

Elective Primary Hip Arthroplasty Measure

Measure Justification Form

June 2019



Contents

1.0 Introduction	4
1.1 Project Title and Overview	4
1.2 Measure Name	4
1.3 Type of Measure	4
2.0 Importance.....	5
2.1 Evidence to Support the Measure Focus	5
2.1.1 Measure Description	5
2.1.2 Evidence for Measure Focus	5
2.2 Performance Gap	6
2.2.1 Rationale	6
2.2.2 Performance Scores.....	7
3.0 Scientific Acceptability	8
3.1 Data Sample Description	8
3.1.1 Type of Data Used for Testing.....	8
3.1.2 Specific Dataset Used for Testing	8
3.1.3 Dates of the Data Used in Testing.....	8
3.1.4 Levels of Analysis Tested.....	8
3.1.5 Entities Included in the Testing and Analysis.....	8
3.1.6 Patient Cohort Included in the Testing and Analysis.....	8
3.1.7 Sample Differences	10
3.1.8 Social Risk Factors Included in Analysis	10
3.2 Reliability Testing.....	10
3.2.1 Level of Reliability Testing	10
3.2.2 Method of Reliability Testing	10
3.2.3 Statistical Results from Reliability Testing	11
3.2.4 Interpretation	12
3.3 Validity Testing	12
3.3.1 Level of Validity Testing.....	12
3.3.2 Method of Validity Testing	12
3.3.3 Statistical Results from Validity Testing	14
3.3.4 Interpretation	15
3.4 Exclusions Analysis	15
3.4.1 Method of Testing Exclusions.....	15
3.4.2 Statistical Results from Testing Exclusions.....	16
3.4.3 Interpretation	17
3.5 Risk Adjustment or Stratification	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.....	20
3.5.4 Statistical Results	20
3.5.5 Analyses and Interpretation in Selection of Social Risk Factors.....	20
3.5.6 Method for Statistical Model or Stratification Development.....	21
3.5.7 Statistical Risk Model Discrimination Statistics	22
3.5.8 Statistical Risk Model Calibration Statistics	22
3.5.9 Statistical Risk Model Calibration – Risk Decile.....	23
3.5.10 Interpretation.....	23
3.6 Identification of Meaningful Differences in Performance.....	23
3.6.1 Method	23

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

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Elective Primary Hip Arthroplasty 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

Elective Primary Hip Arthroplasty Episode-Based Cost Measure

1.3 Type of Measure

Cost/Resource Use

¹ CMS, “Elective Primary Hip Arthroplasty 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, “Elective Primary Hip Arthroplasty 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 Elective Primary Hip Arthroplasty cost measure (also referred to as “the Total Hip Arthroplasty (THA) measure”) evaluates clinicians’ risk-adjusted cost to Medicare for beneficiaries who receive this 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 Elective Primary Hip Arthroplasty 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, the Elective Primary Hip Arthroplasty measure represents an area where there are significant opportunities for improvement, including mitigating the use and variation of institutional post-acute care (PAC); reducing overutilization; and increasing the use of less invasive surgical techniques.

The use of PAC accounts for a significant portion of the costs and cost variability for a hip arthroplasty episode,³ and institutional PAC is particularly costly. Sabeh et al. found the cost of a hip arthroplasty episode was higher for patients receiving rehabilitation at inpatient rehabilitation facilities or skilled nursing facilities compared to a home health agency,⁴ evidence in favor of non-institutional PAC.

Developing appropriate criteria for hip arthroplasty could help identify patients who truly need the procedure and mitigate overutilization.⁵ Findings show significant variation in joint replacement usage by geography, patient preference, and clinical criteria followed by

² Fred, Herbert L. (2016). “Cutting the Cost of Health Care: The Physician’s Role.” Texas Heart Institute Journal, vol. 43, no. 1, pp. 4-6.

³ Shubeck, S. P., et al. (2018). “Hot Spotting as a Strategy to Identify High-Cost Surgical Populations.” Annals of Surgery.

⁴ Sabeh, K. G., et al. (2017). “The Impact of Discharge Disposition on Episode-of-Care Reimbursement after Primary Total Hip Arthroplasty.” The Journal of Arthroplasty 32(10): 2969-2973.

⁵ Ghomrawi, Hassan M. K., et al. (2012). “Appropriateness Criteria and Elective Procedures – Total Joint Arthroplasty.” The New England Journal of Medicine 367: 2467-2469.

surgeons.^{6,7} One study found surgeons followed different criteria when recommending surgery to patients with different severity levels and that 25 percent of hip replacements performed could be considered inappropriate,⁸ pointing to the possibility of significant cost savings if hip replacements are performed more selectively.

The relative invasiveness of different surgical approaches offers additional opportunity for performance improvement. Surgeons have developed less invasive methods of inserting the hip arthroplasty prosthesis, including one that effectively spares the muscles around the hip. One study found the use of said method resulted in a shorter length of acute hospital stay, increased discharges to home, and improved pain and Harris Hip Scores at three and six months post-surgery, with no differences in complication rates.⁹ A second study analyzing non-trauma-related hip replacements found cost savings resulted from using two newer, minimally invasive techniques: modified lateral minimally invasive and anterior-lateral muscle-sparing.¹⁰

This measure aims to address these example areas of potential improvement. Given the frequency of hip arthroplasty among Medicare beneficiaries, the use of this episode-based cost measure can provide clinicians with information to improve care outcomes and reduce future health care costs.

2.2 Performance Gap

2.2.1 Rationale

The 2010 prevalence of total hip arthroplasties in the United States population was 0.8 percent, increasing with age to 1.5 percent at sixty years and 5.9 percent by ninety years of age.¹¹ There were an estimated 2.5 million individuals with a total hip arthroplasty in 2010, and the demand for primary hip arthroplasties is estimated to grow by 174 percent between 2005 and 2030.¹² Opportunities for improvement for elective primary hip arthroplasty include appropriate use of institutional PAC (e.g., having patients receive post-procedure treatment in a home health or outpatient therapy setting), improving adherence to correct treatment guidelines, and increasing the use of optimal surgical techniques. The Elective Primary Hip Arthroplasty episode-based cost measure was recommended for development by an expert clinician committee—the Musculoskeletal Disease Management – Non-Spine 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.

⁶ Mujica-Mota, Ruben E., et al. (2012). "Determinants of Demand for Total Hip and Knee Arthroplasty: A Systematic Literature Review." *BMC Health Services Research* 12: 225.

⁷ Cobos, Raquel, et al. (2010). "Variability of Indication Criteria in Knee and Hip Replacement: An Observational Study." *BMC Musculoskeletal Disorders* 11: 249.

⁸ Ibid.

⁹ Sibia, U. S., et al. (2017). "The Impact of Surgical Technique on Patient Reported Outcome Measures and Early Complications After Total Hip Arthroplasty." *The Journal of Arthroplasty* 32(4): 1171-1175.

¹⁰ Goldstein, J. P., et al. (2016). "The Cost and Outcome Effectiveness of Total Hip Replacement: Technique Choice and Volume-Output Effects Matter." *Applied Health Economics and Health Policy* 14(6): 703-718.

¹¹ Kremers et al. (2015). "Prevalence of Total Hip and Knee Replacement in the United States." *Journal of Bone and Joint Surgery* 97(17):1386-97.

¹² Kurtz et al. (2007). "Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030." *Journal of Bone and Joint Surgery* 89(4):780-5.

2.2.2 Performance Scores

Performance scores are provided for 2,030 clinician group practices (identified by Tax Identification Number [TIN]) and 5,957 practitioners (identified by combination of TIN and National Provider Identifier [NPI]). These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a Merit-based Incentive Program (MIPS) eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced APM participation). This table uses a testing volume threshold of 10 episodes.

Table 1: Distribution of Performance Scores

Metric	TIN	TIN-NPI
Mean score	\$19,801	\$19,116
Standard deviation	\$2,278	\$2,315
Score IQR	\$3,034	\$3,211
Score percentile		
10 th	\$17,006	\$16,379
20 th	\$17,833	\$17,048
30 th	\$18,451	\$17,676
40 th	\$19,031	\$18,252
50 th	\$19,606	\$18,824
60 th	\$20,147	\$19,472
70 th	\$20,850	\$20,187
80 th	\$21,651	\$21,001
90 th	\$22,668	\$22,191

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 Elective Primary Hip Arthroplasty 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 Elective Primary Hip Arthroplasty 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

2,030 clinician group practices and 5,957 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 10 or more THA episodes during the measurement period. Episodes from all 50 States and D.C. in the following settings were included: acute inpatient (IP) hospitals, hospital outpatient departments (HOPD), ambulatory/office-based care centers, and ambulatory surgical centers (ASC).

3.1.6 Patient Cohort Included in the Testing and Analysis

108,895 Medicare beneficiaries (from 111,434 episodes) were included in TIN level testing and analysis, and 98,001 beneficiaries (from 100,340 episodes) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the Elective Primary Hip Arthroplasty measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who received an elective primary hip replacement procedure during the measurement period as identified by the episode trigger Current Procedural Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) code 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 receiving elective total hip replacement 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 ambulatory/office-based care, IP hospital, OP hospital, or ASC setting.
- Where there is an IP stay concurrent with the trigger, it occurs in a short-term stay acute hospital as defined by subsection (d).¹³
- If the trigger event was performed inpatient, the IP stay was billed as a major joint replacement or reattachment of lower extremity.
- The trigger event was performed unilaterally, i.e., without a second procedure on the opposite hip occurring in the 90 days following the initial hip replacement.
- The beneficiary was not diagnosed with congenital deformity of the hip during the lookback period.
- The beneficiary was not diagnosed with osteomyelitis of the hip or femur during the lookback period.
- The beneficiary was not diagnosed with a septic hip joint during the lookback period.
- The trigger claim did not include a diagnosis of cancer of the hip or femur.
- The trigger claim did not include a diagnosis of hip or femur fracture or other trauma.
- The episode is not an outlier case.

To determine whether the THA 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 THA measure's inclusion criteria have only a minimal effect on the percentage of beneficiaries of any particular demographic or patient characteristic. The difference between beneficiaries being included or not included in the measure is less than between zero and 1.9 percentage points across each of the characteristics in the analysis at TIN level testing, and between zero and 2.7 percentage points at TIN-NPI level testing. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 29.1 percent, compared to 28.5 percent with the inclusion criteria applied at TIN level testing and 28.6 percent at TIN-NPI level testing. The difference in the percentage of beneficiaries for race with and without the inclusion criteria is less than one percentage point for all categories at the TIN level and most categories at the TIN-NPI level, and is between 1.0 and 1.2 percentage points for two TIN-NPI level categories (white and black). The breakdown of male and female beneficiaries remains the same when comparing the use of inclusion criteria at both levels of

¹³ 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](#).

testing, with approximately 61 percent female and 39 percent male either with or without the application of inclusion criteria, with percentage point shifts of less than 0.2. These results indicate that there is minimal shift in patient characteristics when applying the inclusion criteria listed above at both TIN and TIN-NPI level testing.

3.1.7 Sample Differences

n/a

3.1.8 Social Risk Factors Included in Analysis

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

- Income (ACS)
 - Low Income: median income < 33rd percentile nationally
 - Medium Income: median income in the interval spanning the 33rd percentile to the 66th percentile nationally
 - High Income: median income > 66th percentile
- Education (ACS)
 - Education < High School: when % with < high school education is the highest for a given Census Block Group
 - Education = High School: when % with only high school is the highest
 - Education > High School: when % with > high school is the highest
- Employment (ACS)
 - Unemployment Rate > 10%
 - Unemployment Rate <= 10%
- Race (EDB)
 - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
 - Female, male
- Dual status (CME)
 - Full dual, partial dual, non-dual

3.2 Reliability Testing

3.2.1 Level of Reliability Testing

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

3.2.2 Method of Reliability Testing

Data Element Reliability

The THA 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.¹⁴ The fiscal year 2018 Medicare fee-for-service program proper payment rate was 91.9 percent.¹⁵ CMS continues to perform successful corrective actions and give providers additional education to ensure accurate billing.

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

Measure Reliability

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

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

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

Where:

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

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

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

3.2.3 Statistical Results from Reliability Testing

Measure Reliability

At a testing volume threshold of at least 10 episodes, the mean reliability is 0.85 for TINs and 0.78 for TIN-NPIs. The mean reliability continues to increase at the 20 and 30-episode volume thresholds. 100 percent of TINs and TIN-NPIs have reliability greater than or equal to 0.4 at the 10, 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	0.85	100.0%	0.78	100.0%

¹⁴ 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>

¹⁵ Ibid.

Volume Threshold (# episodes)	TIN		TIN-NPI	
	Mean Reliability	% ≥ 0.4	Mean Reliability	% ≥ 0.4
20	0.90	100.0%	0.86	100.0%
30	0.93	100.0%	0.89	100.0%

3.2.4 Interpretation

Measure Reliability

Overall reliability of the THA measure is 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.¹⁶

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 THA measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. These expert panels methodically assessed 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 Musculoskeletal Disease Management – Non-Spine Clinical Subcommittee, (ii) the Elective Primary Hip Arthroplasty 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 29 members with clinical experience in non-spinal musculoskeletal disease management, affiliated with 26 specialty societies. The Clinical Subcommittee provided input at an in-person meeting in April 2018 on the measure to develop, scope, and suggested composition of a smaller, targeted workgroup to provide detailed input on each aspect of measure specifications. The Elective Primary Hip Arthroplasty workgroup was composed of 15 members, affiliated with 14 specialty societies, including the American Academy of Orthopaedic Surgeons, the American Association of Hip and Knee Surgeons, the American Occupational Therapy Association, and the American Society of Anesthesiologists. The workgroup considered empirical analyses and their clinical expertise to provide input during

¹⁶ 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

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

One of the key roles of the measure-specific workgroup was to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in hip replacement, thus preventing inclusion of unrelated cost variation in this measure. Assigned services were defined separately for the pre- and post-trigger windows and include hip replacement surgery, evaluation, testing, treatment, complications, and follow-up care. Assigned services could occur in the emergency department; outpatient facilities; inpatient facilities, including long-term care hospitals; inpatient rehabilitation facilities; and home health service settings or could relate to durable medical equipment, prosthetics, orthotics, and supplies.

Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the THA measure by examining differences in risk-adjusted cost for known indicators of resource or service utilization based on a literature review, specifically complications related to the procedure that occur in the post-trigger period. For this analysis, we compared the ratio of observed to expected spending (henceforth called the "O/E cost ratio") for Elective Primary Hip Arthroplasty episodes with and without acute (re)admissions and episodes with and without PAC occurring in the post-trigger period. This analysis sought to confirm the expectation that the THA 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 THA measure were classified into clinically coherent groups of services, called "clinical themes." The THA measure clinical themes are:

- **Preoperative Work-Up:** includes services such as routine chest x-rays; electrocardiograms; laboratory testing, such as blood counts, electrolytes and basic metabolic testing, and coagulation testing; other diagnostic studies, e.g., x-rays of the knee; diagnostic procedures; and office or outpatient evaluations linked to the surgery
- **Post-Acute Care and Rehabilitation:** includes procedures or therapy performed after the surgery, e.g., physical and occupational therapy, arthrocentesis, x-rays, laboratory testing, other diagnostic or therapeutic procedures, and durable medical equipment (DME) related to after-care, such as wheelchairs, canes, walkers, or crutches
- **Wound Care:** includes DME and supplies related to the after-care, e.g., tape, wound filler, dressing, wound care sets, canisters, or negative wound therapy electrical pumps
- **Imaging:** includes CT scans of the leg or abdomen, routine chest x-rays, or ultrasound scans of veins in the outpatient setting
- **Surgical Site Infection (SSI):** includes procedures or testing to diagnose and treat SSI, e.g., inpatient or outpatient (including emergency department) care for cellulitis,

infections or inflammatory reactions; laboratory testing; insertion of an infusion device; or removal of liners

- **Thromboembolism (DVT/PE):** includes emergency department visits, inpatient care, critical care, and other diagnostic or therapeutic procedures to diagnose and treat pulmonary embolism or other venous embolism and thrombosis, e.g., outpatient visits, major chest procedures, blood tests, or ultrasounds
- **Transfusions / Bleeding:** includes laboratory testing, CT scans, x-rays, or diagnostic procedures to diagnose bleeding or the need for transfusions in the inpatient or outpatient (including emergency department) setting
- **Sepsis:** includes emergency department visits and inpatient hospital and critical care (including ventilator support) to treat sepsis after the procedure
- **Cardiovascular Complications:** includes emergency department and inpatient hospital care, diagnostic cardiac catheterization, imaging (including ultrasounds), cardiac monitoring, and other diagnostic or therapeutic procedures to address cardiac complications, as well as coronary bypass or other cardiovascular procedures within a week of the original procedure
- **Prosthetic Complication / Revision:** includes wound dehiscence and infections leading to procedures; therapeutic procedures on joints; imaging; surgery to remove or revise the prosthetic; physical therapy when linked to a complication; and other diagnostic procedures, in addition to DME to address complication or revision of a prosthetic

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

We expected that the four themes most related to complications (SSI, Sepsis, Cardiovascular Complications, and Prosthetic Complication/Revision) would have the highest correlation with risk-adjusted episode cost, as complications are likely associated with high cost even after accounting for beneficiary characteristics.¹⁷ We would expect similar trends for the Transfusions/Bleeding and DVT/PE themes, as they contain services relating to complications. By contrast, we expected that the Preoperative Work-Up, Imaging, and Post-Acute Care and Rehabilitation themes have more nuanced, offsetting effects. While higher costs for these types of visits can directly increase the costs of an episode, research indicates that pre- and post-surgical interventions can be associated with lower total resource use by saving on later costs, also discussed in Section 2.1.2.¹⁸ Therefore, it is possible the correlation of the measure with these types of costs is lower than for complications.

3.3.3 Statistical Results from Validity Testing

For the first analysis of validity, the mean O/E cost ratio for all episodes is 1.00. The mean O/E cost ratio for episodes with complications that result in hospital (re)admission during the post-trigger period is 1.67, compared with 0.98 for episodes without readmissions services relating to complications during the post-trigger period. In addition, the mean O/E cost ratio for episodes with PAC in the post-trigger period is 1.07, compared with 0.84 for episodes without PAC in the post-trigger period. Table 3, below, provides further detail, including a percentile distribution of

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

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

the O/E cost ratios for all episodes and for episodes with and without services relating to complications from THA.

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	0.29	0.62	0.70	0.74	0.82	0.92	1.08	1.37	1.61	2.13
Episodes with Downstream Acute (Re)admission	1.67	0.46	0.83	0.98	1.09	1.29	1.65	2.03	2.30	2.43	2.67
Episodes without Downstream Acute (Re)admission	0.98	0.26	0.62	0.70	0.74	0.82	0.91	1.06	1.30	1.52	1.95
Episodes with PAC	1.07	0.31	0.64	0.73	0.78	0.88	1.00	1.16	1.48	1.72	2.20
Episodes without PAC	0.84	0.17	0.59	0.66	0.70	0.77	0.83	0.88	0.94	0.98	1.64

The clinical themes analysis demonstrates that there is moderate correlation between the clinical themes and risk-adjusted cost. Of these correlations, the highest are found with the SSI (correlation: 0.38), Sepsis (correlation: 0.34), and Cardiovascular Complications (correlation: 0.33) themes, which does agree with our initial expectations. By contrast, the Imaging (correlation: -0.15) and Preoperative Work-Up (correlation: 0.04) themes had weaker correlation with risk-adjusted cost, as expected. Unlike the other clinical themes, results for Imaging indicate a weak negative correlation with risk-adjusted cost, meaning that lower Imaging costs are weakly associated with higher risk-adjusted cost.

3.3.4 Interpretation

As expected, the average O/E cost ratio for episodes with post-trigger complications is much higher than for episodes without downstream complications. Similarly, the mean O/E cost ratio for episodes with PAC is higher than for those without PAC, which is appropriate given the often substantial cost of PAC. These results demonstrate that the Elective Primary Hip Arthroplasty measure is able to capture accurately higher resource use.

The clinical themes analysis demonstrates that high risk-adjusted cost is moderately associated with themes related to complications, and is less correlated to themes relating to testing, imaging, and physician visits. 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 laboratory testing and x-rays or other imaging. Importantly, we see that correlation with risk-adjusted cost is relatively consistent across both high-cost themes such as PAC (average cost: \$2,841) and lower-cost themes such as DVT/PE (average cost: \$582). 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 THA measure to capture a homogenous patient population within the scope of the measure focus on hip replacement and ensure that episodes provide meaningful information to attributed clinicians or as part of data processing to ensure that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each

episode. For the exclusions analysis, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- *Episodes where beneficiary death date occurred before the episode end.*
 - These episodes are excluded for all measures due to the potential to reflect inaccurately a clinician's performance. Episodes where the beneficiary died may be unusually high-cost, due to perimortem treatment costs, or unusually low-cost, due to the truncated episode window. Neither of these cases accurately reflects the efficiency of the clinician performing the treatment.
- *Episodes in which the beneficiary underwent a staged or same-day bilateral hip replacement procedure.*
 - The recovery care for patients with bilateral hip arthroplasty—same-day or staged—is very different compared to unilateral hip arthroplasty due to the loss of mobility.
- *Episodes where the hip replacement is performed due to cancer, hip fracture, or trauma.*
 - Cancer treatment is an unusual reason for a hip arthroplasty, indicating potentially atypical treatment patterns. Similarly, patients who suffer from a hip fracture or other trauma that is treated with a THA may experience different treatment and are less likely to undergo the procedure on an elective basis. Both cancer and hip fracture or trauma patients may also be higher-risk or otherwise require additional care.
- *Episodes where the beneficiary has congenital deformity of the hip, osteomyelitis of the hip or femur, or a septic joint.*
 - Beneficiaries with these disorders may require substantively different services, such as long-term antibiotic treatment, and typically more complex care that differs from the routine care for this elective procedure.
- *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 O/E cost ratios (calculated by applying existing risk factor coefficients to the excluded episodes) for excluded episodes. We then compared the cost characteristics of the excluded episodes to those of final episodes included in measure calculation to assess the distinctness between the two patient cohorts. A full list of the exclusions and details used for the THA measure is provided in the Measure Codes List.¹⁹

3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics and O/E cost ratios for the THA measure exclusions. Cost statistics are also provided for the set of final episodes included in the THA

¹⁹ CMS, "Elective Primary Hip Arthroplasty 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>.

measure for comparison, with a testing volume threshold of 10 episodes at the TIN and TIN-NPI levels.

Table 4: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost			O/E		
	#	%	Mean	Percentile		Mean	Percentile	
				10 th	90 th		10 th	90 th
All Episodes Meeting Triggering Logic	132,835	100.00%	\$20,393	\$14,265	\$31,394	1.04	0.73	1.49
Beneficiary Death in Episode	839	0.63%	\$27,710	\$14,473	\$48,551	1.12	0.55	1.89
Bilateral Hip Arthroplasty, Primary and Staged	1,823	1.37%	\$28,854	\$6,748	\$46,491	2.00	0.85	3.54
Congenital Deformity of the Hip	473	0.36%	\$22,230	\$14,259	\$35,059	1.19	0.74	1.75
Hip Arthroplasty for Cancer	144	0.11%	\$32,493	\$16,074	\$54,137	1.58	0.76	2.81
Hip Fracture/Trauma (Reason for Hip Arthroplasty)	6,412	4.83%	\$29,731	\$15,248	\$48,290	1.35	0.72	2.13
Osteomyelitis of Hip and Femur	39	0.03%	\$27,218	\$14,227	\$62,157	1.47	0.56	3.05
Septic Joint	213	0.16%	\$22,352	\$3,067	\$44,468	1.27	0.27	2.83
Outlier	2,390	1.80%	\$41,877	\$14,203	\$75,911	1.82	0.50	3.52
<i>Final Episodes (TIN)</i>	111,434	83.89%	\$19,091	\$14,257	\$27,935	0.98	0.73	1.33
<i>Final Episodes (TIN-NPI)</i>	100,340	75.54%	\$18,907	\$14,244	\$27,423	0.97	0.73	1.30

3.4.3 Interpretation

The statistical results indicate that many excluded episodes differ in both mean observed costs and mean O/E cost ratios and have larger variation compared to the final set of episodes and support the exclusion of these episodes to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for each exclusion is provided below.

Episodes ending in death: Results show a large difference between mean observed episode cost for episodes ending in death and the final set of episodes: \$27,710 compared to \$19,091 at the TIN level and \$18,907 at the TIN-NPI level. This difference becomes more pronounced in the right tail of episodes, with the episode cost at the 90th percentile at \$48,551 compared to \$27,935 and \$27,423 for the final episodes at the TIN and TIN-NPI levels, respectively. This suggests that there can be high variation in perimortem costs resulting in higher cost episodes, despite the episodes being by definition truncated. As such, episodes ending in death are excluded to avoid the potential of clinicians being incentivized to avoid treating complex, high-risk patients.

Episodes in which the beneficiary underwent a staged or same-day bilateral hip replacement procedure: In addition to being approximately 50 percent more expensive than final episodes at both the TIN and TIN-NPI levels, bilateral total hip replacements are infrequent (1,823 episodes, compared to more than 100,000 final episodes at both TIN and TIN-NPI levels). Removing these episodes ensures the measure compares only patients who require similar treatment plans—as patients undergoing bilateral procedures may be sicker, the procedure is more expensive, and the recovery is different—while still being representative of the majority of THA patients.

Episodes where the hip replacement is performed due to cancer, hip fracture, or trauma: Episodes where THA was indicated due to cancer, hip fracture, or trauma are also substantially more expensive than the final episodes (mean observed cost of \$32,493 and \$29,731, respectively) with this difference becoming more pronounced at the 90th percentile. Such a difference is in line with expectations, given the supplemental care these more complex patients

may require, such as admittance through the emergency department. These episodes are excluded as THA due to cancer, hip fracture, or trauma is typically not elective and so is outside the scope of the measure intent.

Episodes where the beneficiary has congenital deformity of the hip, osteomyelitis of the hip or femur, or a septic joint: While the mean observed cost of episodes in which the beneficiary was diagnosed with these conditions are only slightly more expensive (mean observed costs of \$22,230, \$27,218, and \$22,352, respectively) than TIN or TIN-NPI level final episodes, the episodes at the 90th percentile are markedly different from the final set of episodes. The observed episode cost at the 90th percentile for congenital deformity is \$35,059; Septic Joint is \$44,468, and for osteomyelitis is \$62,157. In comparison, the observed cost at the 90th percentile for final episode is under \$28,000 at TIN and TIN-NPI levels. As such, these small and unique patient groups are excluded from the measure, as they are not clinically comparable to the majority of the Elective Primary Hip Arthroplasty patient cohort.

Outlier cases: The ratio of observed to expected episode cost ranges from 0.5 at the 10th percentile to 3.52 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 121 risk factors. This measure's risk adjustment model is not stratified by risk categories.

The risk adjustment model for the Elective Primary Hip Arthroplasty 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 is currently using anticoagulants, which may indicate the potential for additional post-surgical complications;
- whether the beneficiary has been diagnosed with chronic pain or avascular necrosis of the hip, both of which may necessitate different post-surgical care and analgesia;
- whether the beneficiary is diagnosed with opioid dependence, possibly indicating different post-surgical analgesia and increased monitoring;
- whether the beneficiary has a history of DVT/PE, antiplatelet use, post-traumatic arthritis of the hip, or a condition which puts them at higher risk of dislocating their hip, any of which would require additional monitoring and possibly supplemental treatment;
- whether the beneficiary has been diagnosed with one or more inflammatory arthropathies, possibly requiring different post-surgical care because of the potential for multi-joint involvement;
- whether the beneficiary is obese, posing additional surgical, post-surgical, and rehabilitation challenges;
- whether the beneficiary has a sickle-cell disorder, requiring more complex pain management;
- whether the beneficiary is a current or former smoker, which may impair their healing and compromise their respiratory function;
- whether the beneficiary has a history of spinal disorders, which may make post-operative recovery more difficult; and
- whether the beneficiary seems frail, likely necessitating additional PAC and closer monitoring, as indicated by the presence of one or more of the following diagnoses or treatments:
 - Anemia;
 - Dementia;
 - Home health;
 - Home hospital bed;
 - Home oxygen;
 - Nursing physician facility visits;
 - Recent admission to long-term care hospital;
 - Recent all-cause admission;
 - Walking aid; or
 - Wheelchairs.

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 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.

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

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 THA measure methodology.

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

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 the National Quality Forum (NQF) #2158: MSPB-Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Pope et al (2011) report.²² 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.

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

²⁰ CMS, "Elective Primary Hip Arthroplasty 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>.

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

²² Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

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.

61.4 percent of the beneficiaries in the THA measure are female. The majority of the beneficiaries (91.8%) 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 1.3 percent of beneficiaries are classified below a high school education level, the vast majority of episodes are classified at greater than high school level (88.2%). Finally, 19.4 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 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.

Secondly, we analyzed the impact of adding social risk variables on overall model performance by looking at the differences in the O/E 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 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 ratio of observed over expected costs for 94.0 percent of TINs and 94.7 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 0.993 for both. These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

Due to the limited impact of social risk factor effects under the current risk adjustment model, we believe the THA measure risk adjustment model sufficiently accounts for the effects of social risk factors 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) O/E cost ratios and predictive ratios to examine the fit of the models at different levels of patient complexity, and (3) coefficient estimates, standard errors, and p-values.

- 1) *R-squared and adjusted R-squared* were calculated for the overall measure. The results should be evaluated in the context of the service assignment rules, which indicate which

costs are counted in the measures and which costs are not counted. This is an important distinction from all-cost measures, as a low R-squared does not necessarily indicate that a measure reflects variation unrelated to clinical care, while a high R-squared does not necessarily indicate the opposite; instead, the risk adjustment models must be evaluated in concert with the service assignment rules. These results are provided in Section 3.5.7.

- 2) *Predictive ratios and O/E cost ratios* were calculated for each “risk decile” for the episode group. A “risk decile” is based on the risk scores, which indicate how costly episodes are expected to be, as predicted through risk adjustment. After arranging episodes into deciles based on their risk score, we calculated the predictive ratios and average O/E cost ratios for each decile. The predictive ratio aims to examine the fit of the model at different levels of patient complexity to examine the model’s ability to predict both very low and high cost episodes, and is calculated using the formula of average (expected cost)/average (observed cost) for all episodes in each decile. Similarly, the O/E cost ratio demonstrates the model’s prediction accuracy, and is calculated using the formula of average (observed cost/expected cost) for all episodes in each decile. These are discussed in Sections 3.5.8 and 3.5.9.
- 3) *Coefficient estimates, standard errors, and p-values* were run to consider the extent to which the coefficients for the risk factor covariates are predictive of episode cost. Results for individual risk adjustment variables should be viewed in the context of the entire model, rather than being analyzed individually. For instance, coefficients indicate the incremental effect of a model variable, holding all other variables fixed. As another example, interactions between model variables must be interpreted in concert with the effects of those variables in isolation.

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

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Elective Primary Hip Arthroplasty cost measure, calculated by dividing explained sum of squares by total sum of squares, is 0.17. The adjusted R-squared is also 0.17.

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. 2011.²⁴

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 very close to one across risk deciles, indicating that the model is accurately predicting actual episode cost. Full results can be seen the NSDR Addendum.

²³ 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>.

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.

3.5.9 Statistical Risk Model Calibration – Risk Decile

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

3.5.10 Interpretation

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

As demonstrated in Section 3.5.8 and 3.5.9, the average O/E cost ratios and the predictive ratios for all risk deciles are very 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.

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

3.6.2 Statistical Results

Key findings show that, generally, there is a large performance difference among clinicians in the THA measure:

- (i) the 99th percentile of the measure score is nearly 1.7 times the 1st percentile at both the TIN level and TIN-NPI levels and
- (ii) the THA measure score at the 90th percentile is approximately 33 percent greater than the score at the 10th percentile at both the TIN and TIN-NPI level.

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

The results also show that there is not systemic regional difference in clinician score. For instance, the mean scores for clinicians across nine census divisions (excluding 'Unknown') are within a less than \$1,900 range at the TIN level (i.e., \$18,853 – \$20,760) and \$2,000 range at the TIN-NPI level (i.e., \$18,352 to \$20,318). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas with a difference of \$322 at the TIN level and \$20 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 hip arthroplasties. 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 \$19,093 to \$21,261 and a range in mean TIN-NPI score of \$18,197 to \$20,910, 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 THA measure scores, indicating the measure's ability to capture differences in performance. Our findings regarding variation in measure scores are consistent with expert clinician input. Overall, as expected, results show that clinicians are not being systematically penalized or rewarded due to risk score decile given the current THA measure design (i.e., the differences in measure score are not due to the risk profile of patients).

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the THA 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 the EDB, the beneficiary does not appear in the EDB, or the beneficiary death date occurs before the episode trigger date.

The THA 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 THA measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the number of TINs and TIN-NPIs who had at least one episode excluded due to missing data. The missing data categories are:

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

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>.

Table 5: Missing Data Categories for the Elective Primary Hip Arthroplasty Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	0	0	0
Death before trigger	*	*	*
Other primary payer	13,920	2,724	8,829
Not continuously enrolled	8,070	2,273	6,679

*asterisk indicates that there were fewer than 11 episodes

3.7.3 Interpretation

As the THA measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency of missing data as noted above. Acumen takes measures to address cases of missing or inaccurate information in claims data.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

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

4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

4.3 Data Collection Strategy

4.3.1 Data Collection Strategy Difficulties

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

4.3.1.1 Data Collection

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

4.3.1.2 Missing Data

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

4.3.1.3 Sampling

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

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

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

5.1.2 Feedback on the Measure and Development Process

5.1.2.1 Technical Assistance Provided During Development or Implementation

Development: Field Testing

Acumen and CMS conducted a national field test of 11 episode-based cost measures developed during 2018, including the Elective Primary Hip Arthroplasty 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 episode-based cost measures: 14,237 TINs; 63,984 TIN-NPIs
- Testing sample for Elective Primary Hip Arthroplasty: 2,041 TINs; 5,992 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, a total of 2,388 field test reports for episode-based cost measures were downloaded by 403 clinician groups (TINs) and 1,985 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the workgroup to consider in recommending refinements to the measures based on the testing data and feedback.

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

Data Provided During Field Testing

Each field test report contained the following sheets:

- High-level summary results across all episode-based cost measures being field tested
- Results for each measure including cost measure score and breakdown of episode cost compared to the national average and TIN/TIN-NPIs with a similar patient case mix (or risk profile)

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>.

- 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 five 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 end of the feedback period.

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 Elective Primary Hip Arthroplasty cost measure. After the Measure Applications Partnership (MAP) Clinician Workgroup meeting in December 2018, there was another public comment period on their preliminary recommendations, which received 23 comments across the 11 measures, with three comments specific to the Elective Primary Hip Arthroplasty cost measure.³⁰ These public comment periods were facilitated by NQF. Stakeholders were able to submit their comments via the NQF website.

5.1.2.4 Feedback from Providers being Measured

Field Testing

The Field Testing Feedback Summary Report presents all feedback gathered during the field testing period. The following list synthesizes some of the key points that were raised through the field testing feedback period:

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

³⁰ Measure Applications Partnership, *National Quality Forum*.

https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.

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

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

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

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

5.1.2.5 Feedback from Other Users

Pre-Rulemaking

The MAP recognized the importance of cost measures to the MIPS program and conditionally supported the Elective Primary Hip Arthroplasty 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. The 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 Elective Primary Hip Arthroplasty measure made after consideration of field testing analyses and stakeholder feedback are:

- **Triggers:** Removed CPT/HCPCS 27132 (conversion to THA) as trigger code
- **Service Assignment:** Added new assigned services pre- and post-trigger:
 - Pre-Trigger:
 - Hip imaging and cardiovascular tests within 30 days
 - Post-Trigger
 - Sepsis related to THA within 30 days
 - Sepsis unspecified within seven days

- **Risk Adjustment:**
 - Modified existing risk adjustors as follows:
 - Set lookback period for all risk adjustors to 120 days to align with the Knee Arthroplasty measure
 - Removed F11 (opioid related disorders) from high-risk dislocators and create a new risk adjustor for that code alone
 - Refined coding for high-risk dislocators and inflammatory arthropathies to be more hip-specific
 - Combined risk adjustors for spine surgery, intervertebral disc disorders, and spinal stenosis into one risk adjustor for spinal disorders
 - Added DGN D6832 to the anticoagulant use risk adjustor
 - Created additional risk adjustors for:
 - Antiplatelet therapy
 - The following frailty indicators:
 - Anemia
 - Dementia
 - Home health
 - Home hospital bed
 - Home oxygen
 - Nursing physician facility visits
 - Recent admission to long-term care hospital
 - Recent all-cause admission
 - Walking aid
 - Wheelchairs
 - Removed septic joint as a risk adjustor (changed to an exclusion)
- **Exclusions:** Excluded patients with any of the following diagnoses:
 - Congenital deformity of the hip
 - Osteomyelitis of the hip and femur (distinct from the inflammatory arthropathies risk adjustor)
 - Septic joint

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

Elective Primary Hip Arthroplasty Measure-Specific Workgroup Members:

Adam Rana, American Association of Hip and Knee Surgeons

Adolph Yates, American Academy of Orthopaedic Surgeons

Andrew Gordon, American Academy of Physical Medicine and Rehabilitation

Anita Bemis-Dougherty, American Physical Therapy Association

David Jevsevar, American Association of Hip and Knee Surgeons

Dennis Rivenburgh, American Academy of Physician Assistants

Dheeraj Mahajan, AMDA - The Society for Post-Acute and Long-Term Care Medicine

Edward Mariano, American Society of Anesthesiologists

Harold Rees, American Association of Hip and Knee Surgeons

Jeremy Furniss, American Occupational Therapy Association

Judy Dusek, National Association of Clinical Nurse Specialists

Marc DeHart, American Academy of Orthopaedic Surgeons

Mark Levine, The American Geriatrics Society

Robin Kamal, American Academy of Orthopaedic Surgeons

Vasili Karas, American Academy of Orthopaedic Surgeons

The Elective Primary Hip Arthroplasty workgroup is composed from the larger Musculoskeletal Disease Management – Non-Spine 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>.