

Non-Emergent Coronary Artery Bypass Graft (CABG) 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	9
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	20
3.5.3 Conceptual Model of Impact of Social Risks.....	21
3.5.4 Statistical Results	21
3.5.5 Analyses and Interpretation in Selection of Social Risk Factors.....	21
3.5.6 Method for Statistical Model or Stratification Development.....	22
3.5.7 Statistical Risk Model Discrimination Statistics	23
3.5.8 Statistical Risk Model Calibration Statistics	23
3.5.9 Statistical Risk Model Calibration – Risk Decile.....	23
3.5.10 Results of Risk Stratification Analysis.....	23
3.5.11 Interpretation.....	24
3.6 Identification of Meaningful Differences in Performance.....	24

3.6.1 Method	24
3.6.2 Statistical Results	25
3.6.3 Interpretation	25
3.7 Missing Data Analysis and Minimizing Bias.....	25
3.7.1 Method	25
3.7.2 Missing Data Analysis	26
3.7.3 Interpretation	26
4.0 Feasibility.....	27
4.1 Data Elements Generated as Byproduct of Care Processes	27
4.2 Electronic Sources	27
4.3 Data Collection Strategy	27
4.3.1 Data Collection Strategy Difficulties.....	27
5.0 Usability and Use	28
5.1 Use	28
5.1.1 Current and Planned Use	28
5.1.2 Feedback on the Measure and Development Process	28
5.2 Usability	32
5.2.1 Improvement	32
5.2.2 Unexpected Findings.....	32
5.2.3 Unexpected Benefits	32
6.0 Related and Competing Measures	33
6.1 Relation to Other Cost Measures	33
6.2 Harmonization.....	33
6.3 Competing Measures	33
Contact Information	34

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Non-Emergent Coronary Artery Bypass Graft (CABG) 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 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

Non-Emergent Coronary Artery Bypass Graft (CABG) Episode-Based Cost Measure

1.3 Type of Measure

Cost/Resource Use

¹ CMS, “Non-Emergent Coronary Artery Bypass Graft (CABG) 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, “Non-Emergent Coronary Artery Bypass Graft (CABG) 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 Non-Emergent Coronary Artery Bypass Graft (CABG) cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who undergo a CABG procedure. 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 Non-Emergent CABG 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. Opportunities for improvement for non-emergent CABG exist within two primary performance gaps: reducing readmissions and encouraging cost-effective post-acute care.

While the cost of CABG itself is significant, readmissions related to complications of the procedure and perioperative care reduce quality of life and increase costs. A study of CABG surgeries performed in California in 2009 found that more than 13 percent of patients were readmitted within 30 days, with the most frequent reasons for readmission being heart failure and post-operative infections, accounting for 15.3 and 12.9 percent of readmissions, respectively.³ Studies in other US states have found similar readmission rates.⁴ A 2015 study investigating causes for readmissions in an 11-hospital network in the US found that 68.7 percent of readmissions were classified as either clearly related or possibly related to the surgical procedure.⁵ In this analysis, the mean hospital stay for clearly related readmissions was 15.5 days. The study also found readmission rates varied across hospitals from 6.1 percent to 18.0 percent; there was also variation in the likelihood of the readmission being related to the CABG procedure.

Hospital readmission rates can be used as an indicator of the quality of care during index hospitalizations. CABG accounted for 87 percent of the 30-day readmission rate (15.6 percent)

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 Li, Zhongmin, Ehrin J. Armstrong, et al. "Hospital Variation in Readmission after Coronary Artery Bypass Surgery in California." *Circulation: Cardiovascular Quality and Outcomes* (2012).

⁴ Ibid.

⁵ Lancey, Robert, Paul Kurlansky, et al. "Uniform Standards Do Not Apply to Readmission Following Coronary Artery Bypass Surgery: A Multi-Institutional Study." *The Journal Of Thoracic And Cardiovascular Surgery* 149, no. 3 (2015): 850-7.e.

for all cardiac procedures recorded in New York from 2005 to 2007.⁶ According to a 2010 study, each 30-day readmission for CABG resulted in additional costs up to \$13,256. Although mortality rates have steadily declined in recent years, the readmission rates have remained steady or increased in some cases, possibly indicating insufficient coordination efforts between inpatient and outpatient care providers.⁷

Post-acute care (PAC) also contributes to overall costs and cost variation with CABG surgery. Following a CABG procedure, patients may: (i) go without PAC services, (ii) receive only home health care services, (iii) receive home health and outpatient physical therapy, or (iv) be transferred to a skilled nursing facility (SNF). The type of PAC received will depend on the availability of PAC services (e.g., not all communities have SNFs), patients' clinical and functional profiles (e.g., comorbidities and post-surgical rehabilitation needs), and patients' residential environments (e.g., presence of informal care givers, presence of stairs in the home). Different PAC settings provide care at different levels of intensity, and they result in different costs to the Medicare program.⁸ A 2017 study that examined PAC spending for FFS Medicare beneficiaries following CABG, colectomy, and total hip replacement from 2009 to 2012 found that at least half of FFS Medicare beneficiaries who underwent CABG went on to receive PAC services.⁹

A 2017 study using 2009 to 2012 data on Medicare beneficiaries found that the average spending on PAC services was under \$4,000 for hospitals in the bottom quintile of 90-day PAC episode spending and over \$10,000 for the upper quintile. After patient-level risk adjustment and price-standardization, the majority of cost variation remained. This variation was more dependent on the type of PAC selected (i.e., inpatient rehabilitation and skilled nursing) rather than the intensity of care at a given PAC setting, such as the length of stay.¹⁰ These data point to the potential for improvement in care quality, patients' quality of life, as well as Medicare cost savings.

2.2 Performance Gap

2.2.1 Rationale

An average of approximately 100,000 Medicare beneficiaries underwent CABG surgery annually between 2000 and 2012.¹¹ More than 13 percent of CABG patients are readmitted within 30 days, and each 30-day readmission for CABG resulted in additional costs up to \$13,256.^{12,13} The Non-Emergent CABG episode-based cost measure was recommended for development by an expert clinician committee—the Cardiovascular Disease Management

⁶ Price, Jonathan D., Jamie L. Romeiser, et al. "Risk Analysis for Readmission after Coronary Artery Bypass Surgery: Developing a Strategy to Reduce Readmissions." *Journal of the American College of Surgeons* 216, no. 3 (2013): 412-19.

⁷ Birkmeyer, J.D., C. Gust., et al. "Medicare Payments for Common Inpatient Procedures: Implications for Episode-Based Payment Bundling." *Health Serv Res* 45, no. 6 Pt 1 (2010): 1783-95.

⁸ Buntin, M. B., A. D. Garten, et al. "How Much Is Postacute Care Use Affected by Its Availability?" [In eng]. *Health Serv Res* 40, no. 2 (Apr 2005): 413-34.

⁹ Chen, L. M., E.C. Norton, et al. "Spending on Care after Surgery Driven by Choice of Care Settings Instead of Intensity of Services." [In eng]. *Health Aff (Millwood)* 36, no. 1 (Jan 01 2017): 83-90.

¹⁰ Ibid.

¹¹ McNeely, Christian, Stephen Markwell, et al. "Trends in Patient Characteristics and Outcomes of Coronary Artery Bypass Grafting in the 2000 to 2012 Medicare Population." *The Annals Of Thoracic Surgery* 102, no. 1 (2016): 132-38.

¹² Li, Zhongmin, Ehrin J. Armstrong, et al. "Hospital Variation in Readmission after Coronary Artery Bypass Surgery in California." *Circulation: Cardiovascular Quality and Outcomes* (2012).

¹³ Birkmeyer, J.D., C. Gust., et al. "Medicare Payments for Common Inpatient Procedures: Implications for Episode-Based Payment Bundling." *Health Serv Res* 45, no. 6 Pt 1 (2010): 1783-95.

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.

2.2.2 Performance Scores

Performance scores are provided for 863 clinician group practices (identified by Tax Identification Number [TIN]) and 2,615 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 Payment System (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 10 episodes.

Table 1: Distribution of Performance Scores

Metric	TIN	TIN-NPI
Mean score	\$42,903	\$42,515
Standard deviation	\$3,670	\$3,771
Score IQR	\$4,157	\$4,124
Score percentile		
10 th	\$39,181	\$38,721
20 th	\$40,107	\$39,640
30 th	\$40,806	\$40,405
40 th	\$41,408	\$41,148
50 th	\$42,175	\$41,861
60 th	\$43,057	\$42,650
70 th	\$43,764	\$43,500
80 th	\$45,489	\$44,863
90 th	\$47,442	\$47,051

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 Non-Emergent CABG 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 Non-Emergent CABG 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

863 clinician group practices and 2,615 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 10 or more Non-Emergent CABG episodes during the measurement period. Episodes from all 50 States and D.C. in the following setting were included: acute inpatient (IP) hospitals.

3.1.6 Patient Cohort Included in the Testing and Analysis

42,600 Medicare beneficiaries (from 42,600 episodes) were included in TIN level testing and analysis, and 39,898 beneficiaries (from 39,898 episodes) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the Non-Emergent CABG measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who underwent a non-emergent CABG procedure during the measurement period as identified by the episode trigger Current Procedural Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) codes on Part B Physician/Supplier 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 undergoing non-emergent CABG 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 IP setting.
- The IP facility is a short-term stay acute hospital as defined by subsection (d).¹⁴
- The beneficiary did not receive a concurrent Cox Maze procedure.
- The CABG procedure was not emergent.
- The beneficiary is not receiving dialysis treatment for end-stage renal disease.
- The episode is not for a reoperation of heart artery bypass or valve procedure more than 1 month after original operation.
- The beneficiary has not experienced cardiogenic shock or other shock on the trigger date, the day before the trigger date, or during the IP stay prior to the procedure.
- The IP stay for the IP procedure contains a relevant Medicare Severity Diagnosis-Related Group (MS-DRGs) code.
- The episode is not an outlier case.

To determine whether the Non-Emergent CABG 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 Non-Emergent CABG 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 6.5 percentage points across each of the characteristics in the analysis at TIN and TIN-NPI level testing. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 27.3 percent, the same at TIN level testing and 27.2 percent at TIN-NPI level testing with the inclusion criteria applied. The difference in the percentage of beneficiaries for race with and without the inclusion criteria is less than 3.0 percentage points for all categories for TIN and TIN-NPI level testing. The breakdown of male and female beneficiaries remains similar when comparing the use of inclusion criteria at both at TIN and TIN-NPI level testing, with a 0.9 percentage difference for females and males when applying the inclusion criteria. 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.

3.1.7 Sample Differences

n/a

¹⁴ 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](#).

3.1.8 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 Non-Emergent CABG 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.¹⁵ The FY 2018 Medicare FFS program proper payment rate was 91.9 percent.¹⁶ 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

As displayed in the table below, 100 percent of TINs and TIN-NPIs at 10, 20, and 30-episode volume thresholds 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 0.82 and for TIN-NPIs is 0.74. 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	0.82	100.0%	0.74	100.0%
20	0.87	100.0%	0.82	100.0%
30	0.89	100.0%	0.86	100.0%

¹⁵ 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>.

¹⁶ Ibid.

3.2.4 Interpretation

Measure Reliability

Overall reliability of the Non-Emergent CABG 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.¹⁷

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 Non-Emergent CABG 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 Cardiovascular Disease Management Clinical Subcommittee, (ii) the Non-Emergent CABG 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 46 members with clinical experience in cardiovascular disease management, affiliated with 31 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 Non-Emergent CABG workgroup was composed of 14 members, affiliated with 14 specialty societies, including the Society of Thoracic Surgeons, the American Medical Association, and the American College of Cardiology. 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 topics to inform the workgroup such as pre- and post-trigger windows for selected episodes, and inclusion of services and costs for attributed

¹⁷ 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.

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 a non-emergent CABG procedure, thus preventing inclusion of unrelated cost variation in this measure. Assigned services occurring in the emergency department, outpatient facility and clinician services, IP – medical, IP – surgical, inpatient rehabilitation facility – medical, durable medical equipment, prosthetics, orthotics, and supplies, and home health were defined separately for the pre- and post-trigger windows, and include non-emergent CABG procedures, evaluation, testing, treatment, complications, 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 Non-Emergent CABG 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 non-emergent CABG. For this analysis, we compared the ratio of observed to expected (henceforth called the "O/E cost ratio") spending for Non-Emergent CABG episodes with and without complications related to non-emergent CABG occurring in the post-trigger period. This analysis sought to confirm the expectation that the Non-Emergent CABG 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 Non-Emergent CABG measure were classified into clinically coherent groups of services, called "clinical themes." The Non-Emergent CABG measure clinical themes are:

- **Renal Failure:** Inpatient and outpatient hospital care including emergency department visits and critical care for acute kidney failure.
- **Preoperative Work-Up:** Includes routine chest x-rays; electrocardiograms; laboratory testing, such as blood tests, and coagulation assessment blood tests; other diagnostic techniques, such as x-rays; or diagnostic procedures, such as office or outpatient evaluations.
- **Postoperative Labs:** Include therapeutic procedures and lab testing following the procedure.
- **Imaging after CABG:** Includes imaging (CT scans, ultrasounds, x-rays, MRIs) and related supplies for aortic aneurysm, pneumonia, pulmonary embolism, sepsis, open wounds, or other complications.
- **Rehabilitation:** Includes services for physical therapy.
- **Wound Care and Surgical Site Infection (SSI):** Includes inpatient and outpatient care and wound care supplies.
- **Coronary Disease [Myocardial Infarction (MI) and Coronary Artery Disease (CAD)]:** Inpatient and outpatient hospital care including emergency department visits and critical care for ischemic or hypertensive heart disease and related procedures including angioplasty, cardiac catheterization, and procedures to treat stenosis, and other therapeutic procedures and related supplies.

- **Cardiovascular Care, Other:** Inpatient and outpatient hospital care including emergency department visits and critical care for heart failure, arrhythmias or other related conditions, including catheterization, and other diagnostic or therapeutic procedures and supplies.
- **Readmissions / Emergency Department (ED) Visits, Other:** Inpatient and outpatient hospital care including emergency department ED visits and critical care for volume depletion, pneumonia, bronchitis, fever, sepsis, etc.
- **Pleural Effusions and Pericardial Effusions:** Inpatient and outpatient hospital care including emergency department visits and critical care for disease of the pericardium or pleural effusion, and related diagnostic, therapeutic, or cardiac procedures and supplies.

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 Coronary Disease (MI and CAD) and Wound Care and SSI themes would have the highest correlation with risk-adjusted episode cost, as they are likely associated with high cost even after accounting for beneficiary characteristics. By contrast, we expected that the Postoperative Labs and Preoperative Work-Up themes after a CABG procedure would have more nuanced, offsetting effects.

3.3.3 Statistical Results from Validity Testing

Table 3 presents an analysis of validity, showing the O/E cost ratio of episode with or without downstream acute (re)admissions and post-acute care. 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 downstream acute readmission during the post-trigger period is 1.20, compared with 0.99 for episodes without downstream acute readmission during the post-trigger period. The mean O/E cost ratio for episodes with post-acute care [Inpatient Rehabilitation Facilities, Long-Term Care Hospitals (LTCH), Home Health (HH), or Skilled Nursing Facilities (SNF)] during the post-trigger period is 1.04, compared with 0.93 for episodes without post-acute care during the post-trigger period.

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	0.22	0.67	0.76	0.80	0.87	0.94	1.07	1.30	1.46	1.78
Episodes with Downstream Acute (Re)admission	1.20	0.26	0.77	0.89	0.93	1.02	1.13	1.31	1.57	1.74	2.05
Episodes without Downstream Acute (Re)admission	0.99	0.21	0.66	0.75	0.80	0.87	0.94	1.05	1.27	1.42	1.74
Episodes with Post-Acute Care	1.04	0.24	0.66	0.75	0.81	0.89	0.99	1.15	1.37	1.52	1.83
Episodes without Post-Acute Care	0.93	0.15	0.68	0.76	0.80	0.85	0.90	0.96	1.05	1.20	1.61

The clinical themes analysis demonstrates that there is a moderate correlation between the Coronary Disease (correlation: 0.54), Rehabilitation (correlation: 0.53), and Wound Care and

Surgical Site Infection (correlation: 0.53) themes and risk-adjusted cost. By contrast, the Postoperative Labs (correlation: 0.02) and Preoperative Work-Up (correlation: 0.03) themes were found to have a weak 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 Non-Emergent CABG measure is able to accurately capture higher resource use.

The clinical themes analysis demonstrates that high risk-adjusted cost is strongly associated with themes related to Coronary Disease and Wound Care and SSI, and also linked – though more weakly, to Renal Failure and Pleural Effusions and Pericardial Effusions themes. This indicates that the measure may penalize clinicians who have higher rates of complications, while not disincentivizing the provision of appropriate pre- and post-operative care, such as postoperative labs and preoperative work-ups. Importantly, we see that correlation with risk-adjusted cost is strong not only for high-cost themes such as Coronary Disease (average cost: \$6,929), but also for lower cost themes such as Rehabilitation (average cost: \$4,291). 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 Non-Emergent CABG measure to ensure a homogenous patient population within the scope of the measure focus on non-emergent CABG procedures 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 that are needed to ensure a homogenous patient population, and which cannot be adequately accounted for through risk adjustment. These exclusions, along with their rationales, are listed below:

- *Episodes where beneficiary death date occurred before the episode end.*
 - 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 receives an emergent CABG procedure.*
 - These patients are more ill with higher costs and rates of complications, including mortality. The variance in costs for this high-risk patient cohort is also higher and would likely not be adequately accounted for by risk adjustment.
- *Episodes where the beneficiary undergoes a concurrent Cox Maze procedure.*
 - This procedure occurs in a minority of patients who undergo CABG surgery. This occurs in a higher risk cohort and increases the complexity of the procedure and associated costs.
- *Episodes where the beneficiary is receiving dialysis for end-stage renal disease.*
 - Patients with severe kidney disease have higher costs and risk of complications as a high-risk patient population.
- *Episodes where the beneficiary has a reoperation sternotomy.*
 - These patients have had prior thoracic surgery and have higher costs and risk of complications.

- *Episodes where the beneficiary receives shock prior to a CABG procedure.*
 - These patients are more ill with higher costs and rates of complications, including mortality.
- *Episodes where the beneficiary receives IP procedures without relevant MS-DRGs.*
 - A minority of CABG surgeries are represented with alternate DRGs. These cases typically represent a second, high-cost procedure also performed during the inpatient stay. Frequently, they represent very high risk patients requiring additional invasive support.
- *Episodes classified as outlier cases.*
 - To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1st percentile and above the 99th percentile are considered outliers and removed from measure calculation.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different risk profile than the included episodes, such as a higher mean cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For the exclusions, we examined the number of episodes and 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 Non-Emergent CABG measure is provided in the Measure Codes List.¹⁸

3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics and O/E cost ratios for the Non-Emergent CABG measure exclusions. Cost statistics are also provided for the set of final episodes included in the Non-Emergent CABG measure for comparison, with a testing volume threshold of 10 episodes at the TIN and TIN-NPI levels.

¹⁸ CMS, “Non-Emergent Coronary Artery Bypass Graft (CABG) 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 th	90 th		10 th	90 th
All Episodes Meeting Triggering Logic	65,520	100.00%	\$44,410	\$26,129	\$70,530	1.01	0.79	1.31
Episodes where the beneficiary dies before the end of the episode	3,960	6.04%	\$54,518	\$4,236	\$105,637	1.11	0.79	1.61
Episodes where the beneficiary receives an emergent CABG procedure	4,191	6.40%	\$49,759	\$25,652	\$78,603	1.01	0.78	1.31
Episodes where the beneficiary undergoes a concurrent Cox Maze procedure	3,451	5.27%	\$38,897	\$4,234	\$78,715	1.05	0.85	1.32
Episodes where the beneficiary is receiving dialysis for end-stage renal disease	3,106	4.74%	\$56,533	\$8,581	\$93,156	0.94	0.61	1.31
Episodes where the beneficiary is having a redo sternotomy	2,139	3.26%	\$50,064	\$8,131	\$84,195	1.07	0.81	1.45
Episodes where the beneficiary receives shock prior to a CABG procedure	1,115	1.70%	\$72,733	\$42,645	\$110,157	1.17	0.80	1.73
Episodes where the beneficiary receives IP procedures without relevant MS-DRGs	5,361	8.18%	\$13,861	\$3,533	\$26,442	1.00	1.00	1.00
Episodes classified as outlier cases	880	1.34%	\$84,591	\$29,889	\$151,176	1.57	0.54	2.85
<i>Final Episodes (TIN)</i>	42,600	65.02%	\$42,192	\$27,539	\$62,047	0.99	0.79	1.28
<i>Final Episodes (TIN-NPI)</i>	39,898	60.89%	\$42,005	\$27,497	\$61,675	0.99	0.79	1.27

3.4.3 Interpretation

Statistical results indicate that the excluded episodes, aside from outliers, have similar mean observed costs to the final set of episodes, but that there is more variation for excluded 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: The difference between mean observed episode cost for death and non-death episodes is large: \$54,518 compared to \$42,192 for final episodes at the TIN-level, and the difference becomes much more marked at the 90th percentile where episodes ending in death are \$105,637 compared to \$62,047 for the TIN-level final episodes. This suggests that episodes ending in death are likely for sicker, more complex patients who can incur many costs in perimortem care. Because of this, including episodes ending in death in measure calculation may create incentives to avoid complex, high-risk patients.

Episodes where the beneficiary has a concurrent Cox Maze procedure. While the mean observed cost of these episodes is approximately \$3,300 less than for the final set of episodes, they are much more expensive in the right tail. At the 90th percentile, episodes with this concurrent procedure are \$78,715 compared to \$62,027 for final episodes at the TIN-level. As such, these cases are not included in the measure to ensure a clinically comparable patient cohort, due to the difficulty of adequately risk adjusting for these differences, particularly for the riskiest patients in this cohort.

Episodes where the beneficiary receives an emergent CABG. The mean observed cost of these episodes is approximately \$7,500 more than for the final set of episodes, with the difference becoming larger at the 90th percentile (\$78,603 compared to \$62,047 for the final episodes at the TIN-level). This is in line with expectations and expert clinical input about the different patient cohort that receives an emergent CABG procedure.

Episodes where the beneficiary is diagnosed with end-stage renal disease and is on dialysis. The mean observed cost of these episodes approximately \$14,300 more than for the final set of episodes, in line with expectations about the higher risk patient cohort that is on dialysis for the treatment of end-stage renal disease.

Episodes where the beneficiary is undergoing a reoperation of heart artery bypass or valve procedure more than one month after original operation. The mean observed cost of these episodes is approximately \$8,000 more than for the final set of episodes. The difference in patient cohort becomes more pronounced at the 90th percentile, where episodes for redo procedures are \$84,195 compared to \$62,057.

Episodes where the beneficiary experiences shock prior to CABG. The mean observed cost of these episodes is approximately \$30,000 more than for the final set of episodes. This reflects the fact that patients in this cohort are highly complex and require different care from the overall patient cohort for non-emergent CABG. The variation in cost is also higher for these patients than for the final set of episodes, with the cost at the 90th percentile almost at \$110,157 compared to the final episodes cost of \$62,047 at the TIN-level.

Episodes where the beneficiary receives IP procedures without relevant MS-DRGs. The observed cost profile of these episodes is very different than for the final set of episodes: the mean observed cost is \$14,861, or less than half of the mean observed cost for the final set of episodes at the TIN-level (\$42,192). At the right end of the tail, this trend continues with the observed cost at the 90th percentile for these episodes at \$26,442 compared to \$61,675 for the final set of episodes at the TIN-level. As such, these episodes are excluded to ensure clinical comparability.

Outlier cases: The O/E cost ratio for outlier cases ranges from 0.54 at the 10th percentile to 2.85 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 110 risk factors and stratification by two risk categories.

The risk adjustment model for the Non-Emergent CABG 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 has long-term (current) anticoagulant use in the 120-day lookback period as the use of anticoagulant indicates higher procedural risk of bleeding and the presence of additional comorbidities;
- whether the beneficiary has long-term (current) antiplatelet use in the 120-day lookback period as the prior use of antiplatelet indicates higher procedural risk of bleeding and often is a surrogate for prior coronary intervention;
- whether the beneficiary has a history of myocardial infarction (MI) or percutaneous coronary intervention (PCI) in the 30 days prior to the trigger date, including the trigger date as patients undergoing CABG surgery after recent MI or PCI are distinct from patients with CABG without recent revascularization or MI;
- whether the beneficiary has dementia in the 120-day lookback period as these patients are more likely to have increased costs following the procedure due to complications and additional required services including post-acute care rehabilitation;
- whether the beneficiary smokes or is dependent on nicotine in the 120-day lookback period as these patients are more likely to have severe coronary artery disease with higher procedural risks and higher post-operative respiratory complications and recurrent coronary events;
- whether the beneficiary uses home oxygen in the 120-day lookback period as this indicator of frailty is associated with higher resource utilization after CABG surgery;
- whether the beneficiary has nursing facility physician visits in the 120-day lookback period as this indicator of frailty is associated with higher resource utilization after CABG surgery;
- whether the beneficiary has been discharged from an IP stay in the 30 days prior to the trigger date, including the trigger date as this indicator of frailty is associated with higher resource utilization after CABG surgery;
- whether the beneficiary has a history of wheelchair use in the 120-day lookback period as this indicator of frailty is associated with higher resource utilization after CABG surgery;
- whether the beneficiary home health services during the 120-day lookback period as this indicator of frailty is associated with higher resource utilization after CABG surgery;

- whether the beneficiary has received outpatient therapy during the 120-day lookback period as this indicator of frailty is associated with higher resource utilization after CABG surgery, and;
- whether the beneficiary has a history of walking aid use in the 120-day lookback period as this indicator of frailty is associated with higher resource utilization after CABG surgery.

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 Non-Emergent CABG measure sub-groups, which represent more granular, mutually exclusive patient populations defined by clinical criteria:

- CABG with Concurrent Aortic Valve Replacement
- Isolated CABG

Full details of the risk adjustment model are in the Measure Codes List File.¹⁹ The National Summary Data Report (NSDR) Addendum includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.²⁰

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 Non-Emergent CABG 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 the influence of the attributed clinician, or any other factors that would help prevent unintended consequences. These additional risk adjusters are listed in the section above.

¹⁹ CMS, "Non-Emergent Coronary Artery Bypass Graft (CABG) 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>.

²⁰ 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>.

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, which are based on the presence or lack of a concurrent aortic valve replacement, are listed in the above section. The stratification for isolated and concurrent aortic valve replacement accounts for the higher risk for complications and costs due to both a more technically difficult surgery along with differences in patient characteristics when a patient requires concurrent aortic valve replacement.

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.²¹ 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 range from 25.0 percent to 27.5 percent across the two sub-groups in this measure. The majority of the beneficiaries (87.4% - 92.0%) 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 1.9 to 2.5 percent of beneficiaries are classified below a high school education level, more than 80 percent of all episodes (82.6% - 84.9%) are classified at a high school level or greater. Finally, 21.5 to 22.7 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 adjusters. Acumen ran a step-wise regression to include gender, dual status, gender + dual status, and

²¹ 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.

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 are likely predictive factors for determining resource use among beneficiaries for the relevant characteristic. However, the analysis also shows that the directions of the effects of social risk factors are not consistent. For example, high income beneficiary episodes may display lower expected spending for the CABG with Concurrent Aortic Valve Replacement sub-group but higher expected spending for the Isolated CABG sub-group. There are also differences in significance between the sub-groups; for instance, the high income coefficient is statistically significant for the CABG with Concurrent Aortic Valve Replacement sub-group, but not for the Isolated CABG sub-group.

Secondly, we analyzed the impact of adding social risk variables on overall model performance by looking at the differences in the ratio of 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 ratios for 99.0 percent of TINs and 99.4 percent of TIN-NPIs changed by ± 0.03 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 levels, with a Spearman correlation coefficient of 0.995 for both levels. 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 direction and limited impact of social risk factor effects under the current risk adjustment model, we believe the Non-Emergent CABG 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 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.²²

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Non-Emergent CABG cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.42. The adjusted R-squared is 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. 2011.²³

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 the average of all deciles having a predictive ratio of 1.00, ranging between 0.97 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, CABG procedures with concurrent aortic valve replacement are more expensive than isolated CABG cases. At the TIN level, the mean score for CABG with concurrent aortic valve replacement episodes is \$53,506 compared to isolated CABG episodes at \$41,294. Results are similar at the TIN-NPI level, where the mean score for CABG with concurrent aortic valve replacement episodes is \$53,219 compared to isolated CABG episodes

²² 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>.

²³ Pope, Gregory C., John Kautter, et al. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

at \$40,870. 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	863	\$42,903	\$37,341	\$39,181	\$40,478	\$42,175	\$44,635	\$47,442	\$55,062
TIN	CABG w/ concurrent aortic valve replacement	806	\$53,506	\$42,098	\$46,596	\$48,852	\$52,031	\$56,039	\$62,862	\$79,247
TIN	Isolated CABG	863	\$41,294	\$35,683	\$37,596	\$38,975	\$40,558	\$43,016	\$45,856	\$54,085
TIN-NPI	All TIN-NPIs	2,615	\$42,515	\$36,330	\$38,721	\$40,012	\$41,861	\$44,136	\$47,051	\$55,348
TIN-NPI	CABG w/ concurrent aortic valve replacement	2,272	\$53,219	\$41,110	\$45,627	\$47,963	\$51,307	\$56,413	\$63,299	\$83,351
TIN-NPI	Isolated CABG	2,615	\$40,870	\$34,662	\$36,993	\$38,402	\$40,280	\$42,518	\$45,451	\$53,208

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.²⁴ 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 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.

²⁴ 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.

3.6.2 Statistical Results

Key findings show that, generally, there is a large performance difference among clinicians in the Non-Emergent CABG measure:

- (i) the 99th percentile of the measure score is approximately 1.5 times the 1st percentile at both the TIN level and TIN-NPI levels;
- (ii) the Non-Emergent CABG measure score at the 90th percentile is nearly 20 percent greater than the score at the 10th percentile at both the TIN and TIN-NPI level; and

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 \$2,500 range (i.e., \$41,493 - \$43,994 at the TIN level and \$41,249 - \$43,391 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas, with mean scores in rural areas being \$1,348 greater at the TIN level and \$527 greater 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 CABG procedures. We also analyzed clinicians by risk score decile, as variation by risk score decile could indicate that the risk adjustment model is over- or under-correcting for clinicians with systematically riskier patients. Measure scores also show little variation by risk score decile, with a range in mean TIN score of \$42,405 to \$43,393 and a range in mean TIN-NPI score of \$41,863 to \$42,879, indicating that the risk adjustment model is overall functioning as intended. Full results can be seen in the NSDR.²⁵

3.6.3 Interpretation

There is clinically and practically significant variation in Non-Emergent CABG 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 Non-Emergent CABG measure-specific workgroup suggested the development of sub-groups based on the presence of concurrent aortic valve replacement, noting the higher risk for complications and cost due a more technically difficult surgery and differences in patient characteristics. The results show a large difference in mean cost between isolated CABG procedures and CABG procedures with a concurrent aortic valve replacement. Risk adjustment variables likely have different impacts between the two sub-groups as well. Overall, as expected, results show that clinicians are not being systematically penalized or rewarded due to risk score decile given the current Non-Emergent CABG measure design (i.e., the differences in cost measure scores are not due to 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 Non-Emergent CABG measure, Acumen expects a high degree of data completeness. To further 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

²⁵ 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>.

the EDB, the beneficiary does not appear in the EDB, or the beneficiary death date occurs before the episode trigger date.

The Non-Emergent CABG 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 Non-Emergent CABG 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 inpatient admission 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 Non-Emergent CABG Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	0	0	0
Death before admission	118	102	171
Other primary payer	7,306	994	3,687
Not continuously enrolled	4,234	930	2,980

3.7.3 Interpretation

As the Non-Emergent CABG 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 Non-Emergent CABG 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.²⁶ 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 11 episode-based cost measures: 14,237 TINs; 63,984 TIN-NPIs
- Testing sample for Non-Emergent CABG: 859 TINs; 2,612 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.²⁷ 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 all the 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

²⁶ The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

²⁷ 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.²⁸

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 summarizing the testing done on the measures. This presentation was followed by a Q&A session.²⁹

²⁸ 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>.

²⁹ 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 two comments for the Non-Emergent CABG 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 Non-Emergent CABG cost measure.³⁰ 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, highlighting that a transparent approach to member selection would ensure an appropriate mix of specialties and clinician types.

³⁰ 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 Non-Emergent CABG 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 Non-Emergent CABG measure made after consideration of field testing analyses and stakeholder feedback are:

- **Service Assignment:** Edited the following assigned services:
 - Assign the cost for infective endocarditis within 30 days of the trigger in the post-trigger period
 - Assign home health services if the principal diagnosis of the claim is strongly associated with CABG surgery (e.g., “Encounter for Surgical aftercare Following Surgery on the Circulatory System” and “Chronic Ischemic Heart Disease”)
 - Assign home health skilled nursing care if the principal diagnosis is strongly associated with CABG surgery whether it occurs in a skilled nursing facility or home health setting

- Assign inpatient rehabilitation facility admissions that have principal diagnoses that are strongly associated with CABG surgery or are potential complications of CABG surgery and are new diagnoses (specifically stroke)
- **Risk Adjustment:**
 - Edited the following measure-specific risk adjustors:
 - Smoking/Nicotine Dependence
 - Recent MI or PCI use
 - Anticoagulant Use
 - Added risk adjustors for the following:
 - Home Oxygen
 - Nursing Facility Physician Visits
 - Recent All-Cause 30-Day Admissions
 - Wheelchair Use
 - Walker Use
 - Recent Home Health Services
 - Outpatient Therapy

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

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

Non-Emergent CABG Workgroup Members:

Adam Kingeter, American Society of Anesthesiologists

Alice Bell, American Physical Therapy Association

Elizabeth Schulwolf, Society of Hospital Medicine

Ewen Nicol, AMDA - The Society for Post-Acute and Long-Term Care Medicine

Faisal Bakaeen, Society of Thoracic Surgeons

James Blankenship, Society for Cardiovascular Angiography and Interventions

James Scharff, Society of Thoracic Surgeons

Jeff Rich, Society of Thoracic Surgeons

Keith Horvath, Society of Thoracic Surgeons

Martha Radford, American Medical Association

Matthew Elrod, American Physical Therapy Association

Peggy Kalowes, American Heart Association, Cardiovascular Nursing Council

Sanjay Samy, Society of Thoracic Surgeons

Suma Thomas, American College of Cardiology

William Van Decker, American Society of Nuclear Cardiology

The Non-Emergent CABG workgroup is composed from the larger Cardiovascular Disease Management Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the [Episode-Based Cost Measures Development Process document](#).³¹

³¹ 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>.