

## Lower Gastrointestinal Hemorrhage Measure

### Measure Justification Form

June 2019



# Contents

<b>1.0</b>	<b>Introduction .....</b>	<b>4</b>
1.1	Project Title and Overview .....	4
1.2	Measure Name .....	4
1.3	Type of Measure .....	4
<b>2.0</b>	<b>Importance.....</b>	<b>5</b>
2.1	Evidence to Support the Measure Focus .....	5
2.1.1	Measure Description .....	5
2.1.2	Evidence for Measure Focus .....	5
2.2	Performance Gap .....	6
2.2.1	Rationale .....	6
2.2.2	Performance Scores.....	7
<b>3.0</b>	<b>Scientific Acceptability .....</b>	<b>8</b>
3.1	Data Sample Description .....	8
3.1.1	Type of Data Used for Testing.....	8
3.1.2	Specific Dataset Used for Testing .....	8
3.1.3	Dates of the Data Used in Testing.....	8
3.1.4	Levels of Analysis Tested.....	8
3.1.5	Entities Included in the Testing and Analysis.....	8
3.1.6	Patient Cohort Included in the Testing and Analysis.....	8
3.1.7	Sample Differences .....	10
3.1.8	Social Risk Factors Included in Analysis .....	10
3.2	Reliability Testing.....	10
3.2.1	Level of Reliability Testing .....	10
3.2.2	Method of Reliability Testing .....	10
3.2.3	Statistical Results from Reliability Testing .....	11
3.2.4	Interpretation .....	12
3.3	Validity Testing .....	12
3.3.1	Level of Validity Testing.....	12
3.3.2	Method of Validity Testing .....	12
3.3.3	Statistical Results from Validity Testing .....	14
3.3.4	Interpretation .....	15
3.4	Exclusions Analysis .....	15
3.4.1	Method of Testing Exclusions.....	15
3.4.2	Statistical Results from Testing Exclusions.....	16
3.4.3	Interpretation .....	17
3.5	Risk Adjustment or Stratification .....	18
3.5.1	Method of Controlling for Differences .....	18
3.5.2	Conceptual, Clinical, and Statistical Methods .....	19
3.5.3	Conceptual Model of Impact of Social Risks.....	20
3.5.4	Statistical Results .....	20
3.5.5	Analyses and Interpretation in Selection of Social Risk Factors.....	20
3.5.6	Method for Statistical Model or Stratification Development.....	21
3.5.7	Statistical Risk Model Discrimination Statistics .....	22
3.5.8	Statistical Risk Model Calibration Statistics .....	22
3.5.9	Statistical Risk Model Calibration – Risk Decile.....	22
3.5.10	Interpretation.....	22
3.6	Identification of Meaningful Differences in Performance.....	22
3.6.1	Method .....	22

3.6.2 Statistical Results .....	23
3.6.3 Interpretation .....	23
3.7 Missing Data Analysis and Minimizing Bias.....	23
3.7.1 Method .....	23
3.7.2 Missing Data Analysis .....	24
3.7.3 Interpretation .....	24
<b>4.0 Feasibility.....</b>	<b>25</b>
4.1 Data Elements Generated as Byproduct of Care Processes .....	25
4.2 Electronic Sources .....	25
4.3 Data Collection Strategy .....	25
4.3.1 Data Collection Strategy Difficulties.....	25
<b>5.0 Usability and Use .....</b>	<b>26</b>
5.1 Use .....	26
5.1.1 Current and Planned Use .....	26
5.1.2 Feedback on the Measure and Development Process .....	26
5.2 Usability .....	30
5.2.1 Improvement .....	30
5.2.2 Unexpected Findings.....	30
5.2.3 Unexpected Benefits .....	30
<b>6.0 Related and Competing Measures .....</b>	<b>31</b>
6.1 Relation to Other Cost Measures .....	31
6.2 Harmonization.....	31
6.3 Competing Measures .....	31
<b>Contact Information .....</b>	<b>32</b>

# 1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Lower Gastrointestinal Hemorrhage measure. The MJF is intended to provide detailed information about the testing conducted on this measure and accompanies the Measure Methodology and Measure Codes List files, which together comprise the specifications for this cost measure.<sup>1</sup>

## 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 “MACRA Episode Groups and Cost Measures.” The contract number is HHSM-500-2013-13002I, Task Order HHSM-500-T0002.

## 1.2 Measure Name

Lower Gastrointestinal Hemorrhage Episode-Based Cost Measure

## 1.3 Type of Measure

Cost/Resource Use

---

<sup>1</sup> CMS, “Lower Gastrointestinal Hemorrhage Measure Methodology,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.

CMS, “Lower Gastrointestinal Hemorrhage Measure Codes List,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.

## 2.0 Importance

### 2.1 Evidence to Support the Measure Focus

#### 2.1.1 Measure Description

The Lower Gastrointestinal Hemorrhage cost measure (also referred to as “the Lower GI Bleed measure”) evaluates clinicians’ risk-adjusted cost to Medicare for beneficiaries who receive inpatient non-surgical treatment for acute bleeding in the lower gastrointestinal tract. The cost measure score is a clinician’s average risk-adjusted cost for the episode group across all episodes attributed to the clinician. This acute inpatient medical condition measure includes costs of services that are clinically related to the attributed clinician’s role in managing care during each episode, from the clinical event that opens or ‘triggers’ the episode through 35 days after the trigger. Beneficiary populations eligible for the Lower Gastrointestinal Hemorrhage measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

#### 2.1.2 Evidence for Measure Focus

Policymakers contend that an estimated 80 percent of overall health care costs are attributable to decisions made by clinicians.<sup>2</sup> However, these same clinicians are often unaware of how their care decisions influence the overall costs of care. One of the goals for using cost measures is to help inform clinicians on the costs attributable to their decision-making, as well as the total cost of their patient’s care. A cost measure offers opportunity for improvement if clinicians can exercise influence on a significant share of costs during the episode, or if lower spending and better care quality can be achieved through changes in clinical practice.

According to the literature and previous feedback received through stakeholder input activities, Lower GI Bleed as a measure represents an area where there are many opportunities for improvement, including reducing the incidence of recurrent bleeding and using early intervention strategies to mitigate the risk for downstream complications.

Closely monitoring a patient’s medical and drug history can reduce the incidence of recurrent bleeding and the associated costs. The use of multiple medications is common in the elderly population, given age-related ailments such as cerebrovascular disease, atherosclerotic heart disease, and arthritis. However, the use of some medications are known to increase the risk of both an initial episode of Lower GI Bleed and re-bleeding,<sup>3,4</sup> leaving clinicians faced with a dilemma of how best to approach patients in need of multi-modal therapy. Clinicians need access to more information on risk comparisons at the individual patient level to make better-informed decisions about management for patients at increased risk for lower GI bleeding.

Establishing effective intervention strategies for lower GI bleeding risk factors may prevent the onset of additional complications. Risk factor models for GI bleeding predictors are not as well studied for lower GI bleeding as they are for upper GI bleeding; however, some studies have found the likelihood of adverse outcomes for lower GI bleeding increases with the number of

---

<sup>2</sup> Fred, Herbert L. “Cutting the Cost of Health Care: The Physician’s Role.” *Texas Heart Institute Journal*, vol. 43, no. 1, 2016, pp. 4 – 6.

<sup>3</sup> Cheung, Ka-Shing and Wai K Leung. “Gastrointestinal Bleeding in Patients on Novel Oral Anticoagulants: Risk, Prevention and Management.” *World Journal of Gastroenterology*, no. 23, vol. 11, 2017, pp. 1954-1963.

<sup>4</sup> Taki, Masato, Tadyuki Oshima, et al. “Analysis of Risk Factors for Colonic Diverticular Bleeding and Recurrence.” *Medicine (Baltimore)*, vol. 96, no. 38, 2017, pp. e8090.

risk factors present.<sup>5</sup> In 2009, there were around 2.7 million outpatient clinical visits for diverticular disease;<sup>6</sup> given that diverticular disease is the most common cause of lower GI bleeding, particularly in the elderly, better management of diverticular disease may significantly reduce the initial onset or severity of bleeding. This holds especially true for patients with high levels of risk factors, including comorbid conditions and multimodal drug regimens, with the median hospital cost for diverticulitis around \$6,000 per patient and an estimated total of \$2.6 billion per year in inpatient costs.<sup>7</sup> Establishing a more standardized approach to evaluating and managing risk factors could potentially moderate downstream complications and costs associated with lower GI bleeding.

This measure aims to address these example areas of opportunities for improvement. Given the prevalence of lower GI bleeding among Medicare beneficiaries, the use of this episode-based cost measure can provide clinicians with information to improve care outcomes and reduce future health care costs.

## 2.2 Performance Gap

### 2.2.1 Rationale

Gastrointestinal bleeding is the most common cause of hospitalizations among gastrointestinal diseases, with over 500,000 patients hospitalized annually.<sup>8,9</sup> Lower GI bleeding is responsible for approximately 30 to 40 percent of all gastrointestinal bleeding cases, with an incidence of around 36 per 100,000 persons.<sup>10,11</sup> Morbidity and mortality due to GI bleeds increase significantly for patients who are older and for those with pre-existing medical conditions, leading to higher costs and resource use.<sup>12</sup> With application of an effective measurement tool, opportunities for improvement in care of lower GI bleeding include better methods for characterizing patients at higher risk for re-bleeding, better approaches to treatment and ongoing management to reduce the incidence of recurrent bleeding, and improved use of early intervention strategies to mitigate the risk for catastrophic bleeding and other associated downstream complications. The Lower Gastrointestinal Hemorrhage episode-based cost measure was recommended for development by an expert clinician committee—the Gastrointestinal Disease Management – Medical and Surgical Clinical Subcommittee—because of its high impact in terms of patient population and Medicare spending and the opportunity for incentivizing cost-effective, high-quality clinical care in this area. Based on the initial recommendations from the Clinical Subcommittee, the subsequent measure-specific workgroup provided extensive, detailed input on this measure.

---

<sup>5</sup> Gralnek, Ian M, Ziv Neeman, et al. "Acute Lower Gastrointestinal Bleeding." *The New England Journal of Medicine*, no. 376, 2017, pp. 1054-1063.

<sup>6</sup> Peery, Anne F, Evan S Dellon, et al. "Burden of Gastrointestinal Disease in the United States: 2012 Update." *Gastroenterology*, vol. 143, no. 5, 2012, pp. 1179-1187.

<sup>7</sup> Ibid.

<sup>8</sup> Gralnek, Ian M, Ziv Neeman, et al. "Acute Lower Gastrointestinal Bleeding." *The New England Journal of Medicine*, no. 376, 2017, pp. 1054-1063.

<sup>9</sup> Strate, Lisa L and Ian M Gralnek. "ACG Clinical Guideline: Management of Patients with Acute Lower Gastrointestinal Bleeding." *The American Journal of Gastroenterology*, vol. 111, 2016, pp. 459-474.

<sup>10</sup> Ibid.

<sup>11</sup> Parekh, Parth J, Ross C Buerlein, et al. "Evaluation of Gastrointestinal Bleeding: Update of Current Radiologic Strategies." *World Journal of Gastrointestinal Pharmacology and Therapeutics*, vol. 5, no. 4, 2014, pp. 200-208.

<sup>12</sup> Jansen, Antje, Sabine Harenberg, et al. "Risk Factors for Colonic Diverticular Bleeding: A Westernized Community Based Hospital Study." *World Journal of Gastroenterology*, vol. 15, no. 4, 2009, pp. 457-461.

### 2.2.2 Performance Scores

Performance scores are provided for 1,274 clinician group practices (identified by Tax Identification Number [TIN]) and 137 practitioners (identified by combination of TIN and National Provider Identifier [NPI]). These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a Merit-based Incentive Program (MIPS) eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced Alternative Payment Model participation). This table uses a testing volume threshold of 20 episodes.

**Table 1: Distribution of Performance Scores**

Metric	TIN	TIN-NPI
Mean score	\$10,564	\$11,663
Standard deviation	\$806	\$1,061
Score IQR	\$996	\$1,331
Score percentile		
10 <sup>th</sup>	\$9,577	\$10,435
20 <sup>th</sup>	\$9,925	\$10,862
30 <sup>th</sup>	\$10,148	\$11,035
40 <sup>th</sup>	\$10,336	\$11,351
50 <sup>th</sup>	\$10,517	\$11,598
60 <sup>th</sup>	\$10,721	\$11,784
70 <sup>th</sup>	\$10,926	\$12,064
80 <sup>th</sup>	\$11,189	\$12,572
90 <sup>th</sup>	\$11,584	\$13,155

## 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), enrollment database (EDB), and Common Medicare Environment (CME)

#### 3.1.2 Specific Dataset Used for Testing

The Lower Gastrointestinal Hemorrhage measure uses Medicare Part A and Part B claims data maintained by CMS. Part A and B claims data are used to build episodes of care, calculate episode costs, and construct risk adjusters. Data from the EDB are used to determine beneficiary-level exclusions and supplemental risk adjusters, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), beneficiary birth dates, and beneficiary death dates. The risk adjustment model also accounts for expected differences in payment for services provided to beneficiaries in long-term care based on the 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 American Census, American Community Survey (ACS), and CME are used in analyses evaluating social risk factors in risk adjustment.

#### 3.1.3 Dates of the Data Used in Testing

The measurement period includes Lower Gastrointestinal Hemorrhage episodes ending from January 1, 2017 to December 31, 2017.

#### 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

1,274 clinician group practices and 137 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 20 or more Lower GI Bleed episodes during the measurement period. Episodes from all 50 States and D.C. in the following settings were included: acute inpatient (IP) hospitals.

#### 3.1.6 Patient Cohort Included in the Testing and Analysis

54,114 Medicare beneficiaries (from 58,389 episodes) were included in TIN level testing and analysis, and 2,978 beneficiaries (from 3,086 episodes) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the Lower Gastrointestinal Hemorrhage measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who received inpatient non-surgical treatment for acute bleeding in the lower gastrointestinal tract during the measurement period as identified by the episode trigger Medicare Severity Diagnosis-Related Group (MS-DRG) codes on IP claims. Beneficiaries and their episodes were included in the sample if they met a set of inclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients receiving inpatient treatment for lower GI bleeding.



The inclusion criteria are:

- The beneficiary has Medicare as their primary payer for the entire episode window, as well as the 120 days prior to the trigger day (the 120-day lookback period).
- The beneficiary was continuously enrolled in Medicare Parts A and B, and not enrolled in Part C, for the entirety of the episode window and the 120-day lookback period.
- The beneficiary has a sufficient 120-day lookback period.
- The beneficiary date of birth is not missing.
- The beneficiary death date did not occur before episode end.
- At least one TIN is attributed the episode.
- The episode can be attributed to at least one clinician.
- The episode trigger IP stay occurs in a short-term stay acute hospital as defined by subsection (d).<sup>13</sup>
- The episode trigger IP stay does not have the same admission date as another IP stay.
- The beneficiary does not have inflammatory bowel disease.
- The beneficiary does not leave against medical advice.
- The episode trigger claim does not list as principal diagnosis a condition most commonly associated with upper GI bleeding.
- If the episode trigger claim lists nonspecific GI bleed as principal diagnosis, there is no additional upper GI bleed diagnosis on the claim.
- The episode is not an outlier case.

To determine whether the Lower GI Bleed measure's inclusion 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 inclusion criteria, (ii) episodes without inclusion criteria, (iii) beneficiaries with inclusion criteria, and (iv) beneficiaries without inclusion criteria.

This analysis shows that the Lower GI Bleed measure's inclusion criteria have only a minimal effect on the percentage of beneficiaries of any particular demographic. The difference between beneficiaries being included or not included in the measure is less than or equal to 5.0 percentage points across each of the characteristics in the analysis at TIN level testing, and less than or equal to 4.0 percentage points at TIN-NPI level testing. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 13.8 percent, compared to 10.6 percent at TIN level and 10.3 percent at TIN-NPI level testing. The difference in the percentage of beneficiaries for race with and without the inclusion criteria is less than 1.4 percentage points for all but one category, and is between 1.2 and 3.1 percentage points for the final category (black) for TIN and TIN-NPI testing. The inclusion criteria affects beneficiary percentages to the same extent at both levels of testing, with a difference with and without the inclusion criteria of 2.7 percentage points at the TIN level and 3.7 percentage points at the TIN-NPI level. These results indicate that there is minimal shift in patient characteristics as a result of using the inclusion criteria listed above at both TIN and TIN-NPI level testing.

---

<sup>13</sup> 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](#).

### 3.1.7 Sample Differences

n/a

### 3.1.8 Social Risk Factors Included in Analysis

The social risk factors analyzed were variables from the ACS, EDB, and CME. All ACS variables are at the Census Block Group level. Social risk variables analyzed include the following:

- 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%
- Race (EDB)
  - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
  - Female, male
- Dual status (CME)
  - Full dual, partial dual, non-dual

## 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 Lower GI Bleed 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 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 2017, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3 to 96.4 percent of total payments

each year.<sup>14</sup> The fiscal year 2018 Medicare fee-for-service program proper payment rate was 91.9 percent.<sup>15</sup> 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  $j$

$\sigma_b^2$  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.

### 3.2.3 Statistical Results from Reliability Testing

#### Measure Reliability

44.2 percent, 74.6 percent, and 100 percent of TINs at 10, 20, and 30-episode volume thresholds, respectively, have mean reliability greater than or equal to 0.4. At a testing volume threshold of at least 10 episodes, the mean reliability for TINs is 0.41. The mean reliability continues to increase at the 20 and 30-episode volume thresholds.

**Table 2: Reliability Results at Various Volume Thresholds**

Volume Threshold (# episodes)	TIN		TIN-NPI	
	Mean Reliability	% ≥ 0.4	Mean Reliability	% ≥ 0.4
10	0.41	44.2%	0.12	0.0%
20	0.51	74.6%	0.20	0.0%
30	0.59	100.0%	0.28	0.0%

<sup>14</sup> Comprehensive Error Rate Testing (CERT) Program. "Appendices Medicare Fee-for-Service 2018 Improper Payments Report". Table A6. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/Downloads/2018MedicareFFSSupplementalImproperPaymentData.pdf>

<sup>15</sup> Ibid.

### 3.2.4 Interpretation

#### Measure Reliability

Overall reliability of the Lower GI Bleed measure is moderate at a volume threshold of 20 episodes or more for TINs, due to the team-based nature of inpatient treatment for acute conditions. The low reliability for TIN-NPIs is largely a consequence of said treatment methodology, which results in very few TIN-NPIs being attributed Lower GI Bleed episodes. CMS generally considers 0.4 as the threshold indicating 'moderate' reliability, which is supported by previous work into reliability.<sup>16</sup>

While higher volume thresholds yield even higher reliability results, it is at the cost of further reducing the number of clinicians and clinician groups able to receive a measure score.

### 3.3 Validity Testing

#### 3.3.1 Level of Validity Testing

We conducted performance measure score validity testing, which included empirical validity testing and systematic assessment of face validity.

#### 3.3.2 Method of Validity Testing

##### Face Validity

The Lower GI Bleed measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. These expert panels were convened to methodically assess the extent to which the measure: (i) captured what it was intended to capture, and (ii) differentiated between provider performance. Experts in this clinical area evaluated specifications in an iterative process 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 and refining this measure, Acumen incorporated input from (i) the Gastrointestinal Disease Management – Medical and Surgical Clinical Subcommittee, (ii) the Lower Gastrointestinal Hemorrhage workgroup, (iii) a Technical Expert Panel (TEP), (iv) a Person and Family Committee (PFC), and (v) stakeholder feedback from national field testing.

The Clinical Subcommittee comprised 53 members with clinical experience in gastrointestinal disease management, affiliated with 32 specialty societies. The Clinical Subcommittee provided input at an in-person meeting in April 2018 on which measure to develop, on the measure scope, and on the composition of a smaller, targeted workgroup to provide detailed input on each aspect of measure specifications. The Lower Gastrointestinal Hemorrhage workgroup was composed of 13 members affiliated with 11 specialty societies, including the American Society for Gastrointestinal Endoscopy, the American College of Gastroenterology, and the American Gastroenterological Association. The workgroup considered empirical analyses and their clinical expertise to provide input during an in-person meeting and several webinars between June to December 2018. Input was gathered in a structured manner including the use of a polling process requiring greater than 60 percent consensus.

The TEP provided high-level guidance and input on the overall direction of measure development and the framework for episode-based cost measures, while the PFC provided a patient and family perspective. PFC input included concepts of healthcare quality and value,

---

<sup>16</sup> 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](http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf)

guiding principles and measure-specific input to inform the workgroups such as pre- and post-trigger windows for selected episodes, and inclusion of services and costs for attributed clinicians. In addition, the national field testing feedback period in October and November 2018 offered all stakeholders an opportunity to review and provide input on draft measure specifications and measure feedback reports for attributed clinicians and clinician groups. During this period, 78,221 field test reports for TINs and TIN-NPIs were available for download and review for 11 episode-based cost measures developed throughout 2018.

One of the key roles of the measure-specific 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 treatment of lower GI bleeding, thus preventing inclusion of unrelated cost variation in this measure. Assigned services occurring in the inpatient setting include non-surgical lower GI bleed treatment, evaluation, testing, complication management, and follow-up.

### Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the Lower GI Bleed measure by examining differences in risk-adjusted cost for known indicators of resource or service utilization based on a literature review, specifically complications related to downstream acute (re)admission, as well as episodes with and without post-acute care (PAC), including care provided at inpatient rehabilitation facilities, long-term care hospitals, home health, and skilled nursing facilities. For this analysis, we compared the ratio of observed to expected cost (henceforth called "O/E cost ratio") for Lower Gastrointestinal Hemorrhage episodes with and without complications related to acute (re)admissions occurring in the post-trigger period and episodes with and without PAC. This analysis sought to confirm the expectation that the Lower GI Bleed measure captures variation in service utilization.

In the second approach, we evaluated how different types of cost impact risk-adjusted measure scores. Certain services or costs included in the Lower GI Bleed measure were classified into clinically coherent groups of services, called "clinical themes". The Lower GI Bleed measure clinical themes are:

- **Physical Therapy / DME:** includes physical and occupational therapy and supplies, e.g., ostomy pouch, walkers, hospital beds, and wheelchairs
- **Post-Discharge Imaging:** includes MRIs, x-rays, and other imaging for the abdomen or chest
- **Post-Acute Care:** reflects treatment related to physical therapy, occupational therapy, speech therapy, and home health services
- **Recurrent GI Bleed / Anemia:** includes laboratory testing, procedures to stop bleeding, and associated inpatient and outpatient hospital care
- **Cardiovascular Complications:** accounts for inpatient and outpatient hospital care including emergency department visits and critical care for cardiac arrest, atrial fibrillation and flutter, hypotension or cardiac arrhythmias, e.g., imaging, intubation and mechanical ventilation, and diagnostic and therapeutic procedures
- **Pulmonary / Respiratory Complications:** reflects inpatient and outpatient hospital care including emergency department visits or critical care related to pneumonia, respiratory failure or disorders, bronchitis, and other pulmonary disorders and diagnostic and therapeutic procedures
- **Renal Failure / Electrolyte Abnormalities:** accounts for inpatient and outpatient hospital care including emergency department visits or critical care related to kidney

failure, volume depletion, or other disorders of fluid, electrolyte and acid-base balance, e.g., ultrasounds, diagnostic and therapeutic procedures, and hemodialysis

- **Thromboembolism (DVT/PE):** includes procedures to address a pulmonary embolism, thrombosis, or insertion of a vena cava filter
- **Post-Discharge Endoscopic Treatment:** includes colonoscopy, biopsy, ultrasounds, anesthesia, or imaging
- **Post-Discharge Surgical Treatment:** accounts for colorectal resection, anesthesia, major small and large bowel procedures, and anal and stomal procedures for hemorrhoids, disorders of the anus, rectum, intestine, or digestive system

As with the first analysis for validity, the aim of this analysis was to determine whether the measure is capturing variation in provider cost in the manner intended and expected. To measure this, we took the Pearson correlation between the cost of each clinical theme and the overall risk-adjusted cost for an episode.

We expected that the two themes related to complications (Cardiovascular and Pulmonary/Respiratory) would have the highest correlation with risk-adjusted episode cost, as complications are likely associated with high cost even after accounting for beneficiary characteristics.<sup>17</sup> We would expect similar trends for the Recurrent GI Bleed/Anemia, Renal Failure/Electrolyte Abnormalities, and Thromboembolism (DVT/PE) themes, as they contain services relating to complications. By contrast, we expected that Physical Therapy/DME and Post-Discharge Imaging might have more nuanced, offsetting effects. While higher costs for these types of visits can directly increase the costs of an episode, research indicates that pre- and post-treatment interventions can be associated with lower total resource use by saving on later costs.<sup>18</sup> Therefore, it is possible the correlation of the measure with these types of costs is lower than for complications.

### 3.3.3 Statistical Results from Validity Testing

For the first analysis of validity, the mean O/E cost ratio for all episodes is 1.00. The mean O/E cost ratio for episodes with services relating to complications during the post-trigger period is 1.68, compared with 0.94 for episodes without services relating to complications during the post-trigger period. Episodes with PAC have a mean O/E cost ratio of 1.49, compared to 0.88 for episodes without PAC. Table 3 offers additional details on the O/E cost ratios for the various episode types.

**Table 3: Distribution of Observed to Expected Cost Ratios**

Episode Type	O/E										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
All Final Episodes	1.00	0.43	0.50	0.62	0.68	0.76	0.85	1.03	1.60	1.95	2.67
Episodes with Downstream Acute (Re)admission	1.68	0.47	0.87	1.07	1.18	1.37	1.60	1.90	2.30	2.61	3.11
Episodes without Downstream Acute (Re)admission	0.94	0.38	0.50	0.61	0.67	0.75	0.84	0.97	1.35	1.79	2.57

<sup>17</sup> Khan, N.A., Quan, H., Bugar, J.M. et al., "Association of postoperative complications with hospital costs and length of stay in a tertiary care center" J Gen Intern Med (2006) 21: 177.

<sup>18</sup> Devine, Elizabeth C., Cook, Thomas D., "Clinical and cost-saving effects of psychoeducational interventions with surgical patients: A meta-analysis"



Episode Type	O/E										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
Episodes with Post-Acute Care	1.49	0.57	0.65	0.79	0.87	1.04	1.38	1.83	2.27	2.60	3.05
Episodes without Post-Acute Care	0.88	0.29	0.49	0.60	0.67	0.74	0.82	0.92	1.11	1.47	2.14

The clinical themes analysis demonstrates that there is a strong correlation between the Post-Discharge Surgical Treatment (correlation: 0.77) and Post-Acute Care (correlation: 0.73) themes and risk-adjusted cost. By contrast, the Post-Discharge Imaging (correlation: -0.03) and Physical Therapy/DME (correlation: -0.09) themes had very low correlation with risk-adjusted cost.

### 3.3.4 Interpretation

As expected, the average O/E cost ratio for episodes with post-trigger complications is higher than for episodes without downstream complications. Similarly, the average O/E cost ratio for episodes with PAC is higher than for episodes without PAC. These results demonstrate that the Lower Gastrointestinal Hemorrhage measure is able to accurately capture higher resource use.

The output of the clinical themes analysis does not completely align with what was anticipated. On the one hand, there was a reasonably strong correlation with risk-adjusted cost for clinical themes with high average cost—in particular, Post-Discharge Surgical Treatment (average cost: \$5,752.93, correlation: 0.77), Post-Acute Care (average cost: \$7,875.13, correlation: 0.73), Thromboembolism (DVT/PE) (average cost: \$2,899.87, correlation: 0.49), and Recurrent GI Bleed/Anemia (average cost: \$1,277.99, correlation: 0.47)—further confirming the Lower GI Bleed measure's ability to capture high resource use. Also as expected, themes relating to testing and physician visits—including Post-Discharge Imaging (average cost: \$148.29, correlation: -0.03) and Physical Therapy/DME (average cost: \$220.19, correlation: -0.09)—reflect treatment which may add cost but reduce need for subsequent treatment, and so reasonably do not clearly correlate with risk-adjusted cost. However, the complications clinical themes (Cardiovascular and Pulmonary/Respiratory) had both a low correlation with risk-adjusted cost (0.10 and 0.16, respectively) and unexpectedly low average episode cost (\$267.66 and \$343.46, respectively). This may indicate that the clinical themes themselves do not sufficiently capture treatment variation specifically associated with complications, but, given the earlier observations, does not itself indicate that the measure captures only mechanical increases in episode costs.

## 3.4 Exclusions Analysis

### 3.4.1 Method of Testing Exclusions

Exclusions are used in the Lower GI Bleed measure to ensure a homogenous patient population within the scope of the measure focus on acute bleeding in the lower gastrointestinal tract and that episodes provide meaningful information to attributed clinicians or as part of data processing, to ensure that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode. For the exclusions analysis, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- *Episodes where beneficiary death date occurred before the episode end date.*
  - These episodes are excluded for all measures due to the potential to inaccurately reflect a clinician's performance. Episodes where the beneficiary died may be

unusually high-cost, due to perimortem treatment costs, or unusually low-cost, due to the truncated episode window. Neither of these cases accurately reflects the efficiency of the clinician performing the treatment.

- *Episodes where the beneficiary elects to leave against medical advice*
  - Leaving against medical advice prevents the attributed clinician from completing appropriate care for the patient, which leaves the patient at high risk of further complications. Retaining such beneficiaries would put the attributed clinician at risk of being attributed a costly episode for which they did not have the chance to be able to fully treat the patient.
- *Episodes where the patient has inflammatory bowel disease (IBD)*
  - Patients with IBD likely had other reasons for experiencing a lower GI bleed and, as a result, would undergo different hospital and post-hospital courses.
- *Episodes where the principal diagnosis is of upper GI bleed*
  - Upper GI bleeding requires very different treatment relative to lower GI bleeding, so these episodes were excluded in order to focus only on those patients with lower GI bleeds.
- *Episodes where the principal diagnosis is of nonspecific GI bleed and there is a secondary diagnosis of upper GI bleed*
  - Despite the principal diagnosis being “nonspecific,” presence of an upper GI bleed diagnosis indicates that these episodes are not purely lower GI bleeds. Again, given the substantively different treatment required for lower vs. upper GI bleed, these episodes were excluded in order to focus only on those patients with lower GI bleeds.
- *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 1<sup>st</sup> percentile and above the 99<sup>th</sup> 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 beneficiaries affected, as well as the distributions of observed cost and O/E cost 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 two patient cohorts. A full list of the exclusions and details used for the Lower GI Bleed measure is provided in the Measure Codes List.<sup>19</sup>

### 3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics and O/E cost ratios for the Lower GI Bleed measure exclusions. Cost statistics are also provided for the set of final episodes included in the Lower GI Bleed measure for comparison, with a testing volume threshold of 20 episodes at the TIN and TIN-NPI levels.

---

<sup>19</sup> CMS, “Lower Gastrointestinal Hemorrhage Measure Codes List,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.



**Table 4: Cost Statistics for Measure Exclusions**

Exclusion	Episodes		Observed Cost			O/E		
	#	%	Mean	Percentile		Mean	Percentile	
				10 <sup>th</sup>	90 <sup>th</sup>		10 <sup>th</sup>	90 <sup>th</sup>
All Episodes Meeting Triggering Logic	200,341	100.00%	\$12,494	\$6,443	\$23,977	1.01	0.66	1.62
Death in Episode	15,242	7.61%	\$15,553	\$7,121	\$26,333	0.96	0.59	1.54
Inflammatory Bowel Disease	3,238	1.62%	\$13,799	\$6,631	\$25,905	1.01	0.64	1.62
Leaving Against Medical Advice	1,770	0.88%	\$10,834	\$5,855	\$17,609	0.98	0.69	1.51
Principal Diagnosis of Upper GI Bleed	63,607	31.75%	\$12,609	\$6,669	\$23,620	1.01	0.67	1.62
Principal Diagnosis of Nonspecific GI Bleed, Upper GI Bleed in Array	25,064	12.51%	\$11,911	\$6,645	\$22,280	1.02	0.69	1.63
Outlier Cases	1,626	0.81%	\$28,362	\$5,903	\$56,388	2.04	0.35	4.38
<i>Final Episodes (TIN)</i>	58,389	29.14%	\$10,700	\$6,086	\$19,867	0.97	0.67	1.55
<i>Final Episodes (TIN-NPI)</i>	3,086	1.54%	\$10,700	\$6,150	\$20,055	0.99	0.68	1.56

### 3.4.3 Interpretation

Although statistical results indicate that many of the excluded episodes, aside from outliers and episodes where the beneficiary died, have somewhat similar results to the final set of episodes, these episodes were still excluded due to clinical considerations to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for each exclusion is provided below.

*Episodes ending in death:* Typically, excluding episodes that end in death is appropriate based on higher-level concerns, such as worries over truncated episodes giving the appearance of more cost-effective care. However, in this case, the difference between mean observed episode cost for death and non-death final episodes is substantial in and of itself (\$15,553 compared to \$10,700), and because decedents likely necessitated substantively different care—such as surgical treatments or blood transfusions—excluding such episodes is warranted.

*Episodes where the beneficiary elects to leave against medical advice:* The Lower GI Bleed measure is intended to incentivize clinicians to change their behavior and treatment patterns to increase cost-effectiveness. However, the measure cannot accurately reflect such improvements if TINs and TIN-NPIs are penalized for patients who do not take advantage of the offered care. Though leaving against medical advice results in only slightly higher average observed cost (\$10,834 compared to \$10,700), these beneficiaries are excluded to allow the measure to capture the outcome of clinicians' decisions.

*Episodes where the patient has IBD:* These episodes exhibit a higher average cost (\$13,799 compared to \$10,700) along with the clinical reasoning that IBD patients likely require different treatment plans.

*Episodes where the principal diagnosis is of upper GI bleed or is of nonspecific GI bleed and there is a secondary diagnosis of upper GI bleed:* While these episodes are not substantially more expensive than average (\$12,609 and \$11,911, respectively, compared to \$10,700), the primary concern with this exclusion continues to be retaining the clinical homogeneity of the cohort to patients with lower GI bleeding.

*Episodes classified as outlier cases:* The O/E cost ratio ranges from 0.35 at the 10<sup>th</sup> percentile to 4.38 at the 90<sup>th</sup> 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 119 risk factors. This measure's risk adjustment model is not stratified by risk categories.

The risk adjustment model for the Lower Gastrointestinal Hemorrhage 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. Although the MA risk adjustment model includes 24 age/sex variables, this risk adjustment model does not adjust for sex and so only includes 12 age categorical 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 beneficiary'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 beneficiaries without a full 120-day lookback period are excluded from the measure. This 120-day period is used to measure beneficiary health status and ensures that each beneficiary'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 beneficiary qualifies for Medicare through Disability or ESRD. The model also includes an indicator of whether the beneficiary recently required long-term care, defined as 90 days in a long-term care facility without being discharged to community for 14 days. Beneficiaries who need to reside in long-term care facilities typically require more intensive care than beneficiaries 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 beneficiary was diagnosed with anemia during the 120-day lookback period, indicating a higher risk for requiring blood transfusions and prolonged hospital and post-hospital courses;
- whether the beneficiary used anticoagulant or antiplatelet medications during the lookback period, creating a higher risk of readmissions for GI bleed or other complications and thus making episodes more costly;
- whether the diagnosis associated with the current IP admission was rectal bleeding, because such cases are less costly as compared to colonic GI bleeding;
- whether the beneficiary received a blood transfusion during hospitalization, indicating likely more severe GI bleeding; and

- whether the beneficiary experienced any of a number of diagnoses or treatments associated with a higher risk of complications and cost outside of the clinician's control and not captured by the HCC model, including:
  - Angiography without embolization;
  - Left ventricular assist device (LVAD);
  - Portal hypertension;
  - Prior admission for principal diagnosis of rectal bleeding;
  - Recent GI bleed admission;
  - Recent major abdominal surgery (non-bowel);
  - Recent major bowel surgery;
  - Recent myocardial infarction;
  - Syncope;
  - Transfers; or
  - Frailty, as indicated by nursing physician facility visits, recent all-cause admission in the prior 30 days, dementia, use of home oxygen, use of a home hospital bed, or recent admission to a long-term care hospital.

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 0.5<sup>th</sup> 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 noted in the exclusions analysis above, extremely low- or high-cost outlier episodes with residuals below the 1<sup>st</sup> percentile or above the 99<sup>th</sup> 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.

Full details of the risk adjustment model are in the Measure Codes List File.<sup>20</sup> The National Summary Data Report (NSDR) Addendum includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.<sup>21</sup>

### 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 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 Lower GI Bleed 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

<sup>20</sup> CMS, "Lower Gastrointestinal Hemorrhage Measure Codes List," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.

<sup>21</sup> CMS, "National Summary Data Report Addendum: 11 Episode-Based Cost Measures and Revised MSPB Clinician Measure," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

outside of clinician control, or any other factors that would help prevent unintended consequences. These additional risk adjustors are listed in the section above.

### **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 peer-reviewed literature and data analysis.

### **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., ACOs, previous physician QRUR programs, and other measures such as National Quality Forum (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 Pope et al (2011) report.<sup>22</sup> 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.

The results of the statistical analysis used to characterize our risk adjustment model can be found in the NSDR Addendum, which includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.

### **3.5.5 Analyses and Interpretation in Selection of Social Risk Factors**

Acumen analyzed gender, dual status, income, education, and unemployment as social risk factors (more information on these variables can be found in Section 3.1.8). Beneficiary 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.

55.4 percent of beneficiaries are female under this measure. The majority of the beneficiaries (76.7%) 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 33.3 percent of observations. While 3.1 percent of beneficiaries are classified below a high school education level, the vast majority (82.4%) of episodes are classified as greater than high school level. Finally, 26.1 percent of beneficiaries 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 gender, dual status, gender + dual status, and gender + dual + income + education + unemployment + race, on top of the adapted CMS-HCC model. 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.

First, we analyzed the model coefficients and p-values for each of the base and social risk factor models to understand whether any of the social risk factor covariates are predictive of episode cost. The T-test and F-test revealed many significant p-values, indicating that social risk factors

---

<sup>22</sup> 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.

are likely predictive factors for determining resource use among beneficiaries for the relevant characteristic.

Secondly, we analyzed the impact of adding social risk variables on overall model performance by looking at the differences in the O/E cost ratios with and without social factors in the risk adjustment model. When including social risk factors in our risk adjustment regression, the minor differences in the O/E cost ratios, even for providers at high or low extremes of risk, indicates that social risk factor effects on the model performance are likely captured through existing risk adjustment variables. When including the social risk factors in risk adjustment, the O/E cost ratio for 99.5 percent of TINs and 98.7 percent of TIN-NPIs changed by  $\pm 0.01$  or less.

Finally, 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 and TIN-NPI level, with a Spearman correlation coefficient of 0.999 at the TIN level and 0.998 at the TIN-NPI level. These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

Due to the limited impact of social risk factor effects under the current risk adjustment model, we believe the Lower GI Bleed measure risk adjustment model sufficiently accounts for the effects of social risk factor on clinician measure scores.

### **3.5.6 Method for Statistical Model or Stratification Development**

To analyze the validity of current risk adjustment model, we examined three analyses: (1) R-squared and adjusted R-squared for the regression models, (2) predictive ratios and O/E cost ratios to examine the fit of the models at different levels of patient complexity, and (3) coefficient estimates, standard errors, and p-values.

- 1) *R-squared and adjusted R-squared* were calculated for the overall measure. The results should be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted. This is an important distinction from all-cost measures, as a low R-squared does not necessarily indicate that a measure reflects variation unrelated to clinical care, while a high R-squared does not necessarily indicate the opposite; instead, the risk adjustment models must be evaluated in concert with 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.
- 3) *Coefficient estimates, standard errors, and p-values* were run to consider the extent to which the coefficients for the risk factor covariates are predictive of episode cost. Results for individual risk adjustment variables should be viewed in the context of the entire model, rather than being analyzed individually. For instance, coefficients indicate the incremental effect of a model variable, holding all other variables fixed. As another example, interactions between model variables must be interpreted in concert with the effects of those variables in isolation.

The results of these analyses are presented in the NSDR Addendum to aid in the overall assessment of the predictive ability of the risk adjustment models.<sup>23</sup>

### **3.5.7 Statistical Risk Model Discrimination Statistics**

The overall R-squared for the Lower Gastrointestinal Hemorrhage cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.29. The adjusted R-squared is also 0.29.

The NSDR Addendum also includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.<sup>24</sup>

### **3.5.8 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 average O/E cost ratio is generally close to one across risk deciles, indicating that the model is accurately predicting actual episode cost. Full results can be seen the NSDR Addendum.

### **3.5.9 Statistical Risk Model Calibration – Risk Decile**

Analysis of predictive ratios by risk decile for the measure shows that the model has consistent predictive ratios across risk score deciles, with each decile having a predictive ratio in the range of 0.99 to 1.01.

### **3.5.10 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.<sup>25</sup> 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 Sections 3.5.8 and 3.5.9, the average O/E cost ratios and the predictive ratios for all risk deciles are very close or equal 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, regardless of overall risk level

## **3.6 Identification of Meaningful Differences in Performance**

### **3.6.1 Method**

Our method of determining clinically meaningful differences in episode-based cost measure scores consists of stratifying the clinician measure scores by meaningful characteristics and investigating the clinician score distribution by percentile. Stratification is performed for each of the following characteristics: urban/rural, census division, census region, risk score, and the

---

<sup>23</sup> CMS, "National Summary Data Report Addendum: 11 Episode-Based Cost Measures and Revised MSPB Clinician Measure," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

<sup>24</sup> 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.

<sup>25</sup> Ibid.

number of episodes attributed to the clinician. We analyze the distribution of measure scores for clinicians defined by these characteristics, as well as for the overall episode group.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure scores among clinicians to meaningfully determine a 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 large performance difference among clinicians in the Lower GI Bleed measure:

- (i) the 99<sup>th</sup> percentile of the measure score is 1.46 and 1.56 times the 1<sup>st</sup> percentile at the TIN level and TIN-NPI levels, respectively, and
- (ii) the Lower GI Bleed measure score at the 90<sup>th</sup> percentile is 21.0 percent greater than the score at the 10<sup>th</sup> percentile at the TIN level and 26.1 percent greater at the TIN-NPI level.

These results indicate there is large potential for saving Medicare spending.

The results also show that there is not systemic regional difference in clinician score. For instance, the mean scores for clinicians across nine census divisions (excluding 'Unknown') are within a \$618 range (\$10,237 – \$10,855) at the TIN level and \$1,307 range (\$10,893 – \$12,200) at the TIN-NPI level. Similarly, clinicians and clinician groups in urban areas seem to perform comparably to those in rural areas, with a \$6 difference in mean measure scores at the TIN level and \$186 at the TIN-NPI level.

In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform similarly to those who treat fewer acute cases. 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 scores also show little variation by risk score decile at the TIN level, with a range in mean TIN score of \$10,129 to \$10,621, indicating that the risk adjustment model is functioning as intended at the more treatment-appropriate TIN level. Full results can be seen in the NSDR.<sup>26</sup>

### 3.6.3 Interpretation

There are clinically and practically significant variations in Lower GI Bleed measure scores, indicating the measure's ability to capture differences in performance. Our findings regarding variation in measure scores are consistent with expert clinician input. Overall, as expected, results show that clinicians are not being systematically penalized or rewarded due to risk score decile given the current Lower GI Bleed measure design (i.e., the differences in cost measure scores are not as a result of the risk profile of the patient cohort).

## 3.7 Missing Data Analysis and Minimizing Bias

### 3.7.1 Method

Since CMS uses Medicare claims data to calculate the Lower GI Bleed measure, Acumen expects a high degree of data completeness. To further ensure that we have complete and

---

<sup>26</sup> CMS, "National Summary Data Report: 11 Episode-Based Cost Measures and Two Revised Cost Measures, Updated Following Field Testing (Oct-Nov 2018)," *MACRA Feedback Page*, <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-feedback.html>.



accurate data for each beneficiary who opens an episode, Acumen excludes episodes where beneficiary date of birth information (an input to the risk adjustment model) cannot be found in the EDB, the beneficiary does not appear in the EDB, or the beneficiary death date occurs before the episode trigger date.

The Lower GI Bleed measure also excludes episodes where the beneficiary 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 beneficiary needed to capture the clinical risk of the beneficiary in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the beneficiary's care is covered under Medicare Part C.

### 3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the four categories of missing data which caused episodes to be excluded from the Lower GI Bleed 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:

- Beneficiary date of birth is missing
- Beneficiary death date occurred before the trigger date
- Beneficiary has a primary payer other than Medicare during the episode window or in the 120-day lookback period
- Beneficiary was not enrolled in Medicare Parts A and B, or was enrolled in Part C, during the 120-day lookback period and episode window

**Table 5: Missing Data Categories for the Lower Gastrointestinal Hemorrhage Measure**

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	0	0	0
Death before trigger	636	713	955
Other primary payer	22,648	6,810	38,044
Not continuously enrolled	18,072	4,043	16,856

### 3.7.3 Interpretation

As the Lower GI Bleed 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 address cases of missing or inaccurate information in claims data.



## 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 effect’s 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 potential 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 11 episode-based cost measures developed during 2018, including the Lower Gastrointestinal Hemorrhage measure, for a 35-day comment period (October 3 to November 5, 2018). We provided field test reports to a sample of clinician groups and clinicians.<sup>27</sup> Each report included information for all measures for which the clinician or clinician group was attributed 10 or more episodes. The 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 does not determine case minimums used for any potential program implementation.

- Total testing sample across all episode-based cost measures: 14,237 TINs; 63,984 TIN-NPIs
- Testing sample for Lower Gastrointestinal Hemorrhage: 2,133 TINs; 1,833 TIN-NPIs

All stakeholders, including those who did not receive a field test report, could review a mock field test report that was posted on the CMS website. 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.<sup>28</sup> During field testing, Acumen conducted education and outreach activities including a national webinar, office hours with specialty societies, and Help Desk support.

##### *5.1.2.2 Technical Assistance with Results*

##### **Field Testing**

During the feedback period, 2,388 field test reports for episode-based cost measures were downloaded by 403 clinician groups (TINs) and 1,985 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the 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 the following sheets:

- High-level summary results across all episode-based cost measures being field tested
- Results for each 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)

---

<sup>27</sup> The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

<sup>28</sup> 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>.

- Drill-down detail for each measure, including more detailed information on potential cost drivers in the TIN/TIN-NPI's episodes. For example:
  - 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)
- Episode-level table with detailed information for all episodes attributed to the TIN/TIN-NPI across all measures in the report
  - Data across six major categories: (i) episode costs, (ii) beneficiary information, (iii) attributed clinician(s), (iv) evaluation and management visits performed during episode, (v) Physician Fee Schedule costs to Medicare billed during episode, and (vi) other providers rendering care.

A mock field test report can be viewed on the CMS MACRA Feedback webpage.<sup>29</sup>

### 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, and other available listservs. More detail on this outreach can be found in the Field Test Summary Report on the CMS MACRA Feedback webpage.

Acumen and CMS hosted two office hour sessions in October 2018, to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Acumen also hosted two office hour sessions with members of Clinical Subcommittees and workgroups to provide an update on development and field testing. Across all four office hours sessions, there were over 100 attendees.

Acumen worked with the Physician Value helpdesk and QPP Service Center to answer stakeholder questions during field testing and continued to answer questions after the feedback period ended.

Acumen and CMS hosted a national field testing webinar on October 9, 2018 to provide an overview of the measures being field tested and the information available for public comment. The webinar consisted of an hour-long presentation, outlining (i) the cost measure development activities, (ii) field testing activities, (iii) how to access and understand the confidential field test reports, and (iv) the contents of the reports. The presentation was followed by a 30-minute Q&A session. Around 85 comments and questions were received via webinar chat and on the phone.

A post-field testing webinar was held on March 27, 2019 to provide an update on the measures following field testing. The webinar consisted of a 60 minute presentation providing an overview of the basics of measure construction, highlighting refinements made after field testing, and summarizing the testing done on the measures. This presentation was followed by a Q&A session.<sup>30</sup>

---

<sup>29</sup> CMS, "Episode-based Cost Measures Mock Field Test Report," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2018-Mock-report-for-Episode-Based-Cost-Measures.xlsx>.

<sup>30</sup> CMS, Webinar Recordings, Slides, and Transcripts, *QPP Webinar Library*, <https://qpp.cms.gov/about/webinars>.

### **5.1.2.3 Feedback on Measure Performance and Implementation**

#### **Field Testing**

In total, Acumen received 67 survey responses and 25 comment letters, including many 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 37 comments on the 11 episode-based cost measures included in the Measures Under Consideration List released in December 2018. This included four comments for the Lower Gastrointestinal Hemorrhage cost measure. After the Measure Applications Partnership (MAP) Clinician Workgroup meeting in December 2018, there was another public comment period on their preliminary recommendations, which received 23 comments across the 11 measures, with two comments specific to the Lower Gastrointestinal Hemorrhage cost measure.<sup>31</sup> 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 all feedback gathered during the field testing period. The following list synthesizes some of the key points that were raised through the field testing feedback period:

- *Stakeholder engagement and involvement remains an important aspect of the measure development process.* Stakeholders expressed appreciation for the opportunity to provide feedback during field testing and for CMS' continued efforts to involve them in the measure development process. Commenters also valued the decision to operationalize previously collected feedback, as demonstrated through the addition of measure-specific workgroups to the development process.
- *Field test reports present useful information for understanding clinician performance, though reduced complexity could encourage more clinician participation.* Stakeholders praised the presentation and content of the field test reports. However, the complexity of the information presented in the reports was a challenge for some stakeholders.
- *Improved supplemental field testing materials are helpful but can be further refined.* Some stakeholders found the supplemental field testing materials to be informative and thorough, providing useful information on field testing and the specifications of the cost measures. However, many noted that although the materials are comprehensive, they remain lengthy and complex, and they believe the amount of information provided is too overwhelming to be useful.
- *Ample time for review of field testing reports and materials is vital to collecting meaningful stakeholder feedback.* Some stakeholders suggested the field testing period be extended or kept open, given the large amount and complexity of the information that was presented.
- *Transparent Clinical Subcommittee and measure-specific workgroup selection and voting encourages buy-in from stakeholders.* Some stakeholders expressed concern with the selection and voting processes for the Clinical Subcommittees and workgroups,

---

<sup>31</sup> Measure Applications Partnership, *National Quality Forum*, [https://www.qualityforum.org/Setting\\_Priorities/Partnership/Measure\\_Applications\\_Partnership.aspx](https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx).

highlighting that a transparent approach to member selection would ensure an appropriate mix of specialties and clinician types.

- *Field test report access continues to present challenges for stakeholders.* Some stakeholders noted that they faced difficulties creating accounts and downloading their field test reports from the CMS Enterprise Portal and these challenges may have negatively impacted the number of clinicians that were able to participate in field testing. Stakeholders urged CMS to communicate directly with clinicians receiving field test reports and to find an alternative for delivering and accessing the reports.

The report additionally contains measure-specific feedback, which was used as the basis for the post-field testing refinements that were made to the measures, summarized below:

- Refinements to trigger codes, attribution, episode windows, assigned services, risk adjustment variables, exclusions, and alignment of cost with quality
- Adding/removing certain trigger codes and assigned services and revising the attribution methodology
- Stakeholders also noted that the level of clinician engagement in the development of these episode-based cost measures is a significant improvement over the development process for earlier cost measures.

#### **5.1.2.5 Feedback from Other Users**

##### **Pre-Rulemaking**

The MAP recognized the importance of cost measures to the MIPS program and conditionally supported the Lower Gastrointestinal Hemorrhage cost measure pending NQF endorsement. Specifically, the MAP encouraged the NQF endorsement Cost and Efficiency Standing Committee to consider the appropriateness of the risk adjustment model to ensure clinical and social risk factors are reviewed and included when appropriate. MAP cautioned about the potential stinting of care and noted that appropriate risk adjustment could help safe guard against this practice. The MAP also encouraged the Standing Committee to examine the exclusions in this measure to ensure appropriate attribution.

#### **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 an expert clinician 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 expert clinician 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 Lower Gastrointestinal Hemorrhage measure made after consideration of field testing analyses and stakeholder feedback are:

- **Episode Window:** Shortened post-trigger period to 35 days
- **Triggers:** Modified triggering logic to better focus on lower GI bleeding episodes, as follows:
  - Removed diagnosis codes K550 and K551 as trigger codes for the medical DRGs 377-379
  - For all trigger logic, decided to look only at the principal diagnosis on the trigger claim

- Added trigger exclusions to more comprehensively identify upper GI bleeding: diagnosis codes K5501, K5502, K221, K20, K319, K3189, I85, I864, K228, K226, K838, and K8689
- Added a trigger exclusion logic set for interventional radiology DRGs 356-358 with the same rules as for the medical DRGs 377-379
- **Exclusions:** Removed exclusion for colon cancer patients
- **Service Assignment:** Added new assigned services, as follows:
  - Esophagogastroduodenoscopy within 35 days
  - Services related to MI within 15 days
- **Risk Adjustment:** Added new risk adjustors for:
  - Angiography without embolization
  - Blood transfusion receipt during hospitalization within 48 hours
  - Current admission for rectal bleeding
  - The following frailty indicators:
    - Dementia
    - Home hospital bed
    - Home oxygen
    - Nursing physician facility visits
    - Recent admission to a long-term care hospital
    - Recent all-cause admission within 120 days

## 5.2 Usability

### 5.2.1 Improvement

n/a. The measures have not yet been implemented, and as such have 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.

## **6.0 Related and Competing Measures**

### **6.1 Relation to Other Cost Measures**

There are currently no related NQF-endorsed or non-NQF-endorsed cost measures that address this same measure focus or target population. There are no competing NQF-endorsed or non-endorsed cost measures that address both this same measure focus *and* at this same target population.

### **6.2 Harmonization**

n/a

### **6.3 Competing Measures**

n/a



## Contact Information

### Measure Steward Point of Contact

Organization: Centers for Medicare & Medicaid Services

Name: Joel Andress

Email Address: [joel.andress@cms.hhs.gov](mailto:joel.andress@cms.hhs.gov)

Phone Number: (410) 786-5237

### Developer Point of Contact

Organization: Acumen, LLC

Email Address: [macra-cost-measures-info@acumenllc.com](mailto:macra-cost-measures-info@acumenllc.com)

Phone Number: (650) 558-8882

### Other Additional Information

#### ***Lower Gastrointestinal Hemorrhage Measure-Specific Workgroup Members:***

Ammar Sarwar, Society of Interventional Radiology

C. Matthew Hawkins, Society of Interventional Radiology

Caroll Koscheski, American College of Gastroenterology

Colleen Schmitt, American Society for Gastrointestinal Endoscopy

Eric Haas, American Society of Colon and Rectal Surgeons

Gene Lambert, Society of Hospital Medicine

J. Brent Box, Society of Hospital Medicine

James Richter, American Society for Gastrointestinal Endoscopy

Jennifer Broder, American College of Radiology

Lauren Beste, American Gastroenterological Association

Shazia Siddique, American Gastroenterological Association

Susan Nedza, American College of Emergency Physicians

Walter Peters, American Society of Colon and Rectal Surgeons

The Lower Gastrointestinal Hemorrhage workgroup is composed of members from the larger Gastrointestinal Disease Management – Medical and Surgical Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the [Episode-Based Cost Measures Development Process document](#).<sup>32</sup>

---

<sup>32</sup> CMS, “Episode-Based Cost Measure Field Testing Measure Development Process,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2018-measure-development-process.pdf>