts A and B enrollment across the 120-day lookback period (see Field 4.3). Other inputs (procedure codes on the trigger claim, payment amounts, and place of service) are populated by definition for any claim meeting the inclusion criteria.

Because the measure is built entirely from finalized Medicare administrative claims, susceptibility to inaccuracies is limited to the same coding and billing variation present across all CMS claims-based programs. CMS payment and claims-processing edits, the two-month run-out period before measure calculation (see Field 4.3), and payment standardization mitigate the most common sources of variation. Every measure input (trigger codes, exclusion codes, attribution logic, risk adjustment variables, service-assignment logic) is published in the MIF and Codes List, allowing any episode to be reproduced from source claims. Full data sourcing details are provided in MIF Section 1.6.

## 4.1b Implementation Costs and Burden

The Screening/Surveillance Colonoscopy episode-based cost measure imposes minimal reporting burden on clinicians and groups. All data used to calculate measure scores are sourced from CMS administrative systems that exist independently of this measure and are collected as part of standard Medicare payment operations. No additional data collection, reporting, or abstraction is required from clinicians or groups. Clinicians do not submit any measure-specific data; performance is computed entirely from existing administrative records by CMS.

Because clinicians and groups do not collect, abstract, or submit any measure-specific data, the measure imposes no measure-related data collection burden during patient encounters: there is

no added documentation step, no measure-driven prompt or query embedded in the clinician's workflow, no measure-specific input required during diagnostic decision-making, and no measure-related question posed to the patient. Patients are not asked to complete surveys or provide additional information for measure calculation. Any clinical or workflow effects of the measure that may arise from clinicians reviewing their performance feedback are addressed in Section 6 (Use & Usability), not from data collection.

CMS performs all data abstraction (none required from clinicians), all measure calculation, and all performance reporting; clinicians and groups receive measure scores and patient-level detail through the Quality Payment Program (QPP) performance feedback portal (see Field 4.1c) without any local implementation effort. The measure is calculated and reported at scale on existing CMS infrastructure with no clinician- or group-level implementation activity required.

Full data sourcing details are provided in the MIF (Section 1.6).

## 4.1c Confidentiality

CMS data governance frameworks apply to all data used in this measure. Performance scores reported to clinicians and groups represent risk-adjusted observed-to-expected cost ratios at the clinician or group level and do not contain patient-identifiable information.

MIPS performance feedback reports provided through the QPP portal do include beneficiary-level data. CMS provides attributed beneficiary lists and episode-level cost detail to clinicians, groups, and virtual groups who meet the case minimum for the Screening/Surveillance Colonoscopy episode-based cost measure, enabling review of individual attributed patient episodes and associated costs.[1] These patient-level reports contain Protected Health Information (PHI) and personally identifiable information (PII), including the Medicare Beneficiary Identifier (MBI), date of birth, and date of death.[1] Beneficiary-level data are disclosed only to the attributed clinician or group, not to the public.

All disclosures of beneficiary-level data through MIPS feedback are governed by the Privacy Act of 1974,[2] the Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security Rules,[3] and CMS data governance policies applicable to the Medicare program. Access to beneficiary-level QPP portal data is controlled through role-based access in the QPP system, in which a Security Official for each group, virtual group, or Alternative Payment Model entity approves user roles before performance feedback can be viewed.[4]

Footnotes:

[1] Centers for Medicare & Medicaid Services. 2024 MIPS Performance Feedback: Supplemental and Patient-Level Data Reports Guide. Quality Payment Program; 2024.

[2] Privacy Act of 1974, 5 USC §552a.

[3] Health Insurance Portability and Accountability Act of 1996, Public Law 104-191; 45 CFR Parts 160 and 164.

[4] Centers for Medicare & Medicaid Services. 2024 MIPS Performance Feedback FAQs. Quality

Payment Program; 2024.

## 4.3 Feasibility Informed Final Measure

Testing and operational use during initial measure development identified three areas where findings informed modifications to the final specification.

First, testing of claims data finalization timing revealed that Medicare claims can require an extended run-out period before all claims for a given performance period are finalized in the Common Working File. Because it is not practical to wait until all claims have finalized before calculating measure scores, a two-month run-out period following the end of the calendar year was established as the appropriate point at which to pull claims data for measure calculation. This reflects the trade-off between timely data access and the accuracy that comes with allowing additional claims to finalize.

Second, testing identified two categories of episodes with incomplete beneficiary data that required exclusion to ensure accurate and comparable scoring. Episodes are excluded when the beneficiary was not continuously enrolled in Medicare Parts A and B for the 120-day lookback period prior to the episode trigger date, because the risk adjustment model requires that period of claims history to accurately capture comorbidities. Episodes are also excluded when the beneficiary's date of birth cannot be located, because the risk adjustment model includes a categorical age variable that requires this information.

Third, testing found that episodes in which the beneficiary died before the episode end date exhibited materially different cost distributions from other episodes. To prevent this from distorting clinician and group scores, episodes where beneficiary death occurs before the end of the episode window are excluded from measure scoring.

No specification changes have been made to the measure since first implementation in MIPS (PY2019), and no further modifications to data collection procedures, exclusion logic, or sampling have been required. Operational use of the measure in MIPS confirms that the measure specification remains stable and that no new feasibility issues have been identified.[1]

Full specification details are provided in MIF Sections 3.0, 4.1, and 4.2 and in the accompanying Codes List.

Footnotes:

[1] Centers for Medicare & Medicaid Services. 2026 MIPS Summary: Cost Measures. CMS; 2026. https://www.cms.gov/files/document/2026-mips-summary-cost-measures.pdf.

## 4.4 Proprietary Information

Not a proprietary measure and no proprietary components

## 5.1.1 Data Used for Testing

The measure uses Medicare Parts A and B claims data and beneficiary administrative data sources

for all measure testing (see Field 1.25 for data source detail). All cost figures refer to standardized Medicare allowed amounts. The testing dataset comprised final, deduplicated episodes, with outliers excluded unless otherwise noted. Full data source specifications are provided in MIF Section 1.6.

### 5.1.1a Dates of Testing Data

Episodes ending January 1, 2024 through December 31, 2024 (Performance Year 2024). Because the episode window is 14 days post-trigger, trigger events may occur as early as mid-December 2023 for episodes ending in early January 2024.

The 120-day lookback period for risk adjustment and exclusion checks extends back approximately 4 months prior to each episode's trigger date.

### 5.1.2 Differences in Data

None. The same PY2024 Medicare fee-for-service Parts A and B claims dataset was used for all measure testing, including reliability, validity, and risk adjustment analyses. No separate or supplemental data sources were used for any individual testing component.

### 5.1.3 Characteristics of Measured Entities

Measured entities are clinicians and groups that performed screening or surveillance colonoscopy procedures as the main performing clinician during the performance period. The measure is implemented nationally across all Medicare fee-for-service service areas; attribution is based on billing rather than geographic sampling, and no geographic restriction was applied. In PY2024, 3,205 groups and 15,017 clinicians met the 10-episode case minimum and received a measure score, with 972,965 episodes attributed at the group level and 949,157 at the clinician level.

### 5.1.4 Characteristics of Units of the Eligible Population

The testing population consists of Medicare fee-for-service beneficiaries enrolled in Medicare Parts A and B who underwent a screening or surveillance colonoscopy procedure during PY2024 and whose episode was attributed to a MIPS-eligible clinician or group. There is no age-based exclusion; the population is predominantly aged 65 and older but includes beneficiaries who became Medicare-eligible through disability or end-stage renal disease. Standard enrollment-based exclusions remove beneficiaries enrolled in Medicare Advantage and those without continuous Parts A and B enrollment across the episode window and 120-day lookback period.

The testing dataset included 977,369 final attributed episodes. Mean observed episode cost was $1,173 per episode (standard deviation $389). The episode cost distribution ranged from $166 at the minimum to $2,943 at the maximum, with a median of $1,118 and an interquartile range of $918 to $1,454.

The episode population was concentrated at ages 65 to 74, consistent with the primary screening-

eligible age range under current colorectal cancer screening guidelines. Beneficiaries under age 65 accounted for 47,199 episodes (4.8%), reflecting Medicare enrollment through disability or end-stage renal disease. The age distribution across final episodes was: ages 65 to 69, 339,011 episodes (34.7%); ages 70 to 74, 350,828 (35.9%); ages 75 to 79, 189,320 (19.4%); ages 80 to 84, 45,153 (4.6%); and ages 85 and older, 5,858 (0.6%).

The episode population reflects Medicare beneficiaries undergoing routine colorectal cancer screening or surveillance; beneficiaries with inflammatory bowel disease, Crohn's disease, ulcerative colitis, microscopic colitis, or diverticular disease are excluded. Among the final attributed population, notable prevalent comorbid conditions include anticoagulation use (9.5%), asthma or obstructive sleep apnea (13.0%), diabetes with chronic complications (11.7%), and specified heart arrhythmias (9.5%), consistent with the comorbidity burden typical of an older Medicare population.

Episodes were distributed across three care-setting sub-groups: ambulatory surgery center (ASC), 508,891 episodes (52.1%); hospital outpatient department (HOPD), 436,112 (44.6%); and office, 32,366 (3.3%).

### 5.2.1 Level(s) of Reliability Testing Conducted

Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

### 5.2.2 Method(s) of Reliability Testing

Reliability was estimated using the signal-to-noise method, which quantifies how much of the total observed variation in entity-level performance scores is attributable to systematic differences in true performance between entities (the signal) versus random variation or measurement error (the noise). A reliability value of 1.0 indicates that all observed variation reflects true performance differences; a value of 0.0 indicates that all observed variation is due to random error.

This method is consistent with CMS reliability methodology for episode-based cost measures with continuous performance scores. The performance score for this measure is the risk-adjusted average episode cost, which is a continuous measure. The signal-to-noise estimate is computed at both the group and clinician attribution levels.

Entities must have at least 10 attributed episodes to receive a performance score and be included in reliability reporting.

### 5.2.3 Reliability Testing Results

Signal-to-noise reliability statistics for groups and clinicians meeting the 10-episode case minimum are presented in the table below. Full decile-stratified reliability distributions are presented in Tables 2a and 2b.

**Signal-to-Noise Reliability Statistics, Groups and Clinicians Meeting 10 Episode Case Minimum, 2024 Study Period**

| Statistic | Group | Clinician |
|---|---|---|
| # of Entities | 3,205 | 15,017 |
| Reliability Mean | 0.946 | 0.909 |
| Reliability Std Dev | 0.056 | 0.065 |
| Reliability Minimum | 0.788 | 0.749 |
| Reliability 5th percentile | 0.817 | 0.767 |
| Reliability 10th percentile | 0.856 | 0.807 |
| Reliability 25th percentile | 0.918 | 0.868 |
| Reliability 50th percentile (median) | 0.969 | 0.929 |
| Reliability 75th percentile | 0.990 | 0.962 |
| Reliability 90th percentile | 0.996 | 0.977 |
| Reliability 95th percentile | 0.998 | 0.982 |
| Reliability Maximum | 0.999 | 0.995 |

### 5.2.4 Interpretation of Reliability Results

The signal-to-noise reliability results demonstrate that the 10-episode case minimum produces highly reliable performance scores for the Screening/Surveillance Colonoscopy measure. A signal-to-noise reliability score indicates the proportion of observed variation in entity-level performance scores that reflects true differences in performance, rather than random error; higher values indicate greater confidence that observed performance differences reflect actual practice patterns. This interpretation of reliability values as the proportion of observed variation reflecting true performance differences follows the reliability definition and interpretation guidance specified in the PQM Endorsement and Maintenance (E&M) Guidebook.[2]

At the group level, mean reliability is 0.946, indicating that 94.6 percent of the observed variation in group performance scores reflects true performance differences; at the clinician level, mean reliability is 0.909, indicating that 90.9 percent of the observed variation reflects true performance differences. The minimum observed reliability is 0.788 at the group level and 0.749 at the clinician level.

All groups and all clinicians exceed the 0.4 moderate reliability threshold established for MIPS cost measures in the CY2017 Quality Payment Program final rule (81 FR 77169–77170).[1] The measure also satisfies the CBE accountable entity-level threshold for signal-to-noise reliability, which requires that fewer than 30 percent of accountable entities fall below 0.6, as specified in the PQM E&M Guidebook.[2] All groups and clinicians fall well above the 0.6 threshold (group minimum 0.788; clinician minimum 0.749). The consistently high reliability at both attribution levels and the relatively narrow spread of reliability values confirm that the 10-episode case minimum is appropriately calibrated for this measure.

Footnotes:

[1] Centers for Medicare & Medicaid Services, HHS. Medicare program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models. Final rule with comment period. *Fed Regist*. 2016;81(214):77008-77831.

[2] Battelle. *Endorsement and Maintenance (E&M) Guidebook*. Version 3.3. Partnership for Quality Measurement; March 2026.

## Table 2a. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

### Reliability by Denominator Decile, Groups Meeting 10-Episode Case Minimum, 2024 Study Period

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reliability | 0.95 | 0.79 | 0.82 | 0.88 | 0.92 | 0.94 | 0.96 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 |
| Mean Performance Score | 0.96 | 0.89 | 0.90 | 0.93 | 0.92 | 0.94 | 0.95 | 0.96 | 0.98 | 0.98 | 0.99 | 1.01 | 1.03 |
| Number of Entities | 3,205 | 61 | 312 | 314 | 327 | 328 | 323 | 321 | 319 | 320 | 320 | 321 | 1 |
| Number of Persons / Encounters / Episodes | 972,965 | 10 | 3,838 | 6,250 | 9,868 | 14,783 | 22,395 | 33,797 | 51,236 | 86,454 | 168,441 | 575,916 | 28,484 |
| Mean Per Episode Cost | $1,174 | $1,049 | $1,059 | $1,089 | $1,075 | $1,107 | $1,119 | $1,130 | $1,148 | $1,154 | $1,162 | $1,193 | $1,213 |
| Mean Per Entity Cost | $1,123 | $1,049 | $1,058 | $1,089 | $1,074 | $1,105 | $1,117 | $1,131 | $1,149 | $1,155 | $1,161 | $1,188 | $1,213 |

### Reliability by Denominator Decile, Clinicians Meeting 10-Episode Case Minimum, 2024 Study Period

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reliability | 0.91 | 0.75 | 0.77 | 0.83 | 0.87 | 0.90 | 0.92 | 0.94 | 0.95 | 0.96 | 0.97 | 0.98 | 1.00 |
| Mean Performance Score | 1.00 | 0.97 | 0.98 | 0.99 | 0.99 | 0.99 | 1.00 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.03 |
| Number of Entities | 15,017 | 380 | 1,439 | 1,468 | 1,619 | 1,456 | 1,486 | 1,588 | 1,465 | 1,496 | 1,500 | 1,500 | 1 |
| Number of Persons / Encounters / Episodes | 949,157 | 10 | 16,410 | 23,382 | 35,405 | 42,642 | 56,894 | 79,798 | 96,045 | 128,014 | 175,902 | 294,694 | 714 |
| Mean Per Episode Cost | $1,175 | $1,137 | $1,146 | $1,160 | $1,158 | $1,163 | $1,173 | $1,181 | $1,177 | $1,184 | $1,181 | $1,172 | $1,212 |
| Mean Per Entity Cost | $1,170 | $1,137 | $1,145 | $1,160 | $1,157 | $1,163 | $1,173 | $1,181 | $1,177 | $1,184 | $1,181 | $1,174 | $1,212 |

## Table 2b. Accountable Entity Level Reliability Testing Results by Reliability Score

### Reliability by Decile, Groups Meeting 10-Episode Case Minimum, 2024 Study Period

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reliability | 0.95 | 0.79 | 0.82 | 0.88 | 0.92 | 0.94 | 0.96 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 |

**Reliability by Decile, Clinicians Meeting 10-Episode Case Minimum, 2024 Study Period**

| | Overall | Min | Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reliability | 0.91 | 0.75 | 0.77 | 0.83 | 0.87 | 0.90 | 0.92 | 0.94 | 0.95 | 0.96 | 0.97 | 0.98 | 1.00 |

### 5.3.1 Level(s) of Validity Testing Conducted

Accountable entity level (i.e., measure score) (e.g., criterion validity)

### 5.3.2 Type of Accountable Entity Level Validity Testing Conducted

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis)

### 5.3.3 Method(s) of Validity Testing

This analysis assesses whether the Screening/Surveillance Colonoscopy measure differentiates among clinicians and groups in a clinically meaningful way by examining the relationship between group and clinician resource use and outcome event patterns and entity performance scores.
For each event category, the proportion of attributed episodes containing the event was calculated for both groups and clinicians. Spearman rank correlation was then used to assess the relationship between each entity's event share and mean observed-to-expected (O/E) ratio. The Spearman rank method is used to account for non-normal distributions in the event share variable.
Results are reported at the 10-episode case minimum used for measure reporting, at both group and clinician attribution levels.
Entities accountable for a greater share of episodes with a given event category, particularly outcome events such as repeat colonoscopy and adverse events, should have higher O/E ratios, confirming that the measure captures clinician-attributable resource use and outcome patterns.
Two primary validity events were evaluated:

- Repeat Colonoscopy/Flexible Sigmoidoscopy (Resource Use): A repeat colonoscopy or flexible sigmoidoscopy within the episode window is a direct indicator of potential over-utilization; reduction through adherence to evidence-based screening intervals is an actionable improvement lever.
- Combined Adverse Events (Outcome): Complications, emergency department (ED) visits, inpatient (IP) admissions, and post-acute care collectively reflect procedural safety and post-procedure care quality.

### 5.3.4 Validity Testing Results

Validity testing demonstrated positive, statistically significant correlations between entity performance scores and rates of the two primary validity events, repeat colonoscopy/flexible sigmoidoscopy and combined adverse events, at both the group and clinician levels. At the group level (3,205 entities), repeat colonoscopy correlated with O/E ratios at r=0.12 (p<0.001) and combined adverse events at r=0.11 (p<0.001), indicating that groups with higher rates of these events tend to have higher-than-expected costs. At the clinician level (15,017 entities), correlations were positive and statistically significant but smaller in magnitude, consistent with the expectation that individual-level estimates are noisier given smaller episode volumes per entity. Several individual adverse event categories also showed significant positive correlations at the group level, including cardiopulmonary complications (r=0.15), perforation/peritonitis (r=0.11), procedural complications (r=0.10), and ED visits (r=0.10). Inpatient admission and post-acute care events were not computable at the reporting case minimum because no entities had attributed episodes with these events.

Full correlation results across all event categories are presented in the first table below; primary validity event results are summarized in the second table.

**Spearman Correlation Results by Event Category and Attribution Level, Groups and Clinicians Meeting 10 Episode Case Minimum, 2024 Study Period**

| Category | Event | Attribution Level | # of Entities | Entities w/ Event (%) | Mean Event Share | Spearman r | P-value |
|---|---|---|---|---|---|---|---|
| Outcome | Cardiopulmonary Complication | Group | 3,205 | 77.5% | 0.07 | 0.15 | p<0.001 |
| Outcome | Lower GI Hemorrhage | Group | 3,205 | 61.4% | 0.03 | 0.07 | p<0.001 |
| Outcome | Perforation/Peritonitis | Group | 3,205 | 11.4% | 0.01 | 0.11 | p<0.001 |
| Outcome | Procedural Complication | Group | 3,205 | 3.4% | 0.02 | 0.10 | p<0.001 |
| Outcome | ED Visits | Group | 3,205 | 17.3% | 0.01 | 0.10 | p<0.001 |
| Outcome | IP Admission | Group | 3,205 | 0.0% | — | — | — |
| Outcome | Post Acute Care | Group | 3,205 | 0.0% | — | — | — |
| Outcome | Complications (combined) | Group | 3,205 | 86.3% | 0.08 | 0.11 | p<0.001 |
| Outcome | Combined Adverse Events | Group | 3,205 | 86.6% | 0.08 | 0.11 | p<0.001 |
| Resource Use | Repeat Colonoscopy/Flexible Sigmoidoscopy | Group | 3,205 | 26.3% | 0.01 | 0.12 | p<0.001 |
| Outcome | Cardiopulmonary Complication | Clinician | 15,017 | 71.5% | 0.08 | 0.06 | p<0.001 |
| Outcome | Lower GI Hemorrhage | Clinician | 15,017 | 35.0% | 0.04 | 0.00 | p=0.918 |
| Outcome | Perforation/Peritonitis | Clinician | 15,017 | 3.0% | 0.03 | 0.01 | p=0.103 |

| Category | Event | Attribution Level | # of Entities | Entities w/ Event (%) | Mean Event Share | Spearman r | P-value |
|---|---|---|---|---|---|---|---|
| Outcome | Procedural Complication | Clinician | 15,017 | 1.2% | 0.04 | 0.02 | p=0.043 |
| Outcome | ED Visits | Clinician | 15,017 | 5.4% | 0.02 | 0.00 | p=0.665 |
| Outcome | IP Admission | Clinician | 15,017 | 0.0% | — | — | — |
| Outcome | Post Acute Care | Clinician | 15,017 | 0.0% | — | — | — |
| Outcome | Complications (combined) | Clinician | 15,017 | 79.4% | 0.08 | 0.05 | p<0.001 |
| Outcome | Combined Adverse Events | Clinician | 15,017 | 79.8% | 0.08 | 0.05 | p<0.001 |

*Notes: "Entities w/ Event (%)" is the percentage of all entities with at least one attributed episode with the event. "Mean Event Share" is the mean proportion of episodes with the event among entities with the event. Lower GI Hemorrhage (clinician level), Perforation/Peritonitis (clinician level), and ED Visits (clinician level) are not statistically significant at the 10-episode case minimum. IP Admission and Post Acute Care: all entities have zero events at the 10-episode case minimum; no correlation is computable.*

**Primary Validity Events: Spearman Correlation Summary, Groups and Clinicians Meeting 10 Episode Case Minimum, 2024 Study Period**

| Event | Attribution Level | # of Entities | Spearman r | P-value |
|---|---|---|---|---|
| Repeat Colonoscopy/Flexible Sigmoidoscopy | Group | 3,205 | 0.12 | p<0.001 |
| Repeat Colonoscopy/Flexible Sigmoidoscopy | Clinician | 15,017 | 0.03 | p<0.001 |
| Combined Adverse Events | Group | 3,205 | 0.11 | p<0.001 |
| Combined Adverse Events | Clinician | 15,017 | 0.05 | p<0.001 |

### 5.3.5 Interpretation of Validity Results

At the group level, a positive and statistically significant correlation was observed between the entity's share of episodes involving a repeat colonoscopy or flexible sigmoidoscopy within the episode window and the entity's O/E ratio (r=0.12, p<0.001). Groups with higher rates of repeat colonoscopy within the episode window, a pattern inconsistent with evidence-based screening intervals, tend to have higher O/E ratios. Reducing premature re-scoping is a direct, clinician-modifiable improvement lever captured by the measure.

Combined adverse events also showed a positive, statistically significant correlation at the group level (r=0.11, p<0.001). The positive direction confirms that groups with higher adverse event rates tend to have higher O/E ratios, supporting the measure's capacity to reflect procedural safety and post-procedure care quality.

Correlation magnitudes are consistent with the clinical context: screening colonoscopy is primarily a low-complication procedure for average-risk patients,[1] and the events evaluated are rare. Multi-factorial cost variation means that no single event category is expected to dominate the O/E ratio signal. Statistically significant correlations in the hypothesized direction across two independent improvement domains constitute meaningful validity evidence under these conditions.

Clinician-level results are generally consistent with the group-level pattern, showing positive correlations in the same direction for the primary validity events. Some individual adverse event categories do not reach statistical significance at the clinician level given the smaller number of episodes per attributed entity at the 10-episode case minimum; the full clinician-level results are presented in the table above.

Overall, the pattern of positive, statistically significant correlations for Repeat Colonoscopy/Flexible Sigmoidoscopy and Combined Adverse Events at the group level supports the construct validity of the measure: variation in entity performance scores is systematically associated with variation in event patterns attributable to clinician practice in the hypothesized direction.

Footnotes:

[1] Kothari ST, Huang RJ, Shaukat A, et al. ASGE review of adverse events in colonoscopy. *Gastrointest Endosc*. 2019;90(6):863-876. PMID: 31563271.

### 5.4.1 Methods Used to Address Risk Factors

Statistical risk adjustment model with risk factors

### 5.4.2 Conceptual Model Rationale

Risk adjustment is applied to the Screening/Surveillance Colonoscopy measure to isolate variation in episode costs attributable to clinician practice patterns from variation driven by patient clinical factors outside the clinician's reasonable influence. The model estimates expected episode costs given patient characteristics, so that the measure score reflects only the clinician-modifiable component of cost variation.

The 111 risk-adjustment variables span six categories: 86 hierarchical condition category indicators from CMS HCC Version 24, 12 standard interaction terms, 3 standard beneficiary indicators (ESRD, long-term care institutional status, originally disabled), 8 measure-specific clinical comorbidity indicators, 1 patient age categorical variable, and 1 episode-group BY variable. Variables were selected based on three criteria: a direct clinical relationship to colonoscopy episode cost; availability from pre-episode claims data, ensuring that included factors reflect patient characteristics rather than the care provided; and an independent contribution to cost variation beyond what other model variables already capture. The Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee, the expert clinician committee that recommended this measure for development, provided detailed input on the selection of measure-specific risk variables.

The model includes the full set of hierarchical condition category indicators from CMS HCC Version 24, spanning major clinical domains including malignancies, metabolic and endocrine conditions, hepatic and gastrointestinal disorders, musculoskeletal and rheumatologic conditions, neuropsychiatric conditions, cardiopulmonary conditions, renal conditions, and injury and trauma categories. These indicators map patient diagnoses from the prior year to condition categories with similar cost implications and were developed specifically for use with Medicare fee-for-service claims data. Twelve standard interaction terms from the CMS HCC Version 24 structure are also included, capturing cases where the co-occurrence of multiple conditions increases expected costs beyond what individual indicators predict.

Three standard beneficiary indicators are included: end-stage renal disease (ESRD), long-term care institutional status (LTI), and originally disabled status. These capture structural differences in Medicare enrollment and care setting associated with elevated episode costs independent of the colonoscopy procedure itself: patients with ESRD, long-term care institutional status, or disability-based Medicare eligibility carry higher comorbidity burdens and care coordination requirements that increase expected costs regardless of clinician practice patterns.

Eight measure-specific indicators capture patient comorbidities that increase colonoscopy episode cost in ways outside the attributed clinician's control: automated implantable cardioverter-defibrillator (AICD) status, use of anticoagulation therapy, asthma or obstructive sleep apnea, history of anesthesia difficulties, valvular heart disease or hypertrophic obstructive cardiomyopathy, home oxygen use or respiratory failure, poorly controlled hypertension, and pulmonary hypertension. The Clinical Subcommittee identified these factors based on their clinical relationship to procedural complexity, anesthesia requirements, and complication risk. Anesthesiologist involvement is a significant and variable cost driver in colonoscopy episodes, and comorbidities that increase anesthesia complexity represent a primary source of patient-attributable cost variation in this episode group.[1]

Patient age is incorporated as a categorical variable, with the 65–69 age group as the reference category, reflecting the increase in episode costs with patient age due to greater comorbidity burden. Separate risk-adjustment models are estimated within each of the three care-setting sub-groups (ASC, office, HOPD) because episode cost distributions differ substantially across settings, reflecting structural differences in overhead, case mix, and payment rates that are not attributable to clinician practice.[2] Site-of-care selection is among the principal drivers of cost variation in colonoscopy episodes; stratified modeling ensures that each episode is evaluated against episodes from the same care setting.

Social and environmental factors were considered for inclusion in the risk-adjustment model. The model as specified does not include indicators for individual-level social risk or community-level socioeconomic characteristics, for the following reasons.

First, the episode definition includes only costs clinically related to the colonoscopy procedure within the 14-day window. Pathways through which social and environmental factors are most strongly associated with elevated healthcare costs, including emergency department utilization, readmissions for unrelated conditions, and chronic disease complications, fall outside this episode scope. Second, the CMS HCC Version 24 risk model and the measure-specific clinical adjustors capture patient clinical complexity that is substantially collinear with social risk markers, limiting the marginal predictive contribution of adding social risk variables to the model. Third, patient

selection for elective screening procedures, in combination with the clinical exclusion criteria applied to this episode group, reduces the representation of the highest-social-risk beneficiaries in the attributed episode population.

This conceptual rationale is consistent with prior research on MIPS episode-based cost measures, which found that adjustment for social risk factors does not meaningfully change clinician or group performance on these measures.[3]

Full code-level specifications for all risk adjustment variables, including code types and applicable lookback periods, are provided in the attached codes list (RA and RA_Details tabs).

Footnotes:

[1] Khiani VS, Soulos P, Gancayco J, Gross CP. Anesthesiologist involvement in screening colonoscopy: temporal trends and cost implications in the Medicare population. *Clin Gastroenterol Hepatol*. 2012;10(1):58-64. PMID: 21782768.

[2] Wang Y, Wang Y, Plummer E, Chernew ME, Anderson G, Bai G. Facility fees for colonoscopy procedures at hospitals and ambulatory surgery centers. *JAMA Health Forum*. 2023;4(12):e234025. PMID: 38100094.

[3] Sandhu AT, Bhattacharya J, Lam J, et al. Adjustment for social risk factors does not meaningfully affect performance on Medicare's MIPS clinician cost measures. *Health Aff (Millwood)*. 2020;39(9):1495-1503. PMID: 32897780.

### 5.4.2a Attach Conceptual Model

3510-5.4.2a-ConceptualModel-Spring2026.pdf

### 5.4.3 Variable Distribution Across Measured Entities

Episode-level prevalence and clinician-level distribution of each measure-specific risk adjustment variable in the PY2024 final attributed episode population are presented in the two tables below. Episode-level prevalence (first table) reports the population-aggregate exposure to each variable in the 977,369 final attributed episodes. Clinician-level distribution (second table) reports the share of attributed clinicians exposed to each variable, and the mean count of variable-bearing episodes per exposed clinician, at two testing thresholds: a 10-episode case minimum (15,017 clinicians) and a 20-episode case minimum (11,854 clinicians). Variables in both tables are ordered by episode-level prevalence.

**Episode-Level Prevalence of Measure-Specific Risk Adjustment Variables, Final Attributed Episode Population, 2024 Study Period**

| Measure-Specific Variable | Episodes | % of Final Episodes |
|---|---|---|
| Asthma or obstructive sleep apnea | 127,016 | 13.0% |
| Use of anticoagulation therapy | 92,849 | 9.5% |
| Valvular heart disease or hypertrophic obstructive cardiomyopathy | 57,556 | 5.9% |

| Measure-Specific Variable | Episodes | % of Final Episodes |
|---|---|---|
| Automated implantable cardioverter-defibrillator (AICD) status | 15,255 | 1.6% |
| Pulmonary hypertension | 8,595 | 0.9% |
| Home oxygen use or respiratory failure | 8,510 | 0.9% |
| History of anesthesia difficulties | 203 | 0.02% |
| Poorly controlled hypertension | 120 | 0.01% |

Clinician-level exposure to each risk factor varies markedly across attributed clinicians. Common comorbidities (asthma or obstructive sleep apnea, use of anticoagulation therapy, valvular heart disease or hypertrophic obstructive cardiomyopathy) are present in the final attributed episodes of 8.8% to 30.5% of clinicians meeting the 10-episode case minimum, and 1.3% to 11.9% of clinicians meeting the 20-episode case minimum. Rarer comorbidities (AICD status, pulmonary hypertension, home oxygen use or respiratory failure) are present at fewer than 0.2% of clinicians at the 10-episode case minimum, with most clinicians at this threshold treating zero such patients. History of anesthesia difficulties and poorly controlled hypertension are sufficiently rare that no clinicians reach the 10-episode testing threshold for these variables. Mean episode count among exposed clinicians at the 10-episode case minimum ranges from 11.4 to 17.9 variable-bearing episodes per clinician, demonstrating that even within the exposed subset, the volume of variable-bearing episodes varies meaningfully across clinicians. This between-entity heterogeneity in risk-factor exposure, in both share of clinicians exposed and patient-volume per exposed clinician, is the empirical basis for risk adjustment: without adjustment, clinicians whose attributed episodes contain disproportionate shares of higher-cost comorbidities would be penalized in performance comparisons for case-mix differences outside their control.

**Distribution of Measure-Specific Risk Adjustment Variables, Clinicians Meeting 10-Episode and 20-Episode Case Minimums, 2024 Study Period**

| Measure-Specific Variable | # of Clinicians with Variable (10-Episode Case Minimum) | % of Clinicians (10-Episode Case Minimum) | Mean Episodes per Clinician (10-Episode Case Minimum) | # of Clinicians with Variable (20-Episode Case Minimum) | % of Clinicians (20-Episode Case Minimum) | Mean Episodes per Clinician (20-Episode Case Minimum) |
|---|---|---|---|---|---|---|
| Asthma or obstructive sleep apnea | 4,579 | 30.5% | 17.9 | 1,405 | 11.9% | 27.7 |
| Use of anticoagulation therapy | 3,017 | 20.1% | 15.9 | 607 | 5.1% | 26.1 |
| Valvular heart disease or hypertrophic obstructive cardiomyopathy | 1,317 | 8.8% | 14.5 | 160 | 1.3% | 25.9 |
| Automated implantable cardioverter-defibrillator (AICD) status | 26 | 0.17% | 12.9 | 2 | 0.02% | 22.0 |
| Pulmonary hypertension | 10 | 0.07% | 11.4 | 0 | 0% | — |
| Home oxygen use or respiratory failure | 2 | 0.01% | 12.0 | 0 | 0% | — |

| | | | | | | |
|---|---|---|---|---|---|---|
| History of anesthesia difficulties | 0 | 0% | — | 0 | 0% | — |
| Poorly controlled hypertension | 0 | 0% | — | 0 | 0% | — |

Code-level specifications for all 111 risk adjustment variables, including code types and lookback periods, are provided in the attached Measure Codes List (RA and RA_Details tabs). Descriptive statistics for the full set of 111 risk-adjustment variables (including HCC indicators, interaction terms, beneficiary indicators, and patient age categories), with both episode-level prevalence and clinician-level distribution at the 10-episode and 20-episode case minimums, are provided in attachment Section 5.4.3a.

### 5.4.3a Attach Descriptive Statistics for Risk/Case-mix Variables

3510-5.4.3a-RiskFactors-Spring2026.pdf

### 5.4.4 Risk/Case-Mix Adjustment Modeling and/or Stratification Results

Risk adjustment follows the standard CMS episode-based cost measure approach. Patient-level HCC and measure-specific risk adjustors are defined from Medicare claims in the 120-day lookback period prior to the episode trigger date. A separate ordinary least squares (OLS) regression model is run nationally within each care-setting sub-group, using all risk adjustment variables as covariates and standardized observed episode cost as the dependent variable, to estimate the risk-adjusted expected episode cost. Expected costs below the 0.5th percentile are bottom-coded to that threshold, then renormalized so the mean equals the sub-group's average observed cost. Episodes with residuals outside the 1st–99th percentile range are excluded as outliers, and expected costs are renormalized again to match the sub-group's average standardized observed cost post-exclusion.

Mean observed episode cost and mean expected episode cost are both $1,173 per episode across all attributed episodes, confirming that the risk adjustment model is correctly calibrated at the mean. Mean risk-adjusted episode cost, computed as each episode's O/E ratio multiplied by the national average observed cost, is $1,174 per episode; this differs from mean expected cost because the mean of episode-level O/E ratios is 1.00 with small distributional asymmetry that elevates the risk-adjusted mean by approximately one dollar after rounding.

$R^2$ for the within-sub-group OLS regressions is 0.006 in ASC, 0.011 in HOPD, and 0.011 in Office (model fit on pre-residual-trim episode counts of 519,275, 445,011, and 33,026, respectively); $R^2$ for the combined episode group is 0.43, reflecting between-setting cost variation absorbed when sub-groups are pooled. Low within-sub-group $R^2$ is consistent with the measure design: the upstream service assignment rules remove most cost variation unrelated to the colonoscopy procedure before the regression is fit, leaving the regression to address narrower clinical variation among already-similar episodes within each setting; calibration evidence in Section 5.4.5 is the appropriate basis for evaluating model performance.

The table below presents the mean observed and mean risk-adjusted episode cost for each of the 8 measure-specific clinical adjustors, alongside a summary of statistical significance from the within-sub-group regressions. Each variable is associated with elevated observed episode costs

relative to the overall population mean; the risk adjustment model accounts for these cost differentials so that clinicians treating patients with these comorbidities are not systematically disadvantaged in performance comparisons. All 8 measure-specific adjustors were recommended by the Clinical Subcommittee for inclusion in the risk adjustment model based on their clinical relevance to colonoscopy episode cost (Section 5.4.2). Among the 8 adjustors, home oxygen use or respiratory failure was statistically significant in all three sub-groups; use of anticoagulation therapy, asthma or obstructive sleep apnea, and valvular heart disease or hypertrophic obstructive cardiomyopathy were statistically significant in ASC and HOPD but not in the smaller Office sub-group; and automated implantable cardioverter-defibrillator (AICD) status, pulmonary hypertension, history of anesthesia difficulties, and poorly controlled hypertension did not reach statistical significance in any sub-group. Episode counts for the four adjustors that did not reach statistical significance are small in some sub-groups, limiting statistical power to detect effects. The codes list (RA and RA_Details tabs) lists all 111 risk adjustment variables and their definitions.

**Mean Observed and Risk-Adjusted Episode Cost by Measure-Specific Risk Adjustor, 2024 Study Period**

| Measure-Specific Adjustor | Observed Mean Cost | Risk-Adjusted Mean Cost | Significant Sub-Groups |
|---|---|---|---|
| Overall attributed population | $1,173 | $1,174 | — |
| Home oxygen use or respiratory failure | $1,359 | $1,137 | ASC, HOPD, Office |
| Automated implantable cardioverter-defibrillator (AICD) status | $1,303 | $1,168 | None |
| Pulmonary hypertension | $1,282 | $1,146 | None |
| Poorly controlled hypertension | $1,263 | $1,194 | None |
| Use of anticoagulation therapy | $1,258 | $1,161 | ASC, HOPD |
| History of anesthesia difficulties | $1,243 | $1,203 | None |
| Asthma or obstructive sleep apnea | $1,220 | $1,170 | ASC, HOPD |
| Valvular heart disease or hypertrophic obstructive cardiomyopathy | $1,197 | $1,167 | ASC, HOPD |

The model was estimated separately within each of the three care-setting sub-groups. Observed mean episode cost differed substantially across settings: $976 for ASC episodes (508,891 episodes, 52.1% of the final attributed population), $1,446 for hospital outpatient department episodes (436,112 episodes, 44.6%), and $606 for office episodes (32,366 episodes, 3.3%). These structural cost differences reflect differences in facility payments, overhead, and case mix across settings rather than differences attributable to clinician practice patterns. Stratification neutralizes these site-of-service cost differences: each episode's expected cost is estimated from an OLS model fit to episodes performed in the same setting, so episodes are directly compared only against other episodes with procedures performed in the same setting. After separate risk adjustment within each sub-group and renormalization to the national average observed cost of $1,173 per episode, the mean risk-adjusted episode cost is equivalent across all three sub-groups.

## 5.4.4a Attach Risk/Case-mix Adjustment Modeling and/or Stratification

### Specifications

3510-5.4.4a-RiskAdjustmentModel-Spring2026.pdf

## 5.4.5 Calibration and Discrimination

The O/E ratio distribution is well-centered, with a mean of 1.00 and a median of 1.00, confirming no systematic over- or under-adjustment across the attributed population. The interquartile range of approximately 0.79 to 1.09 reflects meaningful variation in entity efficiency that persists after risk adjustment. The distribution is moderately right-skewed, with the upper tail reflecting a relatively small number of high-cost entities.

Predictive ratios by risk decile demonstrate that the risk adjustment model is well calibrated across the full range of patient clinical complexity. Episodes were divided into deciles by expected cost; mean observed and mean expected costs were calculated within each decile, along with the mean ratio of observed to expected cost (mean O/E) and the ratio of mean expected to mean observed cost (mean E / mean O). Across all ten deciles, mean O/E ranges from 0.99 to 1.00, and mean E / mean O ranges from 1.00 to 1.01. The narrow spread of these ratios indicates that the model neither systematically under-predicts costs for low-complexity patients nor over-predicts costs for high-complexity patients; risk adjustment performs consistently across the attributed episode population.

**Predictive Ratios by Risk Decile, Final Attributed Episode Population, 2024 Study Period**

| Risk Decile | Mean Observed Cost | Mean Expected Cost | Mean E / Mean O | Mean (O/E) |
|---|---|---|---|---|
| All | $1,173 | $1,173 | 1.00 | 1.00 |
| 1 (lowest) | $884 | $882 | 1.00 | 1.00 |
| 2 | $965 | $962 | 1.00 | 1.00 |
| 3 | $969 | $966 | 1.00 | 1.00 |
| 4 | $979 | $976 | 1.00 | 1.00 |
| 5 | $993 | $993 | 1.00 | 1.00 |
| 6 | $1,265 | $1,269 | 1.00 | 1.00 |
| 7 | $1,420 | $1,420 | 1.00 | 1.00 |
| 8 | $1,437 | $1,432 | 1.00 | 1.00 |
| 9 | $1,455 | $1,452 | 1.00 | 1.00 |
| 10 (highest) | $1,504 | $1,515 | 1.01 | 0.99 |

## 5.4.5a Attach Calibration and Discrimination Testing Results

3510-5.4.5a-CalibrationDiscriminationTestingResults-Spring2026.pdf

## 5.4.6 Interpretation of Risk/Case-mix Factor Findings

The risk adjustment model is correctly calibrated at the mean: mean observed and expected episode costs are equal ($1,173) for the full attributed episode population, confirming that the renormalization step is functioning as intended. The O/E ratio distribution is centered at 1.00, indicating no systematic over- or under-adjustment across the attributed population. Predictive

ratios across risk deciles (Section 5.4.5) confirm that calibration holds consistently across the full range of patient complexity. The O/E ratio distribution shows meaningful spread, indicating that the measure discriminates among clinicians and groups and that clinician-attributable variation remains after risk adjustment. The risk adjustment variable set is comprehensive, including HCC indicators spanning major clinical domains, standard beneficiary indicators for ESRD, long-term care status, and disability, and eight colonoscopy-specific clinical adjustors capturing patient comorbidities that increase procedural complexity, anesthesia requirements, and complication risk independent of clinician efficiency.

### 5.4.7 Final Approach to Address Risk Factors

Statistical risk adjustment model with risk factors

### 6.1.1 Current Status

In use

### 6.1.2 Current or Planned Use(s)

Public Reporting, Payment Program, Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

### 6.1.3 Program Details

Name of the program and sponsor
Quality Payment Program (QPP) Merit-based Incentive Payment System (MIPS)
URL of the program
https://qpp.cms.gov/mips
Purpose of the program

The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) established the Quality Payment Program. Under QPP, clinicians are incentivized to provide high-quality and high-value care through Advanced Alternative Payment Models (APMs) or MIPS. MIPS eligible clinicians receive a performance-based payment adjustment to their Medicare payment based on a final score assessing evidence-based and practice-specific data across four performance categories: quality, improvement activities, Promoting Interoperability, and cost. This measure is used in the cost performance category.

Geographic area and percentage of accountable entities and patients included

United States (all 50 states and D.C.).

Applicable level of analysis and care setting

Clinician: Individual (TIN-NPI) and Clinician: Group/Practice (TIN). Care settings include ambulatory surgical centers, ambulatory/office-based care, and hospital outpatient departments.

### 6.1.4 Attributes for Accountability Use

The measure is best suited for use within the Merit-based Incentive Payment System (MIPS) cost performance category and is currently implemented in MIPS as well as the Gastroenterology Care

MIPS Value Pathway (MVP).

The target population is Medicare fee-for-service beneficiaries who undergo a screening or surveillance colonoscopy that triggers an episode. The cohort is predominantly aged 65 and older but includes all Medicare beneficiaries who trigger an episode regardless of age, including younger beneficiaries eligible through disability or end-stage renal disease. Beneficiaries enrolled in Medicare Advantage (Part C) for any portion of the lookback period or episode window are excluded, as are beneficiaries whose primary payer is not Medicare. Full target population specifications are provided in Fields 1.15, 1.15c, and 1.15e.

The measure is best suited for accountability at the clinician and group levels, as specified in Field 1.8. Episodes are attributed under a performing physician model: a clinician is attributed for an episode when they bill a trigger code on the day of the procedure as the main performing clinician, and group-level attribution rolls up to the group associated with that trigger claim. Signal-to-noise reliability has been demonstrated at both the clinician and group levels (Section 5.2). The measure is not designed for facility-level (e.g., hospital, ASC) accountability.

The measure is best suited for the ambulatory care settings in which screening and surveillance colonoscopy is routinely performed: Hospital Outpatient Department (HOPD), Ambulatory Surgery Center (ASC), and ambulatory care office settings, as specified in Field 1.9. Episodes triggered in the inpatient hospital or emergency room setting are excluded, consistent with the measure's focus on elective ambulatory colonoscopy. Each episode is compared against episodes performed in the same care setting through the three-sub-group structure, ensuring that performance comparisons are not driven by structural cost differences across settings (Section 5.4).

Risk adjustment for social risk factors is not performed at the measure level. The risk adjustment model includes patient age, CMS HCC Version 24 categories, and 8 measure-specific clinical comorbidities, but does not include indicators for individual-level social risk or community-level socioeconomic characteristics. The rationale, presented in Section 5.4, is that the 14-day episode scope excludes the care pathways through which social and environmental factors are most strongly associated with cost variation; the CMS HCC Version 24 risk model and clinical adjustors capture patient complexity that is substantially collinear with social risk markers; and prior research on MIPS episode-based cost measures has found that adjustment for social risk factors does not meaningfully change clinician or group performance on these measures.[1] At the payment level, social and medical complexity is recognized through the MIPS Complex Patient Bonus, which adds up to 10 points to the MIPS final score for qualifying clinicians, groups, virtual groups, and APM Entities based on the average CMS HCC risk score and the dual-Medicare-Medicaid eligibility proportion of their attributed Medicare patient population.[2]

Footnotes:

[1] Sandhu AT, Bhattacharya J, Lam J, et al. Adjustment for social risk factors does not meaningfully affect performance on Medicare's MIPS clinician cost measures. *Health Aff (Millwood)*. 2020;39(9):1495-1503. PMID: 32897780.

[2] Centers for Medicare & Medicaid Services. Merit-based Incentive Payment System (MIPS) 2024 Complex Patient Bonus Fact Sheet. CMS; December 23, 2024.

### 6.2.1 Actions of Measured Entities to Improve Performance

The measure score reflects the clinician's or group's risk-adjusted average episode cost, with each episode compared against episodes performed in the same care setting, covering services from the procedure through the 14-day episode window. Per-episode cost variation attributable to clinician practice includes, among other factors, three areas where clinical evidence supports targeted action: bowel preparation adequacy, which determines whether the colonoscopy is completed without early repeat within the episode window; anesthesiologist involvement, which directly affects procedure-level cost; and post-procedure adverse event rates, which affect care utilization within the episode window. The actions described here address the primary modifiable cost drivers identified during measure development and are not an exhaustive list of clinician or group decisions that may affect episode cost.

Adequate bowel preparation is a critical determinant of effective colonoscopy and directly affects episode cost when inadequate preparation leads to incomplete procedures or early repeat examinations within the episode window.[1] Clinicians can improve performance by following current evidence-based guidelines on preparation agent selection, including appropriate use of split-dose protocols, which have been shown to reduce early repeat colonoscopy rates relative to same-day regimens,[2] and by implementing systematic patient education prior to the procedure, which meta-analytic evidence shows significantly improves preparation adequacy.[3] The primary difficulty is variability in patient adherence to preparation instructions; systematic reviews identify health literacy, age-related limitations, obesity, diabetes, opioid use, and the complexity of multi-step protocols as independent predictors of inadequate preparation.[4] The risk adjustment model accounts for several of these factors, including patient age and diabetes-related comorbidities through CMS HCC Version 24 categories, limiting their direct impact on captured performance differences across clinicians and groups. Clinicians can address remaining variation by stratifying patients for higher-risk preparation protocols based on these factors and by using validated preparation quality scales, such as the Boston Bowel Preparation Scale, to monitor outcomes at the practice level over time and identify cases requiring intervention.[5]

Anesthesiologist involvement during routine screening colonoscopy is a significant driver of per-episode cost variation. For average-risk patients, anesthesiologist-administered sedation adds meaningfully to episode cost,[6] and ASGE sedation guidelines indicate that anesthesia provider-administered sedation does not appear to improve safety compared with endoscopist-directed sedation for low-risk patients.[7,8] Clinicians and groups can improve performance by aligning anesthesiologist utilization with patient-specific clinical risk, limiting monitored anesthesia care to patients whose comorbidities or procedure complexity warrants it based on documented criteria. The difficulty is that anesthesiologist utilization is often driven by facility-level arrangements, geographic patterns, and established care conventions rather than by individual patient-level clinical indications; studies show that rates of anesthesiologist-assisted colonoscopy vary substantially across regions and are not associated with differences in patient risk, polyp detection rates, or complication rates.[9,10] Overcoming this requires coordination between proceduralists, anesthesia departments, and facility administrators to establish or adopt protocols that define clinical criteria for anesthesiologist involvement, paired with patient education on the safety and appropriateness of moderate sedation for routine procedures.

Adverse events following colonoscopy, including cardiopulmonary complications associated with

sedation, lower gastrointestinal hemorrhage, and perforation, contribute to episode costs through post-procedure care utilization, including emergency department visits and readmissions within the 14-day episode window.[11] Validity testing confirms that groups and clinicians with higher adverse event rates tend to have worse cost performance scores; results are presented in Section 5.3. Clinicians can reduce adverse event rates through pre-procedural risk stratification using the American Society of Anesthesiologists (ASA) physical status classification, evidence-based sedation management, careful polypectomy technique to reduce post-polypectomy bleeding, appropriate anticoagulation management for patients undergoing polypectomy, and adherence to quality standards published by the American Society for Gastrointestinal Endoscopy (ASGE) and the American College of Gastroenterology (ACG).[11,12] The difficulty is that some adverse events reflect patient-level risk factors that are not fully preventable and that post-procedure complications occurring outside the clinical encounter may not be captured through routine channels. Participation in endoscopy quality registries, such as the GI Quality Improvement Consortium (GIQuIC), provides practice-level benchmarking against peer institutions on adverse event rates and quality indicators, supporting identification of improvement opportunities that are not visible from individual encounter data alone.[13]

Footnotes:

[1] Jacobson BC, Anderson JC, Burke CA, et al. Optimizing bowel preparation quality for colonoscopy: consensus recommendations by the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol*. 2025;120(4):738-764. PMID: 40035345.

[2] Wang L, Sprung BS, DeCross AJ, Marino D. Split-dose bowel preparation reduces the need for early repeat colonoscopy without improving adenoma detection rate. *Dig Dis Sci*. 2018;63(5):1320-1326. PMID: 29243102.

[3] Chang CW, Shih SC, Wang HY, et al. Meta-analysis: the effect of patient education on bowel preparation for colonoscopy. *Endosc Int Open*. 2015;3(6):E646-E652. PMID: 26716129.

[4] Gandhi K, Tofani C, Sokach C, Patel D, Kastenberg D, Daskalakis C. Patient characteristics associated with quality of colonoscopy preparation: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2018;16(3):357-369. PMID: 28826680.

[5] Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston Bowel Preparation Scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc*. 2009;69(3):620-625. PMID: 19136102.

[6] Krigel A, Chen L, Wright JD, Lebwohl B. Substantial increase in anesthesia assistance for outpatient colonoscopy and associated cost nationwide. *Clin Gastroenterol Hepatol*. 2019;17(12):2489-2496. PMID: 30625407.

[7] Early DS, Lightdale JR, Vargo JJ, et al. Guidelines for sedation and anesthesia in GI endoscopy. *Gastrointest Endosc*. 2018;87(2):327-337. PMID: 29306520.

[8] Khiani VS, Soulos P, Gancayco J, Gross CP. Anesthesiologist involvement in screening colonoscopy: temporal trends and cost implications in the Medicare population. *Clin Gastroenterol Hepatol*. 2012;10(1):58-64. PMID: 21782768.

[9] Dominitz JA, Baldwin LM, Green P, Kreuter WI, Ko CW. Regional variation in anesthesia assistance during outpatient colonoscopy is not associated with differences in polyp detection or complication rates. *Gastroenterology*. 2013;144(2):298-306. PMID: 23103615.

[10] Adams MA, Saleh A, Rubenstein JH. A systematic review of factors associated with utilization of monitored anesthesia care for gastrointestinal endoscopy. *Gastroenterol Hepatol (NY)*. 2016;12(6):361-370. PMID: 27493596.

[11] Kothari ST, Huang RJ, Shaukat A, et al. ASGE review of adverse events in colonoscopy. *Gastrointest Endosc*. 2019;90(6):863-876. PMID: 31563271.

[12] Rex DK, Anderson JC, Butterly LF, et al. Quality indicators for colonoscopy. *Gastrointest Endosc*. 2024;100(3):352-381. PMID: 39177519.

[13] Shapiro JA, Holub JL, Dominitz JA, Sabatino SA, Nadel MR. Colonoscopy quality measures and adherence to follow-up guidelines among endoscopists participating in a U.S. endoscopy registry. *Gastrointest Endosc*. 2025;101(1):168-177. PMID: 39111394.

### 6.2.2 Feedback on Measure Performance

The Screening/Surveillance Colonoscopy measure has been subject to structured feedback collection at multiple points since initial development, including clinical subcommittee input, national field testing, pre-rulemaking review, public comment on proposed rules, and a formal reevaluation. Each feedback source is described below in chronological order.

During initial measure development in 2016 and 2017, CMS and Acumen convened the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee to obtain structured expert clinical input on measure design. Input was gathered through a series of structured review sessions, including a clinical subcommittee webinar in August 2017. The Subcommittee identified three clinician-attributable drivers of episode cost variation (bowel preparation quality, anesthesiologist use, and adherence to evidence-based screening intervals) and provided guidance on attribution logic, episode window duration, exclusion criteria, and the care-setting sub-group structure. This input directly shaped the measure specifications submitted for national field testing.

In fall 2017, CMS and Acumen conducted a national field test of the measure as part of a set of eight episode-based cost measures proposed for inclusion in MIPS beginning with the 2019 performance year.[1] Feedback was collected through an online survey and written comment letters submitted during a 35-day comment period, two National Provider Calls attended by approximately 1,000 clinicians, and an email and Help Desk channel.[1] In total, 219 survey responses were received including 53 written comment letters, and participants submitted approximately 120 questions during the National Provider Calls.[1] Twenty-one stakeholders provided colonoscopy-specific written feedback, including the American Gastroenterological Association (AGA) and a joint comment from the American College of Gastroenterology, AGA, and the American Society for Gastrointestinal Endoscopy.[1] Commenters noted that the level of

clinician engagement in development represented a significant improvement over the development process for earlier cost measures and praised the navigability of the performance feedback reports, while identifying areas where actionable content could be strengthened. Substantive feedback themes included recommendations on trigger code selection, sub-group structure, service assignment scope, exclusion criteria, and risk adjustment variables.[1] Colonoscopy-specific feedback was shared with the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee prior to December 2017 measure refinement webinars to inform specification changes.

Prior to initial inclusion in MIPS, the measure was reviewed by the NQF Cost and Efficiency Standing Committee under the Consensus Development Process during the Spring 2019 cycle. The Standing Committee recommended the measure for endorsement, and the NQF Consensus Standards Approval Committee approved the measure for endorsement on November 12, 2019.[2] Public comments during NQF review were submitted by specialty society stakeholders, including a joint comment from the American College of Gastroenterology, the American Society for Gastrointestinal Endoscopy, and the American Gastroenterological Association, addressing measure scope, risk adjustment, and the relationship between procedural quality and episode cost.[2] The Committee identified an opportunity for improvement to decrease costs associated with screening/surveillance colonoscopy and noted that the measure is not designed to capture the quality of bowel preparation, bounding the measure's intended scope to the cost-related dimensions of episode performance.[2]

Public comments on the proposed episode-based cost measures were received during the public comment period for the CY2019 Medicare Physician Fee Schedule (PFS) proposed rule, published in the Federal Register.[3] Multiple commenters supported adoption of the eight episode-based measures, citing the significant clinician input embedded in the development process; one commenter specifically commended CMS for convening the Clinical Subcommittee for the Screening/Surveillance Colonoscopy measure, describing it as a successful and deliberative process. Some commenters expressed concern about including the measures in MIPS scoring beginning with the 2019 performance year, recommending additional time for clinicians to familiarize themselves with measure performance data; CMS responded that extensive field testing combined with planned education and outreach were sufficient to support initial inclusion.[3]

In 2022, CMS and Acumen solicited public comment on the reevaluation of eight Wave 1 episode-based cost measures, including the Screening/Surveillance Colonoscopy measure, through a formal comment period.[4] The joint comment submitted by the American College of Gastroenterology, AGA, and the American Society for Gastrointestinal Endoscopy was the only comment that provided feedback specific to the Screening/Surveillance Colonoscopy measure.[4] The commenters recommended aligning the measure with the 2021 USPSTF colorectal cancer screening guidelines, which lowered the recommended screening start age to 45; opposed expanding the measure to include diagnostic colonoscopy, citing clinical heterogeneity between screening and diagnostic procedures, the difficulty of defining pre- and post-trigger periods across the wide range of diagnostic indications, and the availability of other episode-based cost measures (e.g., Lower Gastrointestinal Hemorrhage) that already capture diagnostic colonoscopy in specific clinical contexts.[4] Cross-cutting themes from commenters across all eight Wave 1 measures addressed episode group definition, accounting for patient heterogeneity, attribution

methodology, cost assignment to episode groups, measure development and maintenance, requests for greater public availability of MIPS cost measure performance information, and alignment with broader federal initiatives.[4] Based on the public comments and empirical analyses, CMS approved three of the eight Wave 1 measures for comprehensive reevaluation; the Screening/Surveillance Colonoscopy measure was not selected and continues through the standard annual maintenance process.[4]

Footnotes:

[1] Centers for Medicare & Medicaid Services / Acumen LLC. Field Testing Feedback Summary Report for Eight MACRA Episode-Based Cost Measures. CMS; June 2018.

[2] National Quality Forum, Cost and Efficiency Standing Committee. Cost and Efficiency, Spring 2019 Review Cycle: CDP Report. NQF; February 21, 2020.

[3] Centers for Medicare & Medicaid Services, HHS. Medicare program; revisions to payment policies under the physician fee schedule and other revisions to Part B for CY 2019; Medicare Shared Savings Program requirements; Quality Payment Program; and Medicaid Promoting Interoperability Program. Final rule. *Fed Regist*. 2018;83(226):59452-60303.

[4] Centers for Medicare & Medicaid Services / Acumen LLC. MACRA Episode-Based Cost Measures: Comprehensive Reevaluation — Public Comment Summary Report. CMS; August 2022.

### 6.2.3 Consideration of Measure Feedback

Feedback from the Clinical Subcommittee, the 2017 national field test, and pre-rulemaking review was reviewed and addressed prior to the measure's initial inclusion in MIPS in PY2019. Field test feedback was compiled by Acumen and shared with the Gastrointestinal Disease Management — Medical and Surgical Clinical Subcommittee prior to the December 2017 measure refinement webinars.[1] Pre-webinar summary sheets were prepared for each substantive feedback theme, providing context on commenter recommendations and staff analysis of specification implications. The Clinical Subcommittee convened for a two-part refinement webinar series in which members reviewed and discussed each feedback theme in detail. Specification changes were adopted through a consensus voting process requiring more than 60 percent agreement. Accepted changes were incorporated into the measure specification submitted for NQF endorsement and initial MIPS inclusion beginning with the 2019 performance year.[2] In cases where Clinical Subcommittee members acknowledged the validity of a commenter's concern but determined that the existing specification already addressed it, the concern was noted but no specification change was made.

Field test commenters noted that several trigger codes could apply to both screening and diagnostic colonoscopies, creating a risk that the episode group would capture procedures beyond the screening and surveillance cohort.[1] The Clinical Subcommittee agreed; the specification was revised to require the PT modifier on the five general-use colonoscopy trigger codes, those that may be billed for either screening or diagnostic colonoscopy, as a condition of episode triggering. In colonoscopy billing, the PT modifier identifies a colorectal cancer screening service that has been converted to a diagnostic test or other procedure; requiring it on those codes ensures the

episode group captures only procedures performed as part of colorectal cancer screening. The two remaining trigger codes specifically identify colorectal cancer screening colonoscopy and do not require the modifier. This refinement was incorporated for the PY2019 initial MIPS specification. Full trigger code specifications, including modifier requirements, are provided in the codes list (Triggers tab) and MIF Section 4.1.

Commenters recommended restructuring the place-of-service sub-groups used for performance comparison, noting that ambulatory surgery centers and office settings are clinically and economically distinct enough to warrant separate groupings.[1] The Clinical Subcommittee agreed; the sub-group structure was revised from two categories to three, based on place of service: Hospital Outpatient Department (HOPD), Ambulatory Surgery Center (ASC), and office settings. The revised structure ensures that each episode is compared against episodes performed in the same care setting. Full sub-group definitions are provided in the codes list (Sub_Groups_Details tab) and the MIF Sub-Groups section.

Commenters identified a set of clinical services included in the initial episode that lacked a plausible clinical connection to colonoscopy care.[1] The Clinical Subcommittee reviewed each service and agreed to narrow the service assignment logic; more than 14 service categories were removed. Removed services included codes associated with fracture of the femur, dizziness, malaise, non-productive cough, non-specific gastrointestinal conditions, flatulence, tachycardia, chest pain, bradycardia, palpitations, head injury, and intraoperative and post-procedural complications. The revised service assignment logic restricts episode-attributed costs to services that are clinically consistent with colonoscopy care and its expected post-procedural course. Full service assignment specifications are provided in the codes list (Service Assignment tab) and MIF Section 4.4.

Commenters recommended additional exclusions to remove clinically heterogeneous subpopulations whose cost patterns are not representative of routine screening colonoscopy.[1] Three exclusions were added for the PY2019 initial MIPS specification. Episodes involving patients with active inflammatory bowel disease diagnoses are excluded because colonoscopy costs in that population reflect disease management rather than routine screening. Episodes involving endoscopic mucosal resection are excluded to distinguish complex therapeutic interventions from standard screening procedures. Episodes in which an upper gastrointestinal endoscopy is performed concurrently are excluded to avoid conflating episode costs attributable to different procedures. Full exclusion criteria are provided in MIF Section 3.0.

Commenters recommended adding an anticoagulant use variable to the risk adjustment model to account for clinical cost differences associated with patients on anticoagulant therapy at the time of colonoscopy, where pre-procedure management and post-procedure monitoring may increase episode cost independently of colonoscopist decisions.[1] The Clinical Subcommittee accepted this recommendation; an anticoagulant use adjustor was incorporated into the PY2019 risk adjustment specification. Full risk adjustment variable specifications are provided in MIF Section 5.0.

Attribution logic was refined for the PY2019 initial MIPS specification to incorporate post-operative modifier codes, which identify clinicians who billed services only in a post-operative care role and were not involved in the main procedure. Clinicians billing exclusively in a post-operative role are excluded from episode attribution, ensuring that only the performing proceduralist is held responsible for episode costs.

MAP conditionally supported the measure pending NQF endorsement and raised two concerns: that the risk adjustment model adequately capture clinical and social risk factors, and that the measure not create incentives for care stinting.[2] The risk adjustment model addresses the first concern through a comprehensive set of adjustors, including 86 hierarchical condition category (HCC) indicators and 8 measure-specific comorbidity factors; risk model calibration data are provided in Section 5.4. No reports or other indication of care stinting associated with this measure have been received, and CMS will continue to monitor for this potential consequence.

Public comments received during the CY2019 PFS proposed rule comment period on the Screening/Surveillance Colonoscopy measure were reviewed and did not result in additional specification changes.[3]

The suggestion from the 2022 Wave 1 reevaluation public comment period to align the measure with the 2021 USPSTF colorectal cancer screening guidelines, which lowered the recommended screening start age to 45, has been addressed through the existing specification.[4] The measure's patient cohort is defined by claims codes associated with screening and surveillance colonoscopy rather than a specific age range; changes in USPSTF-recommended screening age affect the prevalence of eligible beneficiaries in the denominator but do not require specification changes. The recommendation to expand the measure to include diagnostic colonoscopy was not implemented, consistent with the Clinical Subcommittee's original design intent to maintain the clinical homogeneity of the screening and surveillance cohort.

Footnotes:

[1] Centers for Medicare & Medicaid Services / Acumen LLC. Field Testing Feedback Summary Report for Eight MACRA Episode-Based Cost Measures. CMS; June 2018.

[2] National Quality Forum, Cost and Efficiency Standing Committee. Cost and Efficiency, Spring 2019 Review Cycle: CDP Report. NQF; February 21, 2020.

[3] Centers for Medicare & Medicaid Services, HHS. Medicare program; revisions to payment policies under the physician fee schedule and other revisions to Part B for CY 2019; Medicare Shared Savings Program requirements; Quality Payment Program; and Medicaid Promoting Interoperability Program. Final rule. *Fed Regist*. 2018;83(226):59452-60303.

[4] Centers for Medicare & Medicaid Services / Acumen LLC. MACRA Episode-Based Cost Measures: Comprehensive Reevaluation — Public Comment Summary Report. CMS; August 2022.

### 6.2.4 Progress on Improvement

The Screening/Surveillance Colonoscopy measure was first implemented in MIPS in PY2019,[1] with episode attribution reaching clinicians and groups performing colorectal cancer screening nationwide. As of PY2024, 977,369 final episodes were attributed to 3,205 groups and 15,017 clinicians across the country, representing 971,558 unique Medicare beneficiaries (the total episode count exceeds the unique-beneficiary count because a small number of beneficiaries had more than one eligible episode in PY2024) and accounting for approximately 0.3 percent of total Medicare Part A and Part B expenditures.[2] This volume reflects the continued prevalence of

colorectal cancer screening in the Medicare population and the breadth of clinicians subject to performance reporting under this measure.

A meaningful performance gap persists across the attributed population, as documented in Field 2.4. In PY2024, observed episode costs ranged from approximately $730 at the 10th percentile to $1,568 at the 90th percentile, a spread of approximately $838, with further variation at the tails. The full decile distribution of performance scores at the group and clinician levels is provided in Table 1. The continued spread in performance scores indicates that substantial clinician-attributable cost variation remains, with measurable differences between high- and low-cost performers that are not explained by patient case mix. Because the measure is a relative cost measure in which each episode is compared against episodes performed in the same care setting rather than against a fixed benchmark, a continuing distribution of performance is expected; performance scores indicate where each clinician or group falls relative to peers, so a range of better and worse performers is the expected result rather than a sign of stalled improvement.

Under the standard deviation-anchored benchmarking methodology effective for the CY2024 performance period, the national median episode cost is anchored to a MIPS cost score of 7.5 out of 10 for this measure.[3] Clinicians and groups at or below the national median receive at least 7.5 MIPS points from this measure. Based on the PY2024 performance distribution, approximately half of attributed clinicians and groups fall at or below the median and receive at least 7.5 points; the remaining half perform above the national median, receive fewer than 7.5 points, and represent a direct opportunity for cost improvement through the actions described in Field 6.2.1.

Sub-group cost distributions in PY2024 demonstrate substantial pre-adjustment variation across care settings, with observed mean episode cost of $976 (SD $238) in Ambulatory Surgery Centers (508,891 episodes), $1,446 (SD $355) in Hospital Outpatient Departments (436,112 episodes), and $606 (SD $214) in Office settings (32,366 episodes). This pre-adjustment variation is normalized through per-sub-group risk adjustment to a near-uniform risk-adjusted mean of approximately $1,174 in each sub-group, confirming that the sub-group structure successfully neutralizes structural cost differences across care settings while preserving meaningful within-sub-group variation for performance comparison.

The measure continues to produce reliable scores that distinguish clinician and group performance across the attributed population. Signal-to-noise reliability results at both attribution levels are presented in Section 5.2.

CMS provides annual performance feedback reports to attributed clinicians and groups. Performance feedback reports available through the Quality Payment Program portal include cost performance category scores, episode sub-group breakdowns comparing entity performance to national averages, utilization and cost detail organized by clinical theme and service category, and attributed beneficiary lists with episode-level cost data enabling review of individual attributed episodes and associated costs.

Footnotes:

[1] Centers for Medicare & Medicaid Services, HHS. Medicare program; revisions to payment policies under the physician fee schedule and other revisions to Part B for CY 2019; Medicare Shared Savings Program requirements; Quality Payment Program; and Medicaid Promoting

Interoperability Program. Final rule. *Fed Regist*. 2018;83(226):59452-60303.

[2] Centers for Medicare & Medicaid Services. 2026 MIPS Summary: Cost Measures. CMS; 2026. https://www.cms.gov/files/document/2026-mips-summary-cost-measures.pdf.

[3] Centers for Medicare & Medicaid Services. Calendar Year (CY) 2025 Medicare Physician Fee Schedule (PFS) Final Rule: Quality Payment Program (QPP) Fact Sheet. CMS; November 2024.

### 6.2.5 Unexpected Findings

Implementation of the Screening/Surveillance Colonoscopy measure has surfaced one operational consideration with the cost scoring benchmark methodology that was addressed through a recent MIPS scoring policy change, alongside concerns about care stinting and selective referral raised during pre-rulemaking review. CMS has not received any indication or report of stinting or selective referral associated with the measure.

Revision to cost scoring benchmarking methodology (CY2025 Final Rule, effective CY2024 performance period): Prior to the CY2025 PFS Final Rule, MIPS cost measure points were assigned using a decile-percentile benchmarking methodology, with 1 to 10 achievement points distributed across 10 percentile ranges. Under this approach, clinicians and groups with episode costs near the national median received scores in approximately the 2 to 3 point range out of 10. CMS observed in the CY2025 PFS rulemaking that this distribution did not optimally differentiate near-median performers and finalized a revised benchmarking methodology in the CY2025 PFS Final Rule, effective beginning with the CY2024 performance period and 2026 MIPS payment year, with final scores released in June 2025.[1] Under the revised methodology, the median episode cost for a measure is anchored to a score equal to the applicable MIPS performance threshold; for the CY2024 performance period, that threshold is 75 points, translating to a cost score of 7.5. Cut-offs for achievement point ranges are calculated based on standard deviations from the median cost rather than decile percentiles. The revised methodology better differentiates cost performance: clinicians and groups with episode costs near the national median receive a score near the performance threshold, while clinicians and groups with meaningfully higher or lower costs receive scores adjusted in proportion to their deviation from the median. The revised benchmarking applies to all clinicians and groups attributed for this measure beginning with PY2024 scoring.

Stinting and selective referral concerns: The MAP raised stinting of care as a potential unintended consequence during review of MIPS episode-based cost measures, noting that cost measures creating financial incentives to reduce resource utilization could discourage clinicians from providing clinically necessary services.[2] Selective referral of lower-risk patients to ambulatory care settings is a related potential consequence; the measure's three care-setting sub-groups and comprehensive risk adjustment model, including patient comorbidity, CMS HCC Version 24 categories, and sub-group indicators, are designed to mitigate this risk. CMS has not received any indication or report of stinting or selective referral associated with the measure, and CMS will continue to monitor for both potential consequences.

Footnotes:

[1] Centers for Medicare & Medicaid Services. Calendar Year (CY) 2025 Medicare Physician Fee Schedule (PFS) Final Rule: Quality Payment Program (QPP) Fact Sheet. CMS; November 2024.

[2] National Quality Forum, Cost and Efficiency Standing Committee. Cost and Efficiency, Spring 2019 Review Cycle: CDP Report. NQF; February 21, 2020.

### 7.1 Supplemental Attachment

2025-12-py2026-mif-ebcm-ss-clnscpy.pdf

### Developer POC email

pcmp-info@acumenllc.com

### Measure Developer POC

PCMP Measure Development Team
Acumen, LLC
Burlingame, CA
United States

### The measure developer is different from the measure steward

Yes

### Steward Address

Raymond Desautels IV
Baltimore, MD
United States

### Steward Organization

Centers for Medicare & Medicaid Services

### Steward Organization URL

https://cms.hhs.gov

### Steward POC email

Raymond.DesautelsIV@cms.hhs.gov