

Femoral or Inguinal Hernia Repair Measure

Measure Justification Form

June 2019



Contents

1.0	Introduction	4
1.1	Project Title and Overview	4
1.2	Measure Name	4
1.3	Type of Measure	4
2.0	Importance.....	5
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
3.0	Scientific Acceptability	8
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	17
3.5.1	Method of Controlling for Differences	17
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	Results of Risk Stratification Analysis.....	22
3.5.11	Interpretation.....	23
3.6	Identification of Meaningful Differences in Performance.....	23

3.6.1 Method	23
3.6.2 Statistical Results	23
3.6.3 Interpretation	24
3.7 Missing Data Analysis and Minimizing Bias.....	24
3.7.1 Method	24
3.7.2 Missing Data Analysis	25
3.7.3 Interpretation	25
4.0 Feasibility.....	26
4.1 Data Elements Generated as Byproduct of Care Processes	26
4.2 Electronic Sources	26
4.3 Data Collection Strategy	26
4.3.1 Data Collection Strategy Difficulties.....	26
5.0 Usability and Use	27
5.1 Use	27
5.1.1 Current and Planned Use	27
5.1.2 Feedback on the Measure and Development Process	27
5.2 Usability	31
5.2.1 Improvement	31
5.2.2 Unexpected Findings.....	31
5.2.3 Unexpected Benefits	31
6.0 Related and Competing Measures	32
6.1 Relation to Other Cost Measures	32
6.2 Harmonization.....	32
6.3 Competing Measures	32
Contact Information	33

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Femoral or Inguinal Hernia Repair measure. The MJF is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology document and Measure Codes List file, which together comprise the specifications for this cost measure.¹

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

Femoral or Inguinal Hernia Repair Episode-based Cost Measure

1.3 Type of Measure

Cost/Resource Use

¹ CMS, “Femoral or Inguinal Hernia Repair 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, “Femoral or Inguinal Hernia Repair 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 Femoral or Inguinal Hernia Repair cost measure (also referred to as “the Hernia Repair measure”) evaluates clinicians’ risk-adjusted cost to Medicare for beneficiaries who receive surgical procedures to repair a femoral or inguinal hernia. The cost measure score is a clinician’s average risk-adjusted cost for the episode group across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician’s role in managing care during the 30 days prior to the clinical event that opens or ‘triggers’ the episode, through 90 days after the trigger. Beneficiary populations eligible for the Hernia Repair 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.² 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, this measure represents an area where there are opportunities for improvement. These opportunities for improvement are primarily found within the variation in hernia repair techniques, where standardization of techniques and protocols may improve outcomes and lower the cost of care during the episode.

Abdominal wall hernia repair surgery, including femoral and inguinal hernia repair, is a surgical procedure that can be treated by either suturing the defect closed or placing a synthetic mesh over the defect without tension. The treatment can be done laparoscopically or through an open surgical operation, which each have their own merits.³ The Lichtenstein technique is an example of an open surgical method which uses mesh, and is widely used because it is considered relatively easy to perform, low risk, and low cost. Laparoscopic repair is considered more complicated to perform, tends to cost more, and may lead to more complications, though recovery time and pain are reduced.⁴ Inguinal hernia repair treated laparoscopically is frequently completed using a transabdominal preperitoneal (TAPP) procedure and a meta-analysis of randomized controlled trials found that the TAPP procedure resulted in significantly less inguinal pain postoperatively compared to the Lichtenstein technique. In a paper analyzing data from 8 randomized studies involving 425 patients who received TAPP repair and 411 patients who received a Lichtenstein repair, researchers found that the patients who received the TAPP

² 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.

³ Vale, Luke, Adrian Grant, et al., “Cost-Effectiveness of Alternative Methods of Surgical Repair of Inguinal Hernia.” International Journal Of Technology Assessment In Health Care 20, no. 2 (2004): 192-200

⁴ Arregui, Maurice E., and Susan B. Young. “Groin Hernia Repair by Laparoscopic Techniques: Current Status and Controversies.” World Journal Of Surgery 29, no. 8 (2005): 1052-57.

procedure had less pain in the 12 hours after surgery and less chronic pain, with no difference in complications between the two procedures.⁵ According to a 2009 study, although laparoscopic treatment of hernia repair is influenced by clinical decisions, financial considerations may also influence the choice of surgical approach. The study of 58,172 inguinal hernia repairs in Florida in 2002 and 2003 found that only 11,351 (19.5 percent) were performed laparoscopically and their cost was significantly higher (\$12,087 for laparoscopic repairs compared with \$7,580 for open repairs).⁶

Other areas of variation in elective femoral or inguinal hernia repair include bilateral hernia repair and patient comorbidities. Data from the German Herniated Registry with data on 15,176 open hernia repairs, between 2009 and 2014, revealed a significantly higher percentage of postoperative complications in patients who had bilateral procedures compared to unilateral procedures (1.9 percent compared to 0.9 percent, with an odds ratio of 2.13).⁷ This information was developed to guide surgeons who propose bilateral surgery on a unilateral inguinal hernia as prophylaxis for the hernia developing on the other side.⁸ Endoscopic hernia repair tends to reduce the risk of secondary bleeding and complication-related reoperation when a delicate dissection technique is employed, compared to open procedures. Bilateral hernia operations also had a higher chance of requiring a complication-related reoperation.⁹

2.2 Performance Gap

2.2.1 Rationale

In the US, more than 1 million hernias are repaired annually, the majority of which are inguinal hernias.¹⁰ On average, these hernia repair procedures cost approximately \$2,000 to \$2,500, representing nearly \$2.5 billion in annual health care costs.¹¹ Opportunities for improvement for elective femoral or inguinal hernia repair are primarily found within the variation in hernia repair techniques, where utilization of optimal techniques and protocols may improve outcomes and lower the cost of care during the episode.

The Femoral or Inguinal Hernia Repair 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

⁵ Scheuermann, Uwe, Stefan Niebisch et al., "Transabdominal Preperitoneal (Tapp) Versus Lichtenstein Operation for Primary Inguinal Hernia Repair - a Systematic Review and Meta-Analysis of Randomized Controlled Trials." *BMC Surgery* 17, no. 1 (2017): 55-55.

⁶ Smink, Douglas S., Ian M. Paquette et al., "Utilization of Laparoscopic and Open Inguinal Hernia Repair: A Population-Based Analysis." *Journal Of Laparoendoscopic & Advanced Surgical Techniques. Part A* 19, no. 6 (2009): 745-48.

⁷ Jacob, D. A., JA Hackl et al., "Perioperative Outcome of Unilateral versus Bilateral Inguinal Hernia Repairs in TAPP Technique: Analysis of 15,176 Cases from the Herniated Registry." *Surgical Endoscopy*, vol. 29, no. 12, 2015, pp. 3733–3740., doi:10.1007/s00464-015-4146-5.

⁸ Köckerling, Ferdinand, Christine Schug-Pass et al., "Bilateral and Unilateral Total Extraperitoneal Inguinal Hernia Repair (TEP) Have Equivalent Early Outcomes: Analysis of 9395 Cases." *World Journal of Surgery*, vol. 39, no. 8, 2015, pp. 1887–1894., doi:10.1007/s00268-015-3055-z.

⁹ Köckerling, Ferdinand, Andreas Koch et al., "Open Repair of Primary Versus Recurrent Male Unilateral Inguinal Hernias: Perioperative Complications and 1-Year Follow-Up." *World Journal of Surgery*, vol. 40, no. 4, 2015, pp. 813–825., doi:10.1007/s00268-015-3325-9.

¹⁰ Matthews, R. Douglas, and Leigh Neumayer. "Inguinal Hernia in the 21st Century: An Evidence-Based Review." *Current Problems In Surgery* 45, no. 4 (2008): 257-59.

¹¹ Rutkow, Ira M. "Demographic and Socioeconomic Aspects of Hernia Repair in the United States in 2003." *The Surgical Clinics Of North America* 83, no. 5 (2003): 1045.

Subcommittee, the subsequent measure-specific workgroup provided extensive, detailed input on this measure.

2.2.2 Performance Scores

Performance scores are provided for 2,249 clinician group practices (identified by Tax Identification Number [TIN]) and 3,328 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 MIPS eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced APM participation). This table uses a testing volume threshold of 10 episodes.

Table 1. Distribution of Performance Scores

Metric	TIN	TIN-NPI
Mean score	\$4,111	\$4,102
Standard deviation	\$402	\$454
Score IQR	\$374	\$439
Score percentile		
10 th	\$3,511	\$3,387
20 th	\$3,868	\$3,803
30 th	\$4,059	\$4,037
40 th	\$4,161	\$4,163
50 th	\$4,218	\$4,240
60 th	\$4,272	\$4,300
70 th	\$4,324	\$4,353
80 th	\$4,381	\$4,416
90 th	\$4,479	\$4,516

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 Femoral or Inguinal Hernia Repair 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 Femoral or Inguinal Hernia Repair episodes ending from January 1, 2017 through 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

2,249 clinician group practices and 3,328 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 10 or more Hernia Repair episodes during the measurement period. Episodes from all 50 States and D.C. in the following settings were included: Acute inpatient (IP) hospitals, hospital outpatient departments (HOPD), ambulatory/office-based care centers, and ambulatory surgical centers (ASC).

3.1.6 Patient Cohort Included in the Testing and Analysis

74,964 Medicare beneficiaries (from 75,531 episodes) were included in TIN level testing and analysis, and 50,307 beneficiaries (from 50,678 episodes) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the Hernia Repair measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who undergo surgical procedure to repair a femoral or inguinal hernia that triggers a Femoral or Inguinal Hernia Repair episode. 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 hernia repair procedures.

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.
- The episode can be attributed to at least one main clinician.
- The episode trigger claim was in an inpatient, office, HOPD, or ASC setting based on its place of service.
- Where there is a concurrent inpatient stay with the episode trigger code, it occurs in a short-term stay acute hospital as defined by subsection (d).¹²
- Where the beneficiary had an emergency department visit on or one day prior to the trigger date, which suggests emergency hernia repair.
- The episode trigger procedure was not a part of a concurrent major surgery.
- Where there is a concurrent inpatient stay with the trigger, the inpatient stay has a Medicare Severity Diagnosis Related Group (MS-DRG) relevant to femoral or inguinal hernia (MS-DRG 350, 351, 352).
- The episode is not an outlier case.

To determine whether the Hernia Repair 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 Hernia Repair measure's inclusion criteria have a small effect on the percentage of beneficiaries of any particular demographic or patient characteristic. The difference between beneficiaries being included or not included in the measure is less than 2.7 percentage points across each of the characteristics in the analysis at TIN level testing, and less than 5.5 percentage points at TIN-NPI level testing. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 26.2 percent, compared to 25.8 percent at TIN level testing. The difference in the percentage of beneficiaries for race, with and without the inclusion criteria, is within 2 percentage points for all categories at both TIN and TIN-NPI level testing, with the exception of white. For the white beneficiary category, the difference is 1.8 and 2.9 percentage points at the TIN and TIN-NPI level, respectively. The share of female beneficiaries remains close when comparing the use of inclusion criteria, with 13.5 percent without inclusion criteria, compared to 11.3 percent and 11.1 percent at TIN and TIN-NPI level testing, respectively, when inclusion criteria are applied. Similarly, the share of male beneficiaries changes minimally when inclusion criteria are applied, from 86.5 percent without inclusion criteria to 88.7 percent and 88.9 percent at TIN and TIN-NPI level testing, respectively with inclusion criteria. These results indicate that there is minimal shift in patient characteristics when applying the inclusion criteria listed above at both TIN and TIN-NPI level testing.

¹² Only stays at IP facilities that are 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 Hernia Repair 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.¹³ The fiscal year 2018 Medicare fee-for-service program proper payment rate was 91.9 percent.¹⁴ 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.

3.2.3 Statistical Results from Reliability Testing

Measure Reliability

At the 10, 20, and 30 episode thresholds, 100 percent of TINs and TIN-NPIs have mean reliability greater than or equal to 0.4. At a volume threshold of at least 10 episodes, the mean reliability is 0.86 for TINs and 0.81 for TIN-NPIs. 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	0.86	100.0%	0.81	100.0%
20	0.91	100.0%	0.88	100.0%
30	0.94	100.0%	0.91	100.0%

¹³ 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>

¹⁴ Ibid.

3.2.4 Interpretation

Measure Reliability

Overall reliability of the Hernia Repair measure is very high at a volume threshold of 10 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.¹⁵

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 systematic assessment of face validity and empirical validity testing.

3.3.2 Method of Validity Testing

Face Validity

The Hernia Repair 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 Femoral or Inguinal Hernia Repair 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 52 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 the measure to develop, scope, and suggested composition of a smaller, targeted workgroup to provide detailed input on each aspect of measure specifications. The Femoral or Inguinal Hernia Repair workgroup was composed of nine members, affiliated with eight specialty societies, including American College of Surgeons, American Society of Colon and Rectal Surgeons, and Society of Abdominal Radiology. 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 caregiver perspective. PFC input included concepts of healthcare quality and value, guiding principles and measure-specific input to inform the workgroups such as pre- and post-

¹⁵ 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.

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 hernia repair, thus preventing inclusion of unrelated cost variation in this measure. Assigned services occurring in the outpatient and clinician service setting were defined separately for the pre- and post-trigger windows, and include only services directly related to treatment, follow-up, or complications from an inpatient hospitalization for femoral or inguinal hernia repair. Services in the emergency department, home health, inpatient medical and surgical, inpatient rehabilitation facility and durable medical equipment categories are assigned only in the post-trigger period so as to capture additional services related to the surgery that the beneficiary might require.

Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the Hernia Repair 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 the procedure. For this analysis, we compared the ratio of observed to expected cost (henceforth called the "O/E cost ratio") for Femoral or Inguinal Hernia Repair episodes with and without complications related to the procedure that occur in the post-trigger period. This analysis sought to confirm the expectation that the Hernia Repair 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 Hernia Repair measure were classified into clinically coherent groups of services, called "clinical themes." The Hernia Repair measure clinical themes are:

- **Preoperative Work-up:** Routine chest x-rays; electrocardiograms; laboratory testing, such as blood tests to assess coagulation; other diagnostic techniques, or diagnostic procedures, such as office or outpatient visits.
- **Major Abdominal Surgery:** Outpatient and inpatient procedures including, small bowel resection, laparoscopy, major and minor small and large bowel procedures, lysis of adhesions or other procedures.
- **Repeat Hernia Repair and Hernia Treatment:** Services to address repairs or treatment for hernias after the initial surgery, such as hernia repairs, CT scans, diagnostic ultrasounds, MRIs, diagnostic radiology, diagnostic or therapeutic procedures, ancillary services, and supplies related to the procedures.
- **Peritonitis/Abdominal Pain/Ileus:** Inpatient and outpatient hospital care, including critical care related to intraoperative and post procedural complications, intestinal disorders, and pain, including imaging, diagnostic radiology and procedures, therapeutic procedures, and ancillary services.
- **Urinary Tract Infection (UTI):** Inpatient and outpatient hospital care related to cystitis, hematuria, urine retention, and other disorders, including CT scans, diagnostic ultrasounds, diagnostic radiology and procedures, laboratory testing, therapeutic procedures, and ancillary services.

- **Sepsis:** Inpatient and outpatient hospital care including critical care related to bacterial infections and sepsis, including imaging, diagnostic radiology and procedures, therapeutic procedures, and ancillary services.
- **Cardiac Complications:** Inpatient and outpatient hospital care including emergency department visits or critical care related to cardiac diagnoses, including CT scans, echocardiograms, diagnostic radiology and procedures, therapeutic procedures, ancillary services, cardiac defibrillator implant, coronary bypass, and percutaneous cardiovascular procedures within the appropriate time frame.
- **Pulmonary/Respiratory Complications:** Inpatient and outpatient hospital care including emergency department visits or critical care related to asthma, bronchitis, pneumonia or other respiratory disorders and infections, including CT scans, diagnostic ultrasounds, diagnostic radiology and procedures, and therapeutic procedures.
- **Renal Failure and Electrolyte Abnormality:** Inpatient and outpatient hospital care including emergency department visits or critical care related to volume depletion, kidney failure, or other disorders of fluid, electrolyte and acid-base balance, including CT scans, diagnostic ultrasounds, diagnostic radiology and procedures, therapeutic procedures.
- **Surgical Site Complication:** Inpatient and outpatient hospital care including emergency department visits or critical care related to intraoperative or post-procedural hemorrhage or hematoma or other complications of the skin or tissue, including imaging, diagnostic or therapeutic procedures, and wound care.

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 calculated the Pearson correlation between the cost of each clinical theme and the overall risk-adjusted cost for an episode.

We expected that the Renal Failure and Electrolyte Abnormalities theme would have the highest correlation with risk-adjusted episode cost, as complications are likely associated with high cost even after accounting for beneficiary characteristics.¹⁶ We would expect similar trends for the Surgical Site Complication theme as it contains services relating to complications, such as treatment for cellulitis and postoperative infections. By contrast, we expected that the Preoperative Work-up theme would have a smaller correlation to risk-adjusted cost, as variations in associated cost would not necessarily reflect downstream costs.

3.3.3 Statistical Results from Validity Testing

For the first analysis of validity, the mean O/E cost ratio for all episodes is 1.0. The mean O/E cost ratio for episodes with services relating to complications during the post-trigger period is 1.11, compared with 0.99 for episodes without services relating to complications during the post-trigger period. Table 3, which provides a percentile distribution of the O/E cost ratios for all episodes and for episodes with and without services relating to complications from hernia repair, indicates that there is more variation in the O/E cost ratios for episodes with complications, compared to episodes without complications.

¹⁶ Khan, N.A., Quan, H., 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.

Table 3: 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	1.00	0.20	0.25	0.63	0.70	0.96	1.04	1.10	1.15	1.23	1.57
Episodes with Services Related to Hernia Repair Complications	1.11	0.26	0.30	0.74	0.90	1.02	1.10	1.19	1.35	1.58	2.01
Episodes without Services Related to Hernia Repair Complications	0.99	0.19	0.25	0.63	0.69	0.95	1.04	1.09	1.14	1.18	1.47

For the second analysis, results indicated that there is a strong correlation between costs for the Renal Failure and Electrolyte Abnormalities (correlation: 0.59) and the Major Abdominal Surgery (correlation: 0.50) clinical themes and risk-adjusted cost. The Surgical Site Complication theme (correlation: 0.48) also strongly correlates to risk adjusted cost. By contrast, the Preoperative Work-up theme (correlation: 0.16) had lower 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. This result demonstrates that the Hernia Repair measure is able to capture, accurately, higher resource use.

The clinical themes analysis demonstrates that high risk-adjusted cost is strongly associated with themes related to complications and renal failure and weakly correlated with themes relating to preoperative services, as expected. This indicates that the measure may penalize clinicians who have higher rates of complications, while not disincentivizing the provision of appropriate pre-operative care, such as laboratory tests, physical examinations and imaging, where appropriate. Importantly, we see that correlation with risk-adjusted cost is strong not only for high-cost themes such as Major Abdominal Surgery (average cost: \$1,743), but also for lower cost themes such as Urinary Tract Infection (average cost: \$332). This indicates that the correlation does not come from a mechanical increase in episode costs from high-cost themes.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Hernia Repair measure either to capture a homogenous patient population within the scope of the measure focus on femoral or inguinal hernia repairs and ensure 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 were excluded for all measures due to the potential to reflect inaccurately 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 received emergency treatment on the trigger day or during a 1-day lookback period*
 - These episodes were excluded, as emergent surgery is likely to incorporate bowel strangulation, which would increase the patient's risk for complications and more costly services outside of the attributed clinician's influence.
- *Episodes where the triggering procedure was during a concurrent major surgery.*
 - These episodes were excluded as a concurrent major surgery is likely to lead to more downstream costs not directly related to the hernia repair.
- *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 substantially higher or lower 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 ratio of observed to expected cost (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 Hernia Repair measure is provided in the Measure Codes List.¹⁷

3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics and observed to expected cost ratios for the Hernia Repair measure exclusions. Cost statistics are also provided for the set of final episodes included in the Hernia Repair measure for comparison, with a testing volume threshold of 10 episodes at the TIN and TIN-NPI levels.

Table 4: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost			O/E		
	#	%	Mean	Percentile		Mean	Percentile	
				10 th	90 th		10 th	90 th
All Episodes Meeting Triggering Logic	103,506	100.00%	\$4,783	\$2,355	\$5,782	1.04	0.64	1.17
Beneficiary Death in Episode	1,413	1.37%	\$10,697	\$1,692	\$28,256	1.67	0.38	3.73
Emergent Hernia Repair	8,347	8.06%	\$11,525	\$3,312	\$24,392	1.58	0.72	3.46
Concurrent Major Surgery	7,067	6.83%	\$6,240	\$448	\$17,505	1.33	0.11	3.93
Outlier Cases	1,750	1.69%	\$10,619	\$609	\$20,889	1.66	0.14	3.70
<i>Final Episodes (TIN)</i>	75,531	72.97%	\$4,077	\$2,616	\$5,266	0.99	0.73	1.03
<i>Final Episodes (TIN-NPI)</i>	50,678	48.96%	\$4,045	\$2,482	\$5,253	0.98	0.73	1.03

¹⁷ CMS, "Femoral or Inguinal Hernia Repair 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>.

3.4.3 Interpretation

The statistical results indicate that the majority of excluded episodes differ substantially in both mean observed cost and mean O/E cost ratio and have larger variation compared to the final set of episodes and 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 each exclusion is provided below.

Episodes ending in death: The difference between mean observed episode cost for episodes ending in death and the final set of episodes is very large: \$10,697 compared to \$4,077 at TIN level testing and \$4,045 at TIN-NPI level testing. The mean O/E cost ratio for these episodes (1.67) is substantially higher than the mean O/E cost ratio for final episodes at both TIN-level testing (0.99) and TIN-NPI level testing (0.98). Finally, the O/E cost ratio ranges from 0.38 at the 10th percentile to 3.73 at the 90th percentile for episodes ending in death, compared to 0.73 at the 10th percentile and 1.03 at the 90th percentile for final episodes. These results suggest that there can be high variation in perimortem costs resulting in higher cost episodes, despite the episodes being by definition truncated. As such, episodes ending in death are excluded to avoid the potential of clinicians being incentivized to avoid treating complex, high-risk patients.

Episodes with emergent hernia repair: Hernia repairs performed emergently are much more costly than non-emergent hernia repair procedures (mean observed cost of \$11,525 compared to \$4,077 for final episodes at TIN level testing and \$4,045 at TIN-NPI level testing) and even after risk adjustment (mean O/E cost ratio of 1.58 compared to 0.99 and 0.98 for TIN-level and TIN-NPI level testing, respectively). The distribution of observed cost (\$3,312 at the 10th percentile and \$24,392 at the 90th percentile) and O/E cost ratios (0.72 at the 10th percentile and 3.46 at the 90th percentile) indicate high variation in the costs for emergent repairs. These results suggest that episodes with emergent hernia repairs represent highly complex cases that are different from both overall hernia repair episodes and the final set of episodes. As such, emergent hernia repair episodes are excluded to ensure the measure captures a homogeneous case-mix.

Episodes where the triggering procedure was during a concurrent major surgery. Episodes where the triggering service occurred during a concurrent major surgery are \$2,100 more expensive than mean observed cost for final episodes at both TIN and TIN-NPI testing levels. Though the mean observed cost is not substantially higher than for final episodes, there is wider variation in cost: observed costs range from \$448 at the 10th percentile to \$17,505 at the 90th percentile for these episodes compared to \$2,616 to \$5,266 and \$2,482 to \$5,253 for final episodes at the TIN and TIN-NPI levels, respectively. Similarly, there is a wide range in the O/E cost ratios for these episodes compared to final episodes. As such, these episodes were excluded to ensure a homogeneous case mix.

Outlier cases: The O/E cost ratio for outlier cases ranges from 0.14 at the 10th percentile to 3.70 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 107 risk factors and stratification by two risk categories.

The risk adjustment model for the Femoral or Inguinal Hernia Repair 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 received a same-day or staged bilateral hernia repair to account for the inherently higher cost of bilateral procedures relative to unilateral procedures.
- Whether the beneficiary had an emergency department visit for inguinal or femoral hernia within a 120-day lookback, as this indicates a more complex patient.
- Whether the beneficiary had dementia, was obese, or had smoking/nicotine dependence, as these beneficiary characteristics can lead to increased episode costs and are factors outside the reasonable influence of the attributed clinician.
- Whether the beneficiary was admitted to an inpatient hospital during a 120-day lookback period, and discharged within 30 days prior to the episode trigger as a recent prior inpatient stay can indicate a patient with comorbidities.
- Whether the episode is for a recurrent hernia repair, to account for the increased complexity of beneficiaries with recurrent hernias.
- Whether the beneficiary had an incarcerated or strangulated hernia repair. While emergent surgeries that might present strangulated hernia repairs were excluded, incarcerated hernias are included and risk-adjusted to address the difference in cost and outcome.

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 noted 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 performed separately for each of the two Hernia Repair measure sub-groups, which are based on the type of surgical approach.

- Laparoscopic Repair
- Open Repair

Full details of the risk adjustment model are in the Measure Codes List File.¹⁸ The National Summary Data Report (NSDR) Addendum includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.¹⁹

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 Hernia Repair measure methodology.

The workgroup provided input on measure-specific risk adjusters 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 adjusters are listed in the section above.

As previously noted, the risk adjustment model is run on episodes stratified into sub-groups, which may qualify as "ordering" of risk factors. Sub-groups were also determined based 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 sub-groups are listed in the above section. Hernia repair procedures performed laparoscopically versus via open surgery were separated into sub-groups to apply the risk adjustment model to similar surgeries and to avoid incentivizing one type of procedure over the other solely based on cost. The concern is that while laparoscopic may be more expensive, it provides benefits that cannot be fully captured within claims, such as faster return to work and slight improvement in quality of life.

¹⁸ CMS, "Femoral or Inguinal Hernia Repair 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>.

¹⁹ 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>.

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 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.²⁰ 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 adjusters and sub-groups.

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.

The percentage of female beneficiaries ranges from 9.8 to 12.2 percent across the two sub-groups in this measure. The majority of the beneficiaries (89.0 percent for Open Repair and 91.5 percent for Laparoscopic Repair) 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. Less than 2.2 percent of beneficiaries across both sub-groups are classified with a below high school education level, and the large majority of beneficiaries (over 86 percent for both sub-groups) are classified as having a level of education higher than high school. Finally, 18.7 and 20.7 percent of beneficiaries in the Open Repair and Laparoscopic Repair sub-groups, respectively, 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 adjusters. 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

²⁰ Pope, Gregory C., John Kautter, et al., "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. However, the analysis also shows that the significance of the effects of social risk factors is not consistent. For example, coefficients for female beneficiaries are statistically significant for the laparoscopic repair sub-group but not for the open repair sub-group.

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 difference between O/E cost ratios changed by ± 0.03 or less for 99.9 percent of TINs and 99.9 percent of TIN-NPIs.

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, with Spearman correlation coefficients of 0.996 at the TIN level and 0.997 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 inconsistent significance and limited impact of social risk factor effects under the current risk adjustment model, we believe the Hernia Repair measure's 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 for each sub-group.

- 1) *R-squared and adjusted R-squared* were calculated for the measure overall as well as for each sub-group. 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 for each sub-group 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 and set of sub-groups, 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.²¹

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Hernia Repair cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.42. The adjusted R-squared is also 0.42.

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. 2011.²²

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 are presented in 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 between 0.99 and 1.01.

3.5.10 Results of Risk Stratification Analysis

Results indicate that the two measure sub-groups have varying measure scores (see below table). Specifically, procedures performed laparoscopically are more expensive than those performed using open surgery. At the TIN level, the mean score for laparoscopic repair episodes is \$4,924 compared to open repair episodes at \$3,749. Results are similar at the TIN-NPI level. Evaluating performance on laparoscopic cases separately from open surgery cases ensures that this demonstrated cost variability does not influence clinical decisions concerning procedure choice. Stratifying episodes into these sub-groups, helps ensure meaningful comparison of clinician resource use.

Table 5: Distribution of Score by Sub-Group

Level	Sub-group	Provider Count	Mean Score	Score Percentile						
				1st	10th	25th	50th	75th	90th	99th
TIN	All TINs	2,249	\$4,111	\$2,802	\$3,511	\$3,977	\$4,218	\$4,351	\$4,479	\$4,795
TIN	Laparoscopic Repair	1,780	\$4,924	\$3,083	\$4,234	\$4,783	\$5,059	\$5,215	\$5,357	\$5,878
TIN	Open Repair	2,209	\$3,749	\$2,501	\$3,117	\$3,606	\$3,839	\$3,973	\$4,132	\$4,803
TIN-NPI	All TIN-NPIs	3,328	\$4,102	\$2,765	\$3,387	\$3,944	\$4,240	\$4,382	\$4,516	\$4,831

²¹ 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>.

²² Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

Level	Sub-group	Provider Count	Mean Score	Score Percentile						
				1st	10th	25th	50th	75th	90th	99th
TIN-NPI	Laparoscopic Repair	2,206	\$4,886	\$2,948	\$4,038	\$4,743	\$5,076	\$5,230	\$5,356	\$5,751
TIN-NPI	Open Repair	3,174	\$3,732	\$2,410	\$2,998	\$3,555	\$3,835	\$3,997	\$4,160	\$4,947

3.5.11 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.²³ 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 all risk deciles are 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, 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 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 and for each sub-group.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure scores 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 large performance difference among clinicians in the Hernia Repair measure:

- (i) the 99th percentile of the measure score is nearly twice the measure score at the 1st percentile for both the TIN level and TIN-NPI levels; and
- (ii) the measure score at the 90th percentile is approximately 30 percent greater than the score at the 10th percentile for both the TIN and TIN-NPI level.

²³ 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.

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 less than \$400 range (i.e., \$3,961-\$4,275 at the TIN level and \$3,935-\$4,297 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas (mean score of \$4,094 for urban and \$4,176 for rural clinicians at the TIN-level, and similar results 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 perform fewer femoral or inguinal hernia repairs. 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. Results indicate little variation in measure score by risk score decile, with a range in mean TIN score of \$4,056 to \$4,132 and a range in mean TIN-NPI score of \$4,031 to \$4,145, indicating that the risk adjustment model is overall functioning as intended. Full results can be seen in the NSDR.²⁴

3.6.3 Interpretation

There is clinically and practically significant variation in Hernia Repair 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. The Femoral or Inguinal Hernia Repair workgroup suggested development of sub-groups based on surgical approach, noting the differences in cost between laparoscopic repairs and open repairs. The results show a 2 percent difference in cost between rural and urban locations at a TIN and TIN-NPI testing level since those differences are already accounted for through the other components of the measure construction.

Overall, as expected, results show that clinicians are not being systematically penalized or rewarded due to risk score decile given the current Hernia Repair measure design (i.e., the differences in cost measure scores are not 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 Hernia Repair measure, Acumen expects a high degree of data completeness. To ensure that we have complete and 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 Hernia Repair 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.

²⁴ CMS, "National Summary Data Report: 11 Episode-Based Cost Measures and Two Re-evaluated Cost Measures, October 2018 Field Testing," *MACRA Feedback Page*, <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-feedback.html>.

3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the four categories of missing data that caused episodes to be excluded from the Hernia Repair 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 6: Missing Data Categories for the Femoral or Inguinal Hernia Repair Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing Beneficiary Birth Date	0	0	0
Primary Payer Other than Medicare	10,708	3,221	7,232
Beneficiary Death before Admission	13	13	16
No Continuous Enrollment in Medicare Parts A and B, and Any Enrollment in Part C	6,444	2,692	5,264

3.7.3 Interpretation

As the Hernia Repair 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 were 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 adjust accurately 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 the Merit-based Incentive Program (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 Hernia Repair 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.²⁵ 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 Femoral or Inguinal Hernia Repair: 2,348 TINs; 3,691 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.²⁶ 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

²⁵ The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

²⁶ 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>.

- 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)
- 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.²⁷

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 hours 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 hours 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

²⁷CMS, "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>.

summarizing the testing done on the measures. This presentation was followed by a Q&A session.²⁸

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 three comments for the Femoral or Inguinal Hernia Repair cost measure. After the 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 Hernia Repair cost measure.²⁹ 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.

²⁸ CMS, Webinar Recordings, Slides and Transcripts, *QPP Webinar Library*
<https://qpp.cms.gov/about/webinars>.

²⁹ Measure Applications Partnership, *National Quality Forum*.
https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.

- *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, sub-groups, episode windows, assigned services, risk adjustment variables, exclusions, and alignment of cost with quality
- Adding/removing certain trigger codes and assigned services, further sub-grouping, 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 Femoral or Inguinal Hernia Repair 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 Hernia Repair measure made after consideration of field testing analyses and stakeholder feedback are:

- **Triggers:** Edited the following triggers:
 - Removed CPT/HCPCS 49659 (hernia repair procedure using an endoscope) and DGNs as trigger codes
 - Added check for presence of DRG 350-352 (inguinal & femoral hernia procedures) for episodes occurring in inpatient settings
- **Exclusions:** Added exclusion for cases where there is another, more complex surgery on the same day.
- **Service Assignment:**
 - Added the following services:

- Bladder perforations/scrotal hemorrhage services that occur within 7 days of trigger code.
- Testicular pain services within 90 days of trigger code.
- Other urologic complications' services within 30 days of trigger code.
- **Risk Adjustment:**
 - Edited risk adjustors for:
 - Strangulated/incarcerated hernia repair
 - Recent all-cause admission and dementia as frailty indicators
 - IP episodes by checking for presence of MS - DRG 350-352 (inguinal & femoral hernia procedures)
 - Bilateral procedures
 - Removed risk adjustor for American Society of Anesthesiologists Physical Status (ASA status)

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* 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

Phone Number: (410) 786-5237

Developer Point of Contact

Organization: Acumen, LLC

Email Address: macra-cost-measures-info@acumenllc.com

Phone Number: (650) 558-8882

Other Additional Information

Femoral or Inguinal Hernia Repair Workgroup Members:

Christopher Senkowski, MD, American College of Surgeons

Donald Fry, MD, American College of Surgeons

Guy Orangio, MD, American Society of Colon and Rectal Surgeons

Jayme Lieberman, MD, MBA, FACS, American College of Surgeons

Jeffrey Cohen, MD, American Society of Colon and Rectal Surgeons

Manjil Chatterji, MD, Society of Abdominal Radiology

Mark Savarise, MD, American College of Surgeons

Mary Cathleen Shellnutt, DNP, RN, AGCNS-BC, CGRN, National Association of Clinical Nurse Specialists

Richard Dutton, MD, MBA, American Society of Anesthesiologists

The Femoral or Inguinal Hernia Repair workgroup is composed from the larger Gastrointestinal Disease Management Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the [Episode-Based Cost Measures Development Process document](#).³⁰

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>.