Quality Payment
PROGRAM

Colon and Rectal Resection

Measure Justification Form

June 2021

Contents

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Colon and Rectal Resection measure. The form is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology[1] and measure Codes List file[2], which together, comprise the specifications for this cost measure.

1.1 Project Title and Overview

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop care episode and patient condition groups for use in cost measures to meet the requirements of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). The contract name is "Physician Cost Measure and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.2 Measure Name

Colon and Rectal Resection Episode-Based Cost Measure

1.3 Type of Measure

Cost/Resource Use

[1]CMS, "Colon and Rectal Resection Measure Methodology," *MACRA Feedback Page,* *https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback*
[2]CMS, "Colon and Rectal Resection Measure Codes List" *MACRA Feedback Page*, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback

2.0 Measure Testing: Importance

2.1 Evidence to Support the Measure Focus

2.1.1 Measure Description

The Colon and Rectal Resection cost measure evaluates clinicians and clinician groups risk-adjusted cost to Medicare for patients who receive colon or rectal resections for either benign or malignant indications. The measure score is a clinician or clinician group's average risk-adjusted cost across all attributed episodes for the episode group. This procedural measure includes services that are clinically related and under the reasonable influence of the attributed clinician during the 15 days prior to the resection procedure that opens or "triggers" the episode and the 90 days after the procedure. Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period are eligible for the measure.

2.1.2 Evidence for Measure Focus

The Colon and Rectal Resection measure was developed for use in the Merit-based Incentive Payment System (MIPS) to meet the requirements of the Social Security Act section 1848(r), added by the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). MIPS aims to reward high-value care by measuring clinician performance through four areas: quality, improvement activities, promoting interoperability, and cost. Each category assesses different aspects of care, and the categories are weighted such that they are combined into one composite score. CMS is introducing MIPS Value Pathways (MVPs) as a way to align and connect quality measures, cost measures, and improvement activities across performance categories of MIPS for different specialties or conditions. MVPs aim to provide a holistic assessment of clinician value for a specific type of care to achieve better healthcare outcomes and lower costs for patients. The use of cost measures is required by statute, and their purpose is to assess resource use. To be effective, they should capture costs related to a clinician's care decisions and account for factors outside of their influence.

This measure provides clinicians with information about their costs of care that they can use to understand the costs associated with their decision-making. Clinicians play an important role in variation in health care expenditures due to their ability to affect costs.[3] A cost measure offers opportunity for improvement if clinicians can exercise influence on the intensity or frequency of a significant share of costs during the episode, or if clinicians can achieve lower spending and better care quality through changes in clinical practice.

According to the literature and feedback received through stakeholder input activities to date, this measure's focus represents an area where there are opportunities for improvement. Primary opportunities for improvement include selecting the appropriate modality of surgery and adopting prevention strategies to mitigate the risk of common postoperative complications.

A clinician's selected method to performing a colorectal surgery has a significant impact on patient outcomes. Colorectal surgery can be performed using 3 different modalities: open, laparoscopic, and robotic. The benefits of performing colon resections laparoscopically or robotically are well established. These minimally invasive approaches are associated with reduced lengths of stay, reduced utilization of post-acute care, lower postoperative readmission

[3]David Cutler et al., "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," American Economic Journal: Economic Policy 11, no. 1 (February 1, 2019): 192–221, https://doi.org/10.1257/pol.20150421.

rates, and lower mortality rates, especially among the older adult population.[4,5] Although the use of such techniques may be more limited in scope for rectal resections due to added technical complexity, recent studies indicate that these techniques may also have a role in reducing postoperative complications following surgery for rectal cancer treatment. Specifically, studies and reviews of meta-analyses have demonstrated that robotic or laparoscopic surgery for rectal cancer treatment may reduce the incidence of postoperative complications when compared to open surgery.[6,7,8,9] There remains wide variation in the utilization of open surgery, laparoscopic, and robotic approaches for different diagnoses.[10] In 2012, open surgeries constituted 65.4% of all colorectal surgeries nationwide, while 31.2% and 3.4% were performed using laparoscopic or robotic techniques, respectively.[11] Efforts to increase adoption of minimally invasive techniques, when appropriate, through surgeon education and training could be effective strategies to curb costs associated with prolonged lengths of stay and readmission.

Colorectal resection accounts for a substantial share of postoperative readmissions among inpatient procedures, with one study approximating a 30-day postoperative readmission rate of 13.7%.[12,13,14] Estimates of the inpatient cost for readmission following colorectal surgery range from $9,000 to $12,000 across studies.[15] One study estimates that readmissions associated

---

[4]Umashankkar Kannan et al., “Laparoscopic vs Open Partial Colectomy in Elderly Patients: Insights from the American College of Surgeons - National Surgical Quality Improvement Program Database,” World Journal of Gastroenterology 21, no. 45 (December 7, 2015): 12843–50, https://doi.org/10.3748/wjg.v21.i45.12843.

[5]Cigdem Benlice et al., “Robotic, Laparoscopic, and Open Colectomy: A Case-Matched Comparison from the ACS-NSQIP,” International Journal of Medical Robotics and Computer Assisted Surgery 13, no. 3 (September 1, 2017), https://doi.org/10.1002/rcs.1783.

[6]Carly R. Richards et al., "Safe Surgery in the Elderly: A Review of Outcomes Following Robotic Proctectomy from the Nationwide Inpatient Sample in a Cross-Sectional Study," Annals of Medicine and Surgery 44 (2019/08/01/ 2019), https://doi.org/10.1016/j.amsu.2019.06.004; S. A. Antoniou et al., "Laparoscopic Colorectal Surgery Confers Lower Mortality in the Elderly: A Systematic Review and Meta-Analysis of 66,483 Patients," Surg Endosc 29, no. 2 (Feb 2015), https://doi.org/10.1007/s00464-014-3672-x.

[7]C. Simillis et al., "Open Versus Laparoscopic Versus Robotic Versus Transanal Mesorectal Excision for Rectal Cancer: A Systematic Review and Network Meta-Analysis," Ann Surg 270, no. 1 (Jul 2019), https://doi.org/10.1097/SLA.0000000000003227; Antonio Biondi et al., "Laparoscopic Vs. Open Approach for Colorectal Cancer: Evolution over Time of Minimal Invasive Surgery," BMC surgery 13 Suppl 2, no. Suppl 2 (2013), https://doi.org/10.1186/1471-2482-13-S2-S12; D. L. Waitzberg et al., "Postsurgical Infections Are Reduced with Specialized Nutrition Support," World J Surg 30, no. 8 (Aug 2006), https://doi.org/10.1007/s00268-005-0657-x.

[8]Meng Tse Gabriel Lee et al., “Trends and Outcomes of Surgical Treatment for Colorectal Cancer between 2004 and 2012- An Analysis Using National Inpatient Database,” Scientific Reports 7, no. 1 (December 1, 2017), https://doi.org/10.1038/s41598-017-02224-y.

[9]Zhong Lin et al., “Short- and Long-Term Outcomes of Laparoscopic versus Open Surgery for Rectal Cancer: A Systematic Review and Meta-Analysis of Randomized Controlled Trials,” Medicine (United States) (Lippincott Williams and Wilkins, December 1, 2018), https://doi.org/10.1097/MD.0000000000013704.

[10]Philipp Kirchhoff, Pierre Alain Clavien, and Dieter Hahnloser, “Complications in Colorectal Surgery: Risk Factors and Preventive Strategies,” Patient Safety in Surgery (BioMed Central, March 25, 2010), https://doi.org/10.1186/1754-9493-4-5.

[11]Meng Tse Gabriel Lee et al., “Trends and Outcomes of Surgical Treatment for Colorectal Cancer between 2004 and 2012- An Analysis Using National Inpatient Database.”

[12]John D. Birkmeyer et al., “Medicare Payments for Common Inpatient Procedures: Implications for Episode-Based Payment Bundling,” Health Services Research 45, no. 6 PART 1 (2010): 1783–95, https://doi.org/10.1111/j.1475-6773.2010.01150.x.

[13]Elizabeth C. Wick et al., “Readmission Rates and Cost Following Colorectal Surgery,” Diseases of the Colon and Rectum 54, no. 12 (December 2011): 1475–79, https://doi.org/10.1097/DCR.0b013e31822ff8f0.

[14]Rachelle N. Damle et al., “Clinical and Financial Impact of Hospital Readmissions after Colorectal Resection: Predictors, Outcomes, and Costs,” Diseases of the Colon and Rectum 57, no. 12 (December 2014): 1421–29, https://doi.org/10.1097/DCR.0000000000000251.

[15]Birkmeyer et al., “Medicare Payments for Common Inpatient Procedures: Implications for Episode-Based Payment Bundling”; Wick et al., “Readmission Rates and Cost Following Colorectal Surgery”; Damle et al., “Clinical and Financial Impact of Hospital Readmissions after Colorectal Resection: Predictors, Outcomes, and Costs.”

with colorectal surgery account for approximately $300 million in costs annually across the nation.[16] Postoperative readmission is strongly associated with the occurrence of common complications such as surgical site infection (SSI), ileus, and urinary tract infections. Occurrence of SSI alone is estimated to contribute an additional estimated cost of $40,500 per patient and an estimated national total of $3 billion per year.[17] Applying prevention strategies to emergency colorectal surgeries based on clinical guidelines for an "Enhanced Recovery After Surgery" (ERAS) protocol can decrease these post-operative complications and reduce morbidity. ERAS is a standard of perioperative care for elective colorectal surgeries; however, there appears to be low implementation of an ERAS protocol in emergent settings. This may be due to the fact that patients undergoing emergent surgeries have more risk factors and comorbidities that must be managed.[18] Expanding the implementation of ERAS protocols has the potential to improve overall quality of care and reduce related services and their associated costs.

A diverting stoma, in which a surgeon externally diverts the flow of feces may be another avenue to mitigate common complications such as anastomotic leaks and the associated costs.[19,20] Although there are benefits and tradeoffs to fecal diversion to protect an anastomosis, certain factors may indicate cases in which a diverting stoma may be the preferred surgical approach. For example, there is generally a consensus among researchers that the presence of a diverting stoma lowers the risk of anastomotic leak and can lower the risk of developing pelvic sepsis for patients who undergo a low anterior resection.[21,22,23,24] Since the risks associated with diverting stomas are well documented, preventative pathways have been developed to address the potential for dehydration and other common causes of readmission due to colorectal surgeries.[25] For example, one study reported reducing the rate of hospital readmissions and entirely eliminating readmissions related to dehydration by employing an educational intervention for patients with new, temporary or permanent ileostomies.[26] This suggests that coupling diverting stomas with robust patient education may result in improved outcomes following colorectal surgery. Fecal diversion is also demonstrated to have a protective effect in terms of decreased mortality and morbidity for other high-risk cases. For example, recent studies have identified primary anastomosis with diversion as the preferred option for cases with active infections, such as peritonitis from diverticular disease, compared to Hartmann's

[16]Wick et al., "Readmission Rates and Cost Following Colorectal Surgery."
[17]Megan C. Turner and John Migaly, "Surgical Site Infection: The Clinical and Economic Impact," Clinics in Colon and Rectal Surgery 32, no. 3 (May 2019): 157–65, https://doi.org/10.1055/s-0038-1677002.
[18]Varut Lohsiriwat and Romyen Jitmungngan, "Enhanced Recovery after Surgery in Emergency Colorectal Surgery: Review of Literature and Current Practices," World Journal of Gastrointestinal Surgery 11, no. 2 (February 27, 2019): 41–52, https://doi.org/10.4240/wjgs.v11.i2.41.
[19]Silvia Palmisano et al., "Diverting Stoma," in Rectal Cancer (Springer, Milano, 2013), 131–37, https://doi.org/10.1007/978-88-470-2670-4_10.
[20]AL Peel and EW Taylor, "Proposed Definitions for the Audit of Postoperative Infection: A Discussion Paper. Surgical Infection Study Group.," Annals of the Royal College of Surgeons of England 73, no. 6 (March 1991): 385–88.
[21]Jeremy Meyer et al., "Reducing Anastomotic Leak in Colorectal Surgery: The Old Dogmas and the New Challenges," World Journal of Gastroenterology (Baishideng Publishing Group Co., Limited, September 14, 2019), https://doi.org/10.3748/wjg.v25.i34.5017.
[22]Kirchhoff, Clavien, and Hahnloser, "Complications in Colorectal Surgery: Risk Factors and Preventive Strategies."
[23]Scott R Steele et al., "Improving Outcomes and Cost-Effectiveness of Colorectal Surgery," n.d., https://doi.org/10.1007/s11605-014-2643-9.
[24]Alexis Plasencia and Heidi Bahna, "Diverting Ostomy: For Whom, When, What, Where, and Why," Clinics in Colon and Rectal Surgery 32, no. 3 (May 2019): 171–75, https://doi.org/10.1055/s-0038-1677004.
[25]Steele et al., "Improving Outcomes and Cost-Effectiveness of Colorectal Surgery."
[26]Deborah Nagle et al., "Ileostomy Pathway Virtually Eliminates Readmissions for Dehydration in New Ostomates," Diseases of the Colon and Rectum 55, no. 12 (December 2012): 1266–72, https://doi.org/10.1097/DCR.0b013e31827080c1.

procedure.[27] As such, diversions may play an important role in improving outcomes and reducing associated downstream costs for select high-risk colorectal cases.

2.2 Performance Gap

2.2.1 Rationale

Colorectal resection, or colectomy, is a common treatment for colorectal cancer and complications related to diverticular disease. According to the Agency for Healthcare Research and Quality, about 320,000 colorectal resection procedures were performed annually between 2001 and 2011.[28] Colorectal cancer is the second leading cause of cancer-related deaths and the third most common cancer in both men and women in the United States. Colorectal cancer is especially common in the 85 and older adult population, with an incidence of 237 per 100,000 persons in 2016.[29] Similarly, diverticular disease primarily affects older adults, occurring in 50 to 70% of those aged 80 or older. Diverticular disease accounts for more than $2 billion in treatment costs annually. While diverticular disease is usually an asymptomatic condition, the incidence of complications, such as colonic diverticulitis, increases with age.[30,31] Morbidity and the risk of postoperative complications following colorectal resection also increase significantly for patients above age 65.[32] According to the literature, a single colectomy is estimated to cost $25,000, and this cost can increase to nearly $50,000 with post-operative complications.[33,34] Estimates of index hospitalization costs for colorectal surgery are similar and have been shown to range between about $18,000 to $21,000 among a cohort of Medicare patients, with variation in the cost of care provided within a year of the surgery largely driven by readmissions and post-acute care utilization.[35] Given the variation and frequency of treating colorectal cancer and complications related to diverticular disease with colectomy procedures in Medicare patients,

---

[27]Amy L. Lightner and John H. Pemberton, "The Role of Temporary Fecal Diversion," Clinics in Colon and Rectal Surgery 30, no. 3 (July 1, 2017): 178–83, https://doi.org/10.1055/s-0037-1598158; Valerie Bridoux et al., "Hartmann's Procedure or Primary Anastomosis for Generalized Peritonitis Due to Perforated Diverticulitis: A Prospective Multicenter Randomized Trial (DIVERTI)," Journal of the American College of Surgeons 225, no. 6 (December 1, 2017): 798–805, https://doi.org/10.1016/j.jamcollsurg.2017.09.004; Plasencia and Bahna, "Diverting Ostomy: For Whom, When, What, Where, and Why."

[28]Audrey J Weiss and Anne Elixhauser, Trends in Operating Room Procedures in U.S. Hospitals, 2001–2011: Statistical Brief #171, Healthcare Cost and Utilization Project (HCUP) Statistical Briefs (Rockville, MD: Agency for Healthcare Research and Quality, 2014), http://www.ncbi.nlm.nih.gov/pubmed/24851286; Samuel Eisenstein, Sarah Stringfield, and Stefan D. Holubar, "Using the National Surgical Quality Improvement Project (NSQIP) to Perform Clinical Research in Colon and Rectal Surgery," Clinics in Colon and Rectal Surgery 32, no. 1 (2019): 41–53, https://doi.org/10.1055/s-0038-1673353.

[29]U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, "U.S. Cancer Statistics Data Visualizations Tool," n.d., https://gis.cdc.gov/Cancer/USCS/DataViz.html.

[30]Chien Kuo Liu, Hsi Hsien Hsu, and She Meng Cheng, "Colonic Diverticulitis in the Elderly," International Journal of Gerontology (Elsevier (Singapore) Pte Ltd, March 1, 2009), https://doi.org/10.1016/S1873-9598(09)70015-8.

[31]Neda Valizadeh, Kunal Suradkar, and Ravi P Kiran, "Specific Factors Predict the Risk for Urgent and Emergent Colectomy in Patients Undergoing Surgery for Diverticulitis," The American Surgeon 84, no. 11 (November 1, 2018): 1781–86,http://www.ncbi.nlm.nih.gov/pubmed/30747633.

[32]Mehraneh D. Jafari et al., "Colorectal Cancer Resections in the Aging US Population: A Trend toward Decreasing Rates and Improved Outcomes," JAMA Surgery (American Medical Association, June 1, 2014), https://doi.org/10.1001/jamasurg.2013.4930.

[33]Faiz Gani et al., "Bundled Payments for Surgical Colectomy among Medicare Enrollees: Potential Savings vs the Need for Further Reform," JAMA Surgery 151, no. 5 (May 1, 2016): e160202, https://doi.org/10.1001/jamasurg.2016.0202.

[34]David N Flynn et al., "The Impact of Complications Following Open Colectomy on Hospital Finances: A Retrospective Cohort Study," Perioperative Medicine 3, no. 1 (March 7, 2014): 1, https://doi.org/10.1186/2047-0525-3-1.

[35]Zaid M. Abdelsattar, John D. Birkmeyer, and Sandra L. Wong, "Variation in Medicare Payments for Colorectal Cancer Surgery," Journal of Oncology Practice 11, no. 5 (September 30, 2015): 391–95, https://doi.org/10.1200/jop.2015.004036.

the Colon and Rectal Resection cost measure represents an opportunity for improvement on overall cost performance.

The Colon and Rectal Resection episode-based cost measure was recommended for development by an expert clinician committee—the General and Colorectal Surgery Clinical Subcommittee. Based on the initial recommendations from the Clinical Subcommittee, the subsequent measure-specific Clinician Expert Workgroup provided extensive, detailed input on this measure.

2.2.2 Performance Scores

To demonstrate the performance gap captured in the measure, Table 1 below presents a distribution of performance scores for 759 clinician group practices (identified by Taxpayer Identification Number, or TIN) and 502 practitioners (identified by a unique TIN and National Provider Identifier pair, or TIN-NPI) attributed at least 20 episodes in 2019. These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a MIPS eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced Alternate Payment Model participation).

Table 1. Distribution of Observed over Expected (O/E) Ratio

Metric	TIN	TIN-NPI
Mean O/E ratio	1.01	1.01
O/E ratio Interquartile Range (IQR)	0.09	0.10
O/E ratio percentile		
10^{th}	0.93	0.92
25^{th}	0.96	0.95
50^{th}	1.00	1.00
75^{th}	1.05	1.05
90^{th}	1.10	1.10

3.0 Scientific Acceptability

3.1 Data Sample Description

3.1.1 Type of Data Used for Testing

Medicare administrative claims, Long-Term Minimum Data Set (MDS), Medicare Enrollment Database (EDB), Common Medicare Environment (CME), and United States Census Bureau's American Community Survey (ACS).

3.1.2 Specific Dataset Used for Testing

The Colon and Rectal Resection measure uses Medicare Part A and Part B claims data maintained by CMS. Parts A and B claims data are used to build episodes of care, calculate episode costs, and construct risk adjustors. Episode costs are payment standardized and risk adjusted to ensure accurate comparison of cost across clinicians. Payment standardization adjusts the allowed amount for a Medicare service to limit observed differences in costs to those that may result from health care delivery choices. Data from the EDB are used to determine beneficiary-level (or patient-level) exclusions and secondary risk adjustors, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), patient birth dates, and patient death dates. The risk adjustment model also accounts for expected differences in payment for services provided to patients in long-term care based on data from the MDS. Specifically, the MDS is used to create the long-term care indicator variable in risk adjustment.

For measure testing, data from the ACS and CME are used in analyses evaluating social risk factors in risk adjustment.

3.1.3 Dates of the Data Used in Testing

Colon and Rectal Resection episodes ending from January 1, 2019 through December 31, 2019.

3.1.4 Levels of Analysis Tested

Individual clinician (identified by combination of TIN and NPI) and clinician group/practice (identified by TIN).

3.1.5 Entities Included in the Testing and Analysis

After applying exclusions, the final population for analyses included 759 clinician group practices and 502 practitioners who were attributed 20 or more Colon and Rectal Resection episodes during the measurement period. Episodes from all 50 States and the District of Columbia triggered in the following settings were included:

- Ambulatory surgery center (ASC);
- Hospital outpatient department (HOPD); and
- Hospital inpatient acute care facility

3.1.6 Patient Cohort Included in the Testing and Analysis

54,414 Medicare patients, with a mean age of 73.7 (from 54,626 episodes) were included in measure analyses.

The patient population for the Colon and Rectal Resection measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who undergo a procedure for colon or rectal resections for either benign or malignant indications that trigger a Colon and Rectal Resection episode, as identified by the trigger Current Procedural

Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) code(s) on Part B Physician/Supplier claims. If the procedure occurs in an inpatient setting, the concurrent inpatient stay must have a relevant admission (defined as an inpatient claim with Medicare Severity-Diagnosis Related Group [MS-DRG] 329-334). Patients and their episodes were excluded from the sample if they met a set of exclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients who undergo procedures for colon or rectal resections for either benign or malignant indications.

The exclusion criteria are:

- The patient has a primary payer other than Medicare for any time overlapping the episode window or in the 120-day lookback period prior to the trigger day.
- The patient was not enrolled in Medicare Parts A and B for the entirety of the 120-day lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.
- The episode cannot be attributed to a main surgeon.
- The patient death occurs before the end of episode.
- The episode trigger claim was not performed in an Outpatient (OP) hospital, IP hospital, or ASC setting based on its place of service.
- The episode trigger claim occurred in an IP facility that was not a short-term stay acute hospital, as defined by subsection (d).[36]
- Where there is a concurrent inpatient stay with the trigger, the inpatient stay does not have a MS-DRG related to colon and rectal resections (i.e., MS-DRGs 329-334).
- The episode is an outlier case.
- The patient was transferred within 3 days prior to an IP admission.
- The patient had a recent major bowel surgery.
- The patient elected to leave against medical advice.
- The patient received a Left Ventricular Assist Device (LVAD).

To determine whether the Colon and Rectal Resection measure's exclusion criteria distort patient characteristics on episodes, we produced and analyzed distributions of patient characteristics (age, race, sex, dual eligibility status, income, unemployment, hierarchical condition categories [HCCs]) for (i) episodes with exclusion criteria, (ii) episodes without exclusion criteria, (iii) patients with exclusion criteria, and (iv) patients without exclusion criteria.

This analysis shows that the Colon and Rectal Resection measure's exclusion criteria have a minimal to moderate effect on the percentage of patients of any particular demographic category. The difference between patients being excluded and included in the measure is generally less than 7.22 percentage points across each of the characteristics in the analysis at TIN level testing, and less than 4.40 percentage points at TIN-NPI level testing. There is a slightly higher difference for patients with HCC11 "Colorectal, Bladder, and Other Cancers" (12.62 percentage points at the TIN-level and 6.87 percentage points at the TIN-NPI level). The percentage of patients aged 65 to 69 is 23.21% without applying the exclusion criteria, compared to 24.65% at TIN level testing once exclusions are applied. Furthermore, the difference in the percentage of patients across race categories with and without the exclusion criteria is less than 3.14 percentage points at both the TIN and TIN-NPI levels of testing. When

[36]Only stays at IP facilities that are paid under a short-term stay acute hospital as defined by subsection (d) will be included. Subsection (d) hospitals are hospitals in the 50 states and D.C. other than: psychiatric hospitals, rehabilitation hospitals, hospitals whose inpatients are predominantly under 18 years old, hospitals whose average inpatient length of stay exceeds 25 days, and hospitals involved extensively in treatment for or research on cancer. For details on the identification of these hospitals, please refer to the CCN definitions for Short-term (General and Specialty) Hospitals facility types in Chapter 2, Section 2779A1 of the CMS State Operation Manual.

comparing the breakdown of male and female patients at the TIN-NPI level, there is a difference of 0.88 percentage points between the female and male patient populations with and without exclusion criteria. Similarly, at TIN level testing, there is a difference of 0.97 percentage points between the share of male and female patients. These results indicate that there is a minimal to moderate shift in patient characteristics as a result of using the exclusion criteria listed above at both TIN and TIN-NPI levels of testing.

3.1.7 Social Risk Factors Included in Analysis

The social risk factors analyzed were variables from the ACS, EDB, and CME. ACS variables are either at the Census Block Group or Zone Improvement Plan (ZIP) Code level. Social risk variables analyzed include the following:

- Race (EDB)
  - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
  - Female, male
- Dual status (CME)
  - Full dual, partial dual, and non-dual to indicate whether a patient is dually enrolled in Medicare and Medicaid
- Income (ACS)
  - Low Income: median income < 33rd percentile nationally
  - Medium Income: median income in the interval spanning the 33rd percentile to the 66th percentile nationally
  - High Income: median income > 66th percentile
- Education (ACS)
  - Education < High School: when % with < high school education is the highest for a given Census Block Group
  - Education = High School: when % with only high school is the highest
  - Education > High School: when % with > high school is the highest
- Employment (ACS)
  - Unemployment Rate > 10%
  - Unemployment Rate <= 10%
- Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index (ACS)
  - Continuous variable (composite score of multiple community-level metrics, such as property values, density of living spaces, and poverty level) that can theoretically range from 0 to 100[37]

3.2 Reliability Testing

3.2.1 Level of Reliability Testing

The following levels of reliability were tested: critical data elements used in the measure and performance measure score (e.g., signal-to-noise analysis).

3.2.2 Method of Reliability Testing

Data Element Reliability

The Colon and Rectal Resection measure is constructed using CMS claims data, as described in Section 3.1.2. CMS has implemented several auditing programs to assess overall claims code accuracy, ensure appropriate billing, and recoup any overpayments. CMS routinely

[37]Refer to Section 3, page 42 of this AHRQ publication for the scoring algorithm used to calculate the AHRQ SES index variable.

conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in this measure, including diagnosis and procedure codes and other elements that are consequential to payment. Specifically, CMS works with Zone Program Integrity Contractors, and formerly Program Safeguard Contractors, to ensure program integrity; the agency also uses Recovery Audit Contractors to identify and correct for underpayments and overpayments.

CMS also uses the Comprehensive Error Rate Testing (CERT) Program to ensure that Medicare payments are correct in accordance with coverage, coding, and billing rules. Between 2005 and 2019, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3% to 96.4% of total payments each year.[38] The fiscal year 2020 Medicare fee-for-service program proper payment rate was 93.7%.[39] CMS continues to perform successful corrective actions and give providers additional education to ensure accurate billing.

To ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with a three-month claims run-out from the end of the measurement period.

Measure Reliability

Measure reliability is the degree to which repeated measurements of the same entity agree with each other. For measures of clinician performance, the measured entity is the TIN or TIN-NPI, and reliability is the extent to which repeated measurements of the TIN or TIN-NPI give similar results. To estimate measure reliability, we used a signal-to-noise analysis.

This approach seeks to determine the extent to which variation in the measure is due to true, underlying clinician performance, rather than random variation (i.e., statistical noise) within clinicians due to the sample of cases observed. To achieve this, we calculate reliability scores as:

$$R_j = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{w_j}^2}$$

Where:

$\sigma_{w_j}^2$ is the within-group variance of the mean measure score of clinician *j*

σ_b^2 is the between-group variance of clinicians within the episode group

That is, reliability is calculated as the ratio of between-group variance to the sum of between-group variance and within-group variance. Reliability closer to a value of one indicates that the between-group variance is relatively large compared to the within-group variance, which suggests that the measure is effectively capturing the systematic differences between the clinician and their peer cohort.

[38]Comprehensive Error Rate Testing (CERT) Program. “Appendices Medicare Fee-for-Service 2020 Improper Payments Report”. Table A6. https://www.cms.gov/files/document/2020-medicare-fee-service-supplemental-improper-payment-data.pdf-1.
[39]Ibid.

3.2.3 Statistical Results from Reliability Testing

Measure Reliability

At the proposed case minimum of 20 episodes, the mean reliability for TINs is 0.56 and for TIN-NPIs 0.46. The majority of TINs and TIN-NPIs have a mean reliability equal to or greater than 0.4; specifically, 91.6% of TINs and 76.1% of TIN-NPIs meet or exceed this threshold.

3.2.4 Interpretation

Measure Reliability

The mean reliability of the Colon and Rectal Resection measure exceeds 0.4 at a case minimum of 20 episodes or more for both TINs and TIN-NPIs due to the large number of episodes attributed to clinicians. CMS generally considers 0.4 as the threshold indicating 'moderate' reliability, which is supported by previous work into reliability and the threshold was finalized in the CY 2017 Quality Payment Program final rule.[40,41] See the CY2021 Physician Fee Schedule (PFS) proposed rule for further discussion of measure reliability.

3.3 Validity Testing

3.3.1 Level of Validity Testing

Our performance measure score validity testing included systematic assessment of both face validity and empirical validity testing.

3.3.2 Method of Validity Testing

Face Validity

The Colon and Rectal Resection measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. Experts in this clinical area evaluated specifications to ensure that each aspect of the measure (e.g., assigned services) was intentionally capturing only the costs of care within the reasonable influence of the attributed clinician for a defined patient population (i.e., the ability of the measure score to differentiate good from poor performance).

In developing this measure, Acumen incorporated input from the:

(i) a General and Colorectal Surgery Clinical Subcommittee;
(ii) a Colon and Rectal Resection Clinician Expert Workgroup;
(iii) a Technical Expert Panel (TEP); and
(iv) the Person and Family Partners.

This process is detailed in the Episode-Based Cost Measures Development Process document posted on the MACRA Feedback Page.[42]

One of the key roles of the measure-specific Clinician Expert Workgroup was to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in performing colon and rectal resections, thus limiting cost variation unrelated to clinician care for this measure. Assigned services occurring in the emergency department, outpatient facility and clinician services, inpatient medical, inpatient surgical, inpatient rehabilitation facility, and home health settings were defined separately for the pre- and post-

[40]Mathematica, Inc., "Memorandum: Reporting Period and Reliability of AHRQ, CMS 30-Day and HAC Quality Measures – Revised," http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf.
[41]CMS, "CY 2017 Quality Payment Program final rule," 81 FR 77169-77170.
[42]CMS, "Episode-Based Cost Measure Field Testing Wave 3 Measure Development Process," MACRA Feedback Page, https://www.cms.gov/files/document/macra-cmft-ebcm-process-2020.pdf.

trigger windows, and include colon and rectal resection evaluation, testing, treatment, complications, and follow-up.

Empirical Validity Testing

We evaluated the empirical validity of the Colon and Rectal Resection measure by examining correlation with known indicators of resource or service utilization based on a literature review, specifically complications related to colon and rectal resections. For this analysis, we compared the ratio of observed over expected cost (O/E cost ratios) at the provider level for Colon and Rectal Resection episodes with and without complications occurring in the post-trigger period. This analysis sought to confirm the expectation that the Colon and Rectal Resection measure captures variation in service utilization as an indicator of clinician cost performance. We would expect episodes with post-trigger complications, including downstream acute readmissions and post-acute care in an inpatient rehabilitation facility (IRF), skilled nursing facility (SNF), home health (HH), or long-term care hospital (LTCH), would have higher O/E cost ratios since complications like these should yield higher cost, even after accounting for patient clinical characteristics via risk adjustment. Conversely, episodes without these downstream costs should have lower O/E cost ratios, demonstrating that the measure can differentiate good from poor cost performance.

3.3.3 Statistical Results from Validity Testing

Table 2 below presents the results from the first analysis of validity. The mean O/E ratio for all episodes is 1.01. For each type of costly downstream service examined (i.e., acute readmission, post-acute care, and services related to colon and rectal resection complications), we see that there is a higher mean O/E for episodes with those services present than episodes without. For example, the mean O/E ratio for episodes with services relating to colon and rectal resection complications during the post-trigger period is 1.31, compared with 0.98 for episodes without services relating to those complications during the post-trigger period.

Table 2: Distribution of Observed to Expected Ratios

Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
All Final Episodes	1.01	0.34	0.58	0.68	0.73	0.82	0.93	1.09	1.38	1.66	2.35
Episodes with Downstream Acute Readmission	1.45	0.53	0.72	0.86	0.95	1.10	1.33	1.66	2.12	2.52	3.36
Episodes without Downstream Acute Readmission	0.97	0.27	0.57	0.68	0.73	0.81	0.91	1.05	1.26	1.48	2.01
Episodes with Post-Acute Care (IRF, LTCH, HH, SNF)	1.08	0.40	0.58	0.68	0.74	0.82	0.97	1.22	1.59	1.87	2.59
Episodes without Post-Acute Care (IRF,LTCH, HH, SNF)	0.93	0.20	0.58	0.69	0.73	0.81	0.90	1.01	1.13	1.24	1.60

Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
Episodes with Services Related to Colon and Rectal Resection Complications	1.31	0.54	0.61	0.72	0.79	0.92	1.18	1.56	1.98	2.38	3.24
Episodes without Services Related to Colon and Rectal Resection Complications	0.98	0.29	0.58	0.68	0.73	0.81	0.92	1.06	1.29	1.52	2.09

3.3.4 Interpretation

As expected, the average O/E cost ratios for episodes with costly downstream services (i.e., downstream acute readmission, post-acute care, and services related to colon and rectal resection complications) are higher than for episodes without downstream complications. This result demonstrates that the Colon and Rectal Resection measure is able to accurately capture higher resource use, and suggests that episodes with complications (the frequency or severity of which could be reasonably expected to be influenced by the treatment of the attributed clinician) will yield higher costs, even after risk adjustment.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Colon and Rectal Resection measure to ensure a comparable patient population within the scope of the measure focus on colon or rectal resections for either benign or malignant indications, and that episodes provide meaningful information to attributed clinicians. Exclusions are also used as part of data processing so that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode. For the exclusions analysis discussed in this section, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- Episodes where a patient's death date occurred before the episode end date.
    - These episodes were excluded as they may not accurately reflect a clinician's performance as the truncated episode window does not capture the full length of care intended by the measure.
- Episodes where patients have a LVAD
    - Patients with LVAD have different care needs. These episodes were excluded because of the very small number of patients who receive an LVAD.
- Episodes where a patient has had a recent major bowel surgery
    - These episodes were excluded because the cohort of patients who undergo a major bowel surgery shortly before the trigger event may indicate that the procedure is staged. Staged procedures as a result of complications and emergent procedures that require staging may have different costs associated with them. However, the costs of subsequent procedures are assigned to the episode if a procedure becomes staged as a result of complications.

- Episodes where a patient elects to leave against medical advice
    - Leaving against medical advice prevents the attributed clinician from completing appropriate care for a patient, which leaves a patient at high risk of further complications. Retaining such patients would put the attributed clinician at risk of being attributed a costly episode in which they did not have the chance to fully treat a patient.
- Episodes where the patient was transferred within 3 days prior to admission
    - These episodes were excluded to avoid incentivizing institutions to transfer or turn away more complex patients.
- Episodes classified as outlier cases.
    - To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1st percentile and above the 99th percentile are considered outliers and removed from measure calculation.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different risk profile than the included episodes, such as a higher mean cost or a different distribution of costs (e.g., a long tail of high-cost episodes). For the exclusions, we examined the number of episodes and patients affected, as well as the distributions of observed cost and O/E ratios (calculated by applying existing risk factor coefficients to the excluded episodes) for excluded episodes. We then compared the cost characteristics of the excluded episodes to those of final episodes included in measure calculation to assess the distinctness between the 2 patient cohorts. A full list of the exclusions used for the Colon and Rectal Resection measure is provided in the draft Measure Codes List available on the MACRA Feedback Page.[43]

3.4.2 Statistical Results from Testing Exclusions

Table 3 below presents observed cost statistics and O/E cost ratios for the Colon and Rectal Resection measure exclusions. Cost statistics are also provided for the set of final episodes included in the Colon and Rectal Resection measure for comparison, with a case minimum of 20 episodes at the TIN and TIN-NPI levels. For the standard exclusions in the table below, such as the trigger claim not occurring in an inpatient prospective payment system, or IPPS, acute hospital or psychiatric facility, the trigger claim was not performed in an OP, IP, ASC setting, and where there is a concurrent inpatient stay with the trigger, the inpatient stay does not have a related MS-DRG (i.e., MS-DRGs 329-334), these patient cohorts were excluded from the measure to assess episodes in the intended setting and by the measure's intended attribution approach.

Table 3: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost			O/E Cost Ratio		
			Mean	Percentile			Percentile	
	#	%		10th	90th	%	10th	90th
All Episodes Meeting Triggering Logic	61,524	100.00%	$27,376	$12,691	$48,398	1.01	0.71	1.42
Patient Death in Episode	4,380	7.12%	$43,045	$20,545	$68,430	1.04	0.66	1.61
Not an IPPS Acute Hospital	984	1.60%	$26,932	$12,608	$48,049	1.08	0.72	1.62
Not in OP, IP, or ASC Setting	57	0.09%	$35,587	$14,846	$63,091	1.13	0.62	1.91

[43]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

Exclusion	Episodes		Observed Cost			O/E Cost Ratio		
			Mean	Percentile			Percentile	
	#	%		10th	90th	%	10th	90th
No relevant MS-DRG in Episode with Concurrent IP Stay	16,069	0.00%	$11,024	$2,519	$24,118	0.59	0.15	1.34
Transfer within 3 days prior to IP admission	860	1.40%	$47,115	$20,903	$76,581	1.11	0.67	1.84
Leaving AMA	56	0.09%	$24,785	$12,785	$37,162	0.91	0.69	1.15
Recent Major Bowel Surgery	942	1.53%	$33,894	$14,352	$59,290	0.98	0.61	1.52
LVAD	25	0.04%	$34,503	$15,557	$56,851	0.91	0.58	1.16
Outlier cases	1,090	1.77%	$71,384	$17,841	$144,219	1.78	0.40	3.66
Final Episodes (TIN)	12,263	19.93%	$21,968	$12,502	$37,172	0.98	0.72	1.30
Final Episodes (TIN-NPI)	37,878	61.57%	$24,269	$12,604	$42,485	0.99	0.72	1.34

*This table does not include all measure exclusions.

3.4.3 Interpretation

The statistical results indicate that the majority of excluded episodes are different than the final set of episodes at the TIN and TIN-NPI levels of testing. The excluded episodes differ in the mean observed cost (e.g., higher observed cost) or in the mean O/E cost ratio and O/E cost ratio distribution. These results support the exclusion of these episodes to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for exclusions applied based on the clinical validity of the study population are provided below.

Episodes ending in death: Episodes ending in death have a higher mean observed cost ($43,045) compared to the final set of episodes at the TIN level testing ($21,968) and at the TIN-NPI level testing ($24,269). The mean O/E cost ratio for these episodes (1.04) is slightly higher than the mean O/E cost ratio for final episodes at both the TIN-level (0.98) and TIN-NPI level (0.99) of testing. Finally, the O/E cost ratio ranges from 0.66 at the 10th percentile to 1.61 at the 90th percentile for episodes ending in death, compared to 0.72 at the 10th percentile and 1.30 at the 90th percentile for final episodes at TIN level and 0.72 at the 10th percentile and 1.34 at the 90th percentile for final episodes TIN-NPI level, respectively. These results indicate that this patient cohort is distinct in both observed cost and risk profile, and excluding these episodes ensures a fairer cost comparison.

Episodes where patients have an LVAD: Episodes where patients that have an LVAD are clinically different than episodes where patients do not have an LVAD, reflected in the different cost profile of episodes, where the distribution of O/E cost ratio ranges from 0.58 at the 10th percentile to 1.16 at the 90th percentile. There is a very small number of episodes (i.e., 25) with this exclusion.

Episodes where a patient had a recent major bowel surgery: As expected, based on the clinical justification for the exclusion, costs for episodes where the patient had a recent major bowel surgery are substantially higher (mean observed cost of $33,894) in comparison to the final set of episodes. While the mean O/E cost ratio for these episodes is similar (0.98) to the mean O/E cost ratio for final episodes at both TIN (0.98) and TIN-NPI (0.99) levels of testing, the range is more broad, with the O/E cost ratio ranging from 0.61 at the 10th percentile to 1.52 at the 90th percentile (0.72 at the 10th percentile and 1.34 at the 90th percentile for final episodes at TIN level and 0.72 at the 10th percentile and 1.30 at the 90th percentile for final episodes TIN-NPI level). The higher observed costs and higher O/E across the distribution suggest that patients who had a recent major bowel surgery have a different cost profile from other patients.

Episodes where a patient elects to leave against medical advice: This measure is intended to incentivize clinicians to change their behavior and treatment patterns to increase cost-effectiveness. However, the ability of the measure to accurately reflect such improvements is limited if attributed clinicians are held accountable for patients who do not take advantage of the offered care. Although the mean observed costs are similar to the final set of episodes, the mean O/E cost ratio (0.91) is lower than the final set of episodes and the O/E cost ratios range from 0.69 at the 10th percentile to 1.15 at the 90th percentile. These patients are excluded to allow the measure to capture the outcomes of clinicians' decisions.

Episodes where a patient was transferred within 3 days prior to the admission: As expected, these episodes have substantially higher mean observed episode costs ($47,115) compared to the final set of episodes. Additionally, the O/E cost ratios for these episodes range from 0.67 at the 10th percentile to 1.84 at the 90th percentile, a considerably wider range compared to 0.72 at the 10th percentile and 1.34 at the 90th percentile for final episodes at TIN level and 0.72 at the 10th percentile and 1.30 at the 90th percentile for final episodes TIN-NPI level. These results suggest that patients who are transferred within 3 days prior to the admission have both higher mean episode costs and greater variation in risk-adjusted cost. These results, alongside the intention to avoid incentivizing institutions to transfer or deny care to riskier or more complex patients, support exclusion. This exclusion is also in line with other bundled payment programs.

Episodes classified as outlier cases: The mean observed cost of these episodes is around three times greater than for the final set of episodes (e.g., $71,384 compared to $21,968 at the TIN-level and $24,269 at the TIN-NPI-level). The mean O/E cost ratio is substantially higher (1.78) compared to both TIN (0.98) and TIN-NPI (0.99) levels of testing. In addition, the O/E cost ratio for outlier cases ranges from 0.40 at the 10th percentile to 3.66 at the 90th percentile, indicating that the risk adjustment model is currently unable to account for the patient characteristics associated with these high- and low-cost outlier episodes. Excluding outliers based on risk-adjusted cost eliminates the episodes that deviate most from expected spending levels based on patient characteristics.

3.5 Risk Adjustment or Stratification

3.5.1 Method of Controlling for Differences

Differences in case mix are controlled for using a statistical risk model with 145 risk factors and stratification by 2 risk categories.

The risk adjustment model for the Colon and Rectal Resection measure broadly follows the CMS-HCC risk adjustment methodology, which is derived from Medicare Parts A and B claims and is used in the Medicare Advantage (MA) program. Patient age is included as one of 12 age categorical variables derived from the MA risk adjustment model's age/sex variables. Severity of illness is measured using HCCs, indicators of enrollment and long-term care status, and disease interactions. The risk adjustment model also includes variables for factors identified by the expert clinician workgroup as affecting resource use.

The model includes 79 HCC indicators derived from the patient's Parts A and B claims during the period 120 days prior to the episode trigger and are specified in the CMS-HCC Version 22 (V22) 2016 model. Episodes for patients without a full 120-day lookback period are excluded from the measure. This 120-day period is used to measure patient health status and ensures that each patient's claims record contains sufficient fee-for-service data both for measuring spending levels and for risk adjustment purposes.

In addition, the risk adjustment model includes status indicator variables for whether the patient qualifies for Medicare through Disability or ESRD. The model also includes an indicator of whether the patient recently required long-term care, defined as 90 days in a long-term care

facility without being discharged to community for 14 days. Patients who need to reside in long-term care facilities typically require more intensive care than patients who live in the community. These enrollment and long-term care status variables are non-diagnostic indicators of severity of illness.

The model also accounts for disease interactions between HCCs and/or enrollment status variables included in the MA model. These interactions are included because certain combinations of comorbidities increase costs more than is predicted by the HCC indicators alone.

Furthermore, the risk adjustment model includes measure-specific factors intended to further isolate costs that attributed clinicians can reasonably influence, informed by expert clinician input and empirical analyses. The following variables were added to avoid potential unintended consequences:

- Whether the patient had recent chemotherapy or radiation.
- Blood transfusion receipt during hospitalization.
- Whether the patient had an ostomy performed.
- Whether the patient had anemia or secondary anemia.
- Whether the patient had dementia.
- Whether the patient had a smoking or nicotine dependence.
- Whether the patient was previously ventilator dependent.
- Whether the patient had a partial or total laparoscopic colectomy.
- Whether the patient had a recent percutaneous coronary intervention (PCI) or myocardial infarction (MI).
- Whether the patient had a concurrent major abdominal surgery or a recent major abdominal surgery (non-bowel).
- Whether the patient had antiplatelet or anticoagulant use or received a blood transfusion during a hospitalization.
- Whether the patient had cardiomyopathy, valve disease, or recent cardiac arrest.
- Whether the patient had home oxygen or recently received HH services.
- Whether the patient had portal hypertension or pulmonary hypertension.
- Whether the patient had metastatic disease.
- Whether the patient had a recent all-cause admission in 30 days prior to trigger day or a recent all-cause admission in the 120 days prior to the trigger day.
- Whether the patient had an emergent colectomy or inflammatory bowel disease (IBD).
- Whether the patient had a recent admission to a SNF, IRF, or a LTCH.
- Whether the patient had a rectal prolapse.

As with the CMS-HCC model, the risk adjustment approach for this measure uses an ordinary least squares linear regression model. The predicted, or expected, cost is winsorized at 0.5th percentile to make sure episodes with unusually small predicted cost, which would lead to abnormally large O/E cost ratios, do not dominate certain clinicians' final score. The winsorized expected costs are renormalized to ensure the average expected episode cost is the same before and after winsorizing. Then, as presented in the exclusions analysis above, extremely low- or high-cost outlier episodes with residuals below the 1st percentile or above the 99th percentile are excluded to reduce the effect of episodes that deviate the most from their expected values in absolute terms. The expected cost after excluding these outliers is again renormalized to ensure that average expected costs are the same after outlier removal.

Finally, the risk adjustment model outlined above is stratified for each of the 2 Colon and Rectal Resection measure sub-groups, which are based on the type of procedure, below:

- Colon Resection
- Rectal Resection

Full details of the risk adjustment model are in the Measure Codes List File available on the MACRA Feedback Page.[44]

3.5.2 Conceptual, Clinical, and Statistical Methods

We selected the CMS-HCC model based on previous studies evaluating its appropriateness for use in risk adjusting Medicare claims data. This model was developed specifically for use in the Medicare population, meaning that it accounts for conditions found in the Medicare population and is calibrated on Medicare fee-for-service beneficiaries. In addition, the CMS-HCC model is routinely updated for changes in coding practices (e.g., the transition from the International Classification of Diseases [ICD]-9 to ICD-10 codes) and is exhaustive on these code sets. Because the CMS-HCC model has already been extensively tested, we focus our testing on how the CMS-HCC model was adapted to the Colon and Rectal Resection measure methodology.

The workgroup provided input on measure-specific risk adjustors after reviewing empirical analyses on subpopulations of interest to assess whether and if so, how, particular factors should be accounted for in the model. These could include patient characteristics, factors outside of the reasonable influence of the clinician, or any other factors that would help prevent unintended consequences. These additional risk adjustors are listed in the section above.

As previously noted, the risk adjustment model is run on episodes stratified into episode sub-groups, which may qualify as "ordering" of risk factors. Episode sub-groups were also determined based on the workgroup's input, with the goal of ensuring clinical comparability among episodes so that the cost measure fairly compares clinicians with similar patient case-mix. The episode sub-groups are listed in the above section. The Rectal Resection episode sub-group includes episodes that: (i) are triggered by a rectal procedure code, (ii) have a resection procedure trigger code that is accompanied by a rectal or anal cancer diagnosis code, or (iii) are triggered by a lower anterior resection (LAR) (i.e., CPT/HCPCS codes 44145, 44146, 44207, and 44208) when accompanied by an ICD-10 rectal cancer diagnosis code (i.e., C20). However, rectopexies (i.e., CPT/HCPCS code 45400, and 45402) are classified into the Colon Resection episode sub-group to ensure that the rectal procedures captured in the Rectal Resection episode sub-group are of similar complexity and entail comparable risks within the patient cohort. The Colon Resection episode sub-group captures all other cases triggered by the trigger codes not included in the definition for the Rectal Resection episode sub-group.

3.5.3 Conceptual Model of Impact of Social Risks

Our conceptual model of the impact of social risk factors is informed by both published external research and our own data analysis.[45,46,47]

---

[44]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.
[45]Assistant Secretary of Health and Human Services for Planning and Evaluation. Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs. Washington, D.C. December 2016.
[46]Chen LM, Epstein AM, Orav EJ, Filice CE, Samson LW, Joynt Maddox KE. Association of Practice-Level Social and Medical Risk With Performance in the Medicare Physician Value-Based Payment Modifier Program. JAMA. 2017;318(5):453-461.
[47]Medicare Payment Advisory Commission. Beneficiaries Dually Eligible for Medicare and Medicaid. 2018; https://www.macpac.gov/publication/data-book-beneficiaries-dually-eligible-for-medicare-and-medicaid-3/.

3.5.4 Statistical Results

The literature has extensively tested the use of the HCC model as applied to Medicare claims data. Although the variables in the HCC model were chosen to predict annual cost, CMS has also used this risk adjustment model in a number of other settings (e.g., accountable care organizations, or ACOs, previous physician Quality and Resource Use Reports, or QRUR programs, and other measures such as NQF #3512: Knee Arthroplasty, NQF #3509: Routine Cataract Removal with Intraocular Lens (IOL) Implantation, NQF #3510: Screening/Surveillance Colonoscopy, and NQF #2158: MSPB-Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Evaluation of the CMS-HCC Risk-Adjustment Model report[48] and the Report to Congress: Risk Adjustment in Medicare Advantage.[49] For measure-specific factors not included in the CMS-HCC model, we sought expert clinician input through the workgroup, which provided recommendations on additional risk adjustors and measure sub-groups.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

Acumen analyzed sex dual status, income, education, and unemployment as social risk factors (more information on these variables can be found in Section 3.1.7). Patient gender and dual status were obtained from the EDB and CME. Information on income, education, and unemployment was obtained from ACS data and linked to episodes by census block group, where possible, to provide a more granular level of analysis than ZIP code. Patients without geographic information necessary to obtain ACS data were excluded, representing less than 1.5% of episodes.

The percentage of female patients range from 46.9% to 60.5% across the 2 measure sub-groups in this measure. The majority of the patients (81.9% - 85.1%) have non-dual status. Income level is categorized into high, medium, and low from the continuous average income variable in ACS; therefore, each category has 33% of observations. While 1.8% to 2.1% of patients are classified below a high school education level, the overwhelming majority of episodes are classified at a high school level or greater. Finally, 17.0% to 18.6% of patients have high unemployment designation (>10%).

Acumen examined the impact of including social risk factors into our risk adjustment model by running goodness of fit tests when different risk factors are added and compared to the base risk adjustment model, where the base risk adjustment model refers to the full standard set of risk adjustment variables from the CMS-HCC V22 2016 model, disability status, ESRD status, interaction variables, recent long-term care use, and measure-specific clinical risk adjustors. Acumen ran a step-wise regression to include the following additional social risk factors on top of the adapted CMS-HCC model:

- Sex
- Dual status
- Sex + dual status
- Sex + dual status + race
- Sex + dual status + income + education + unemployment
- Sex + dual status + AHRQ SES index score
- Sex + dual status + race + income + education + unemployment

[48]Pope, Gregory C., John Kautter, et al., “Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report.” RTI International: March 2011.
[49]CMS, “Report to Congress: Risk Adjustment in Medicare Advantage,” https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf.

- Sex + dual status + race + AHRQ SES index score

The step-wise regressions help evaluate individual as well as joint significance of the social risk factors. We examined the impact of including social risk factors into our risk adjustment model with T-test of individual significance and F-test of joint significance.

We analyzed the correlation between measure scores calculated with and without the social risk factors. The measure scores calculated with and without these social factors were highly correlated at both the TIN level and the TIN-NPI level, with a Spearman correlation coefficient of 0.99. These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

3.5.6 Method for Statistical Model or Stratification Development

To analyze the validity of the current risk adjustment model, we examined 2 analyses: (1) R-squared and adjusted R-squared for the regression models, and (2) predictive ratios and O/E cost ratios to examine the fit of the models at different levels of patient complexity.

1) R-squared and adjusted R-squared were calculated for the measure. The results should be evaluated in the context of the measure's service assignment rules, which are intended to ensure only clinically associated services are grouped to the episodes. This is an important distinction from all-cost measures, as service assignment leaves less variation for the risk adjustment model to explain. In this context, a low R-squared may indicate the effectiveness of the service assignment rules. These results are provided in Section 3.5.7.
2) Predictive ratios and O/E cost ratios were calculated for each "risk decile" for the episode group. A "risk decile" is based on the risk scores, which indicate how costly episodes are expected to be, as predicted through risk adjustment. After arranging episodes into deciles based on their risk score, we calculated the predictive ratios and average O/E cost ratios for each decile. The predictive ratio aims to examine the fit of the model at different levels of patient complexity to examine the model's ability to predict both very low and high cost episodes, and is calculated using the formula of average (expected cost)/average (observed cost) for all episodes in each decile. Similarly, the O/E cost ratio demonstrates the model's prediction accuracy, and is calculated using the formula of average (observed cost/expected cost) for all episodes in each decile. These are discussed in Sections 3.5.8 and 3.5.9. Statistical Risk Model Discrimination Statistics

The overall R-squared for the Colon and Rectal Resection cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.47. The adjusted R-squared is 0.47. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.[50]

3.5.7 Statistical Risk Model Calibration Statistics

We interpret calibration as how accurately the risk model's predictions match the actual episode cost. We calculate the average O/E cost ratio for each risk decile to demonstrate the model's prediction accuracy. The O/E ratio was between 0.97-1.04 for risk deciles 2 through 10. The O/E ratio is 1.14 for the first decile, showing that the measure is slightly under-predicting costs for less complex patients. This indicates that the risk adjustment model is generally predicting actual episode costs accurately across most risk deciles.

[50]Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

3.5.8 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows that the model generally has consistent predictive ratios across risk score deciles. The predictive ratios range from 0.96 to 1.04 for risk deciles 2 through 10. The first risk decile has a predictive ratio of 0.88, where the risk adjustment model is under predicting expected episode costs. Full results are available in the National Summary Data Report (NSDR) addendum on the MACRA Feedback Page.[51]

3.5.9 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are higher than the values presented in similar analyses of risk adjustment models.[52] As noted in Section 3.5.6, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services, so the resulting variation is reflective of variation related to factors within a clinician's reasonable influence.

As demonstrated in Section 3.5.8 and 3.5.9, the average O/E cost ratios and the predictive ratios for most of the risk deciles are generally close to one. Predictive ratios close to one indicate that expected spending is accurately predicting observed spending. Overall, the results show that the model is accurately predicting observed spending, for most risk levels. Results also show that the risk adjustment model may be slightly under-predicting expected episode spending for less complex patients.

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

Our method of determining clinically meaningful differences in episode-based cost measure performance consists of stratifying clinician measure O/E cost ratios by meaningful characteristics and investigating the clinician O/E cost ratio distribution by percentile. The cost measure score numerator is the sum of the O/E cost ratio for all episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure. The denominator is the total number of episodes from the attributed to a clinician. Using O/E cost ratios allows for direct comparisons of performance at the episode sub-group level since a dollar figure cannot be calculated for those episodes using the national average observed episode cost. Stratification is performed for each of the following characteristics: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician or clinician group. We analyze the distribution of measure O/E cost ratios for clinicians defined by these characteristics.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure O/E cost ratios among clinicians to determine a meaningful difference in performance. In addition, this analysis looks to confirm that the measure behaves as expected with respect to meaningful clinician characteristics.

3.6.2 Statistical Results

Key findings show that, generally, there is a notable performance difference among clinicians in the Colon and Rectal Resection measure:

[51]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.
[52]Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

(i) The 99th percentile of the measure O/E cost ratio is almost 1.5 times the measure O/E cost ratio at the 1st percentile for both the TIN and TIN-NPI levels.
(ii) The Colon and Rectal Resection measure O/E cost ratio at the 90th percentile is approximately 18.28% greater than the O/E cost ratio at the 10th percentile at the TIN level and 19.57% greater at the TIN-NPI level.

These results indicate there is a meaningful potential for Medicare cost savings.

The results also show that there is a limited systemic regional difference in clinician score. For instance, the mean O/E cost ratios for clinicians across 9 census divisions (excluding 'Unknown') are within a 0.04 or less range (i.e., 0.99 to 1.03 at the TIN level and 0.98 to 1.02 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas.

In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform similarly to those who perform fewer procedures. We also analyzed clinicians by risk score decile, as variation by risk score decile could indicate that the risk adjustment model is over- or under-correcting for clinicians with systematically riskier patients. Measure O/E cost ratios also show little variation by risk score decile, with a range in mean TIN O/E cost ratio of 1.00 to 1.02 and a range in mean TIN-NPI O/E cost ratio of 0.99 to 1.04, indicating that the risk adjustment model is overall functioning as intended.

Tables 4-A and 4-B below present the distribution of cost measure O/E cost ratios by a range of clinician/clinician group characteristics, allowing a comparison of O/E cost ratio distributions for these breakdowns. The cost measure O/E cost ratios are presented at the TIN level and the TIN-NPI level.

Table 4-A: Colon and Rectal Resection TIN-Level Cost Measure O/E Ratios

Characteristic	# of TINs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
All TINs	759	1.01	0.87	0.93	1.00	1.10	1.21
Measure Sub-group							
Colon Resection	759	1.01	0.86	0.93	1.00	1.10	1.22
Rectal Resection	717	1.01	0.65	0.80	0.98	1.26	1.72
Urban/Rural							
Urban	697	1.01	0.88	0.93	1.00	1.10	1.21
Rural	62	0.98	0.82	0.90	0.98	1.08	1.17
Unknown	0	-	-	-	-	-	-
Census Region							
Northeast	114	1.02	0.89	0.95	1.01	1.10	1.17
Midwest	181	1.00	0.87	0.92	0.99	1.09	1.22
South	320	1.01	0.87	0.93	1.00	1.10	1.21
West	144	1.02	0.86	0.94	1.01	1.12	1.21
Unknown	0	-	-	-	-	-	-
Census Division							
New England	29	1.00	0.93	0.94	1.00	1.07	1.11
Middle Atlantic	85	1.03	0.85	0.95	1.02	1.11	1.17
East North Central	115	1.01	0.88	0.92	0.99	1.09	1.19
West North Central	66	0.99	0.83	0.92	0.99	1.06	1.22

Characteristic	# of TINs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
South Atlantic	180	1.01	0.87	0.93	1.00	1.10	1.21
East South Central	61	1.00	0.85	0.92	0.99	1.10	1.17
West South Central	79	1.00	0.86	0.92	0.99	1.10	1.27
Mountain	58	1.01	0.82	0.94	1.01	1.09	1.21
Pacific	86	1.03	0.86	0.94	1.02	1.12	1.35
Unknown	0	-	-	-	-	-	-
Provider risk score decile							
1st	75	1.02	0.89	0.93	1.01	1.13	1.22
2nd	76	1.01	0.85	0.94	1.00	1.09	1.21
3rd	76	1.01	0.86	0.93	1.00	1.09	1.35
4th	76	1.00	0.86	0.93	0.99	1.09	1.23
5th	76	1.02	0.88	0.94	1.01	1.10	1.18
6th	76	1.01	0.89	0.94	0.99	1.09	1.27
7th	76	1.00	0.83	0.92	0.98	1.11	1.18
8th	76	1.01	0.82	0.93	1.00	1.12	1.35
9th	76	1.00	0.87	0.92	1.00	1.09	1.22
10th	76	1.02	0.84	0.91	1.01	1.12	1.21
Number of episodes							
10-19 Episodes	0	-	-	-	-	-	-
20-39 Episodes	400	1.01	0.85	0.92	1.00	1.12	1.21
40-59 Episodes	157	1.01	0.89	0.93	0.99	1.09	1.22
60-79 Episodes	78	1.01	0.90	0.94	1.01	1.08	1.35
80-99 Episodes	58	1.00	0.91	0.94	0.99	1.06	1.10
100-199 Episodes	57	1.01	0.95	0.97	1.00	1.05	1.10
200-299 Episodes	8	1.01	0.97	0.97	1.01	1.06	1.06
300+ Episodes	1	0.98	0.98	0.98	0.98	0.98	0.98

Table 4-B: Colon and Rectal Resection TIN-NPI-Level Cost Measure O/E Ratios

Characteristic	# of TIN-NPIs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
All TIN-NPIs	502	1.01	0.87	0.92	1.00	1.10	1.22
Measure Sub-group							
Colon Resection	502	1.01	0.86	0.92	1.00	1.10	1.26
Rectal Resection	484	1.01	0.72	0.80	0.98	1.23	1.59
Urban/Rural							
Urban	480	1.01	0.87	0.92	1.00	1.10	1.22
Rural	22	1.00	0.87	0.90	0.98	1.12	1.32
Unknown	0	-	-	-	-	-	-
Census Region							
Northeast	86	1.00	0.84	0.92	0.99	1.10	1.36
Midwest	100	0.99	0.88	0.92	0.99	1.06	1.17

Characteristic	# of TIN-NPIs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
South	232	1.01	0.88	0.92	1.00	1.12	1.25
West	83	1.01	0.87	0.92	1.00	1.11	1.21
Unknown	1	1.00	1.00	1.00	1.00	1.00	1.00
Census Division							
New England	28	0.98	0.91	0.92	0.98	1.04	1.08
Middle Atlantic	58	1.01	0.84	0.91	1.00	1.14	1.36
East North Central	54	0.99	0.88	0.92	0.99	1.06	1.15
West North Central	46	0.99	0.89	0.92	0.99	1.09	1.18
South Atlantic	134	1.02	0.89	0.93	1.02	1.12	1.19
East South Central	38	1.00	0.88	0.92	0.98	1.12	1.32
West South Central	60	0.99	0.87	0.90	0.98	1.09	1.32
Mountain	36	1.01	0.87	0.89	1.00	1.11	1.17
Pacific	47	1.01	0.90	0.93	1.00	1.14	1.21
Unknown	1	1.00	1.00	1.00	1.00	1.00	1.00
Provider risk score decile							
1st	50	1.01	0.90	0.95	1.00	1.09	1.17
2nd	50	1.02	0.87	0.93	1.01	1.11	1.32
3rd	50	1.00	0.87	0.91	0.99	1.09	1.15
4th	51	1.04	0.88	0.94	1.02	1.13	1.32
5th	50	1.00	0.91	0.93	0.98	1.08	1.36
6th	50	1.00	0.88	0.93	0.99	1.10	1.13
7th	51	1.00	0.86	0.92	0.97	1.13	1.23
8th	50	1.00	0.88	0.91	0.99	1.11	1.25
9th	50	0.99	0.85	0.88	0.99	1.09	1.18
10th	50	1.01	0.84	0.92	1.01	1.09	1.19
Number of episodes							
10-19 Episodes	0	-	-	-	-	--	-
20-39 Episodes	456	1.01	0.87	0.92	1.00	1.11	1.23
40-59 Episodes	42	1.00	0.90	0.94	0.99	1.07	1.12
60-79 Episodes	3	1.03	1.01	1.01	1.02	1.05	1.05
80-99 Episodes	1	1.05	1.05	1.05	1.05	1.05	1.05
100-199 Episodes	0	-	-	-	-	-	-
200-299 Episodes	0	-	-	-	-	-	-
300+ Episodes	0	-	-	-	-	-	-

3.6.3 Interpretation

The results in Tables 4-A and 4-B above indicate that there is no notable variation in the mean cost measure O/E cost ratio across episode sub-groups, the urban/rural divide, census regions, census divisions, provider risk score decile, or episode volume at both the TIN and TIN-NPI levels. For each variable, the largest difference in the mean O/E cost ratio within each category was 0.05 or less. This indicates that the risk adjustment model is overall functioning as intended; it is adjusting cost performance such that there are no substantive differences across the categories for these variables. For episode sub-groups, the model is run separately for each episode sub-group to account for a more fair comparison across episodes in the Colon

Resection episode sub-group and Rectal Resection episode sub-group. These results support that there is meaningful variation in cost performance, even after risk adjustment, across these variables. These results also indicate that there is a meaningful potential for Medicare savings and that there are no systemic differences across geographic region, level of provider risk, and case volume.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the Colon and Rectal Resection measure, Acumen expects a high degree of data completeness. To further ensure that we have complete and accurate data for each patient who opens an episode, Acumen excludes episodes where patient date of birth information (an input to the risk adjustment model) cannot be found in the EDB, the patient does not appear in the EDB, or the patient death date occurs before the episode trigger date.

The Colon and Rectal Resection measure excludes episodes where the patient is enrolled in Medicare Part C or has a primary payer other than Medicare in the 120-day lookback period and episode window. In such situations, Medicare Parts A and B claims data may not capture the complete clinical profile for the patient needed to capture the clinical risk of the patient in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the patient's care is covered under Medicare Part C. Additionally, the Colon and Rectal Resection measure excludes episodes that cannot be attributed to a main surgeon.

3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the 5 categories of missing data which caused episodes to be excluded from the Colon and Rectal Resection measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the number of TINs and TIN-NPIs who had at least one episode excluded due to missing data. The missing data categories are:

- Patient death date occurred before the trigger date.
- Patient has a primary payer other than Medicare during the episode window or in the 120-day lookback period.
- Patient was not enrolled in Medicare Parts A and B, or was enrolled in Part C, during the 120-day lookback period and episode window.
- The episode cannot be attributed to a main surgeon.

Table 5: Missing Data Categories for the Colon and Rectal Resection Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Death before trigger	96	91	113
Other primary payer	7,112	2,271	5,879
Not continuously enrolled	4,470	1,897	4,178
No main surgeon	468	345	419

3.7.3 Interpretation

As the Colon and Rectal Resection measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency of missing data as noted above. Acumen takes measures to ensure that missing or inaccurate information in claims data is not included in the cost measure.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

The data elements used in this measure are generated, collected and/or used by healthcare personnel during the provision of care (e.g., blood pressure, laboratory values, diagnosis, depression score). The data collected during care provision are then translated into the appropriate coding system (e.g. ICD-10 diagnoses, MS-DRGs) for use in Medicare claims.

4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

4.3 Data Collection Strategy

4.3.1 Data Collection Strategy Difficulties

Lessons and associated modifications may be categorized into three types: data collection procedures, handling of missing data, and sampling data associated with beneficiaries who died during an episode of care.

4.3.1.1 Data Collection

Acumen receives claims data directly from the Common Working File (CWF) maintained at the CMS Baltimore Data Center. Medicare claims are submitted by healthcare providers to a Medicare Administrative Contractor (MAC), and are subsequently added to the CWF. However, these claims may be denied or disputed by the MAC, leading to changes to historical CWF data. In rare circumstances, finalizing claims may take many months, or even years. As a result, it is not practical to wait until all claims for a given month are finalized before calculating this measure. As such, there is a trade-off between efficiency (accessing the data in a timely manner) and accuracy (waiting until most claims are finalized) when determining the length of the time (i.e., the “claims run-out” period) after which to pull claims data. To determine the appropriate claims run-out period, Acumen has performed testing on the delay between claim service dates and claims data finalization. Based on this analysis, Acumen uses a run-out period of three months after the end of the calendar year to collect data for development and testing purposes. If this measure is used in a CMS program, calculation and reporting would be done in line with that program’s reporting practices.

4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of episodes with missing data are excluded to ensure completeness of data and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 120 days prior to the episode start date are not included in this measure. This enables the risk adjustment model to accurately adjust for the beneficiary’s comorbidities using data from the previous 120 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary’s date of birth cannot be located are not included in this measure.

4.3.1.3 Sampling

During measure testing, Acumen noted that episodes in which the beneficiary died prior to the episode end date exhibited different cost distributions compared to other episodes. To avoid this potential impact on clinician scores, this measure does not include episodes for which the beneficiary’s date of death occurs prior to the end of the episode window.

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

The measure was developed for use in MIPS, under a contract with CMS.

5.1.2 Feedback on the Measure and Development Process

5.1.2.1 Technical Assistance Provided During Development or Implementation

Development: Field Testing

Acumen and CMS conducted a national field test of 5 episode-based cost measures developed in 2019 and 2020, including the Colon and Rectal Resection measure, for a 5-week comment period (August 17 to September 18, 2020). We provided a Field Test Report to a sample of clinician groups and clinicians.[53] Field Test Reports were provided for each measure that a clinician or clinician group was attributed 10 or more acute inpatient medical condition and procedural episodes or 20 chronic condition episodes. This testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many stakeholders as possible. This sampling technique was used for field testing only and is not indicative of the case minimums used for any potential program implementation.

All stakeholders, including those who did not qualify to receive a Field Test Report, could review a series of mock reports that were representative of each measure and reporting type. Other public documentation posted during field testing included: measure specifications for each measure (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Measure Development Process document, a Frequently Asked Questions document, and a Fact Sheet.[54] During field testing, Acumen conducted education and outreach activities for stakeholders including multiple office hours sessions with specialty societies, a publicly posted field testing webinar recording, and Quality Payment Program Help Desk support.

5.1.2.2 Technical Assistance with Results

Field Testing

During the feedback period, 1,558 Field Test Reports for episode-based cost measures were downloaded by 1,013 clinician groups (TINs) and 545 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the Clinician Expert Workgroup to consider in recommending refinements to the measures based on the testing data and feedback.

The following sections offer more details on the contents of each report and describe the education and outreach efforts associated with the field testing feedback period.

Data Provided During Field Testing

Each Field Test Report contained:

- Detailed performance results for the attributed measure, including cost measure score and breakdown of episode cost compared to the national average and TIN/TIN-NPIs with a similar patient case mix (or risk profile).
- Drill-down detail for each measure, including more detailed information on potential cost drivers in the TIN/TIN-NPI's episodes. For example:

[53]The field test reports were available for download from the Quality Payment Program website: https://qpp.cms.gov/login.

[54]The Measure Development Process, Frequently Asked Questions, and Fact Sheet documents are posted on the MACRA Feedback Page: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html.

- Analysis of utilization and cost for the measure by specific service categories (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)
- Breakdown of costs for Physician/Supplier Part B and inpatient claims (e.g., top 5 most billed services and by risk bracket)
- Accompanying episode-level Comma Separated Value (CSV) file with detailed information for all episodes attributed to the TIN/TIN-NPI. This file provides detailed information on every episode used to calculate your measure score, which includes winsorized observed cost, risk-adjusted cost, facilities and clinicians rendering care, the share of cost by service setting, the patient relationship code (PRC) on the trigger/reaffirming claim line.

Mock Field Test Reports for each measure type that was field tested in 2020 were available for download by eligible clinicians and clinician groups from the CMS MACRA Feedback webpage.[55]

Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of stakeholders using the stakeholder contact list developed through previous education and outreach and clinician engagement efforts, as well as CMS, Quality Payment Program listservs.

Acumen and CMS hosted two office hour sessions between July and August 2020, to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Across both office hours sessions, there were over 35 attendees from targeted specialty societies.

Acumen worked closely with Quality Payment Program Service Center to respond to stakeholder inquiries during field testing and continued to answer questions after the feedback period ended.

Acumen and CMS posted the MACRA Wave 3 Cost Measures Field Testing Webinar to the Quality Payment Program Webinar Library at the start of the field testing period.[56] The webinar recording, slides, and transcript were available for stakeholders to review throughout field testing. The webinar presentation outlined: (i) the cost measure field testing project (ii) the measure development and re-evaluation processes, and (iii) field testing activities. The webinar recording was viewed approximately 450 times during the field testing period.

5.1.2.3 Feedback on Measure Performance and Implementation

Field Testing

In total, Acumen received 24 survey responses and 13 comment letters, including from specialty societies representing large numbers of potentially attributed clinicians.

Survey responses and comment letters were collected via an online survey, which contained general and detailed questions on the reports themselves, questions on the supplemental documentation, and questions on the measure specifications.

Pre-Rulemaking

CMS received 29 comments on the 5 episode-based cost measures included in the Measures Under Consideration List released in December 2020. This included four comments for the

[55]CMS, "Mock Field Test Reports," MACRA Feedback Page, https://www.cms.gov/files/zip/macra-2020-cmft-mock-reports.zip.
[56]MACRA Wave 3 Cost Measures Field Testing Webinar materials are available on the Quality Payment Program Webinar Library: https://qpp.cms.gov/about/webinars.

Colon and Rectal Resection measure. After the Measure Applications Partnership (MAP) Clinician Expert Workgroup meeting in January 2021, there was another public comment period on their preliminary recommendations, which received 25 comments across the 5 measures, with 5 comments specific to the Colon and Rectal Resection cost measure.[57] These public comment periods were facilitated by NQF. Stakeholders were able to submit their comments via the NQF website.

5.1.2.4 Feedback from Providers being Measured

Field Testing

The Field Testing Feedback Summary Report presents stakeholder feedback gathered during the field testing period.[58] The following list synthesizes some of the key points that were raised across all field-tested measures through the field testing feedback period:

- Measure development approach
    - Stakeholders expressed appreciation for the opportunity to provide feedback during field testing and for the incorporation of previous suggestions in an effort to continually improve the measure development and field testing processes.
    - Stakeholders reported that the COVID-19 and wildfire public health emergencies presented challenges to participating in field testing. CMS's inclusion of telehealth services in the cost measures, partly in response to the COVID-19 pandemic, was seen as a positive step that should be continued going forward in an effort to expand access to vulnerable patient populations so long as CMS monitors for unintended consequences.
- Field Test Report access, format, and content
    - Stakeholders didn't report any issues accessing Field Test Reports during the field testing period. Feedback generally was positive regarding the Field Test Report that was updated for 2020 and the supplemental episode-level data file, though some stakeholders preferred the previous Excel format.
- Components of episode-based cost measures
    - Field testing feedback was generally not supportive of the inclusion of Part D drug costs in cost measures, with stakeholders expressing concern that clinicians could be held accountable for transactions that are out of their control or if patients require high-cost medications. Relatedly, stakeholders expressed concern about the lack of transparency for Part D costs.
    - Stakeholder input related to the development and testing of chronic condition measures was mixed. Some stakeholders reported that chronic condition cost measures represent an opportunity to reduce healthcare costs without impeding patient access, choice, or quality of care while others reported it was difficult to evaluate the new measures without measure reliability testing results.
    - Stakeholders maintain that resource use and patient health outcomes are influenced by the social determinants of health and that the cost measures aren't adequately adjusted for these differences when calculating cost measures performance scores.
    - Stakeholders recognize the importance of linking cost and quality, including opportunities to do in the forthcoming MIPS Value Pathways (MVPs), to better evaluate clinician performance and improve patient health outcomes.

[57]Measure Applications Partnership, *National Quality Forum*, https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.
[58]CMS, "2020 Field Testing Feedback Summary Report," MACRA Feedback Page, https://www.cms.gov/files/document/macra-2020-ft-feedback-summary-report.pdf.

The summary report additionally contains measure-specific feedback, which was used as the basis for the post-field testing refinements that were made to the measures. See Section 5.1.2.6 for post-field testing refinements made to the Colon and Rectal Resection measure.

5.1.2.5 Feedback from Other Users

Pre-Rulemaking

In the 2020-2021 MAP review cycle, the MAP recommended "conditional support for rulemaking" for the Colon and Rectal Resection measure, conditional on NQF endorsement. The MAP noted evidence showing that surgical decision making and treatment course related to colon and rectal resection can reduce length of hospital stay, risk of major post-operative complications, and cost. The MAP stated that should testing data show that the measure appropriately assesses costs and can be used alongside quality measures, this measure would be valuable to add to the program measure set. The MAP's final recommendations are available for review on their website.[59]

Person and Family Engagement

Acumen incorporated actionable input from patients and caregivers throughout the Colon and Rectal Resection measure development process. Throughout Wave 3 of measure development, we solicited and considered PFE input on (i) selection of episode groups for development, and (ii) a broad set of questions around constructing measures that will provide meaningful feedback on clinicians' resource use via service assignment, provider attribution, episode length, and more. We also sought comments through a questionnaire during field testing for person and family input. This input was shared with the Colon and Rectal Resection Clinician Expert Workgroup for their consideration as they developed the measure. A discussion of the PFE approach and specific feedback is available on the MACRA Feedback Page.[60]

5.1.2.6 Consideration of Feedback

Field Testing

Careful consideration was given to all feedback gathered during field testing, and several updates were made to the measure based on the recommendations of field testing commenters and the Clinician Expert Workgroup comprised of subject matter and measure-development experts.

After completing field testing, Acumen compiled the feedback provided through the survey and comment letters into a measure-specific report, which was then provided to the Clinician Expert Workgroup, along with empirical analyses to inform their discussion and evaluation of any refinements needed to ensure that the measure is capturing what it was intended to capture.

The changes to the Colon and Rectal Resection measure made after consideration of field testing analyses and stakeholder feedback are:

- Triggers
  - CPT/HCPCS 45550 (treatment of rectal prolapse) added as trigger code
  - CPT/HCPCS 45130 (repair of prolapsed rectum) removed as trigger code
- Exclusions
  - Episodes without an IP component excluded
- Service assignment

[59]Measure Applications Partnership, National Quality Forum, "2020-2021 MAP Final Recommendations" https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=94650

[60]CMS, Summary of Person and Family Engagement (PFE) and Input for Wave 3 Episode-based Cost Measure Development (March 2021). https://www.cms.gov/files/document/summary-person-and-family-engagement.pdf

    - Diagnostic cardiac catheterization, coronary arteriography removed from 15- day pre-trigger period
- Measure-specific risk adjustors
    - Obesity removed
    - Recent Cardiac Arrest removed

5.2 Usability

5.2.1 Improvement

N/A. The measure has not yet been implemented, and as such has not had influence over performance.

5.2.2 Unexpected Findings

N/A. There were no unexpected findings during the development and testing of this measure.

5.2.3 Unexpected Benefits

N/A. There were no unexpected benefits during the development and testing of this measure.

Other Additional Information

Colon and Rectal Resection Clinician Expert Workgroup Members:
Carol Parrish, MS, RDN, Academy of Nutrition and Dietetics
Colleen Schmitt, MD, MHS, American Society for Gastrointestinal Endoscopy
Ezequiel Silva III, MD, FACR, American College of Radiology
Guy Orangio, MD, FACS, FASCRS, American College of Surgeons
Janette Dietzler-Otte, DNP, RN, CWS, CWON, Wound, Ostomy and Continence Nursing Society
Mary Cathleen Shellnutt, DNP, APRN, AGCNS-BC, CGRN, National Association of Clinical Nurse Specialists
Melinda Maggard-Gibbons, MD, MSHS, American College of Surgeons
Michael Sutherland, MD, FACS, American College of Surgeons
Nina Paonessa, DO, FACOS, American Osteopathic Association
Ofor Ewelukwa, MD, MSc, American Gastroenterological Association
Sarah Gebauer, MD, American Society of Anesthesiologists
Scott Regenbogen, MD, MPH, American College of Surgeons
Steven Nurkin, MD, MS, FACS, Society of Surgical Oncology
Tomas Villanueva, DO, MBA, Society of Hospital Medicine
Tracy Young, MSNA, MBA, CRNA, American Association of Nurse Anesthetists
Walter Peters, MD, MBA, American Society of Colon and Rectal Surgeons
Wayne Johnson, DMSc, PA-C, American Academy of Physician Assistants

The Colon and Rectal Resection Clinician Expert Workgroup is composed from the larger General and Colorectal Surgery Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the Episode-Based Cost Measures Development Process document.[61]

[61]CMS, “Episode-Based Cost Measure Field Testing Wave 3 Measure Development Process,” MACRA Feedback Page, https://www.cms.gov/files/document/macra-cmft-ebcm-process-2020.pdf.