Partnership for Quality Measurement
Powered by Battelle

WHAT GOOD LOOKS LIKE – COST MEASURE EXAMPLE

Note: *The information provided in this form is intended to aid the committee and other interested parties in understanding to what degree the items in the measure submission form addresses each of the five PQM Measure Evaluation Rubric domains.*

Some example responses in this form are for illustration only and do not affect the measure's endorsement status. The information in this document has been derived and modified from a measure submission from the Centers for Medicare & Medicaid Services (measure steward) and Acumen, LLC. (measure developer)

Measure Identification

1.0 New or Maintenance*

Select whether this is a new measure (i.e., submitted for initial endorsement) or a maintenance measure (i.e., submitted for maintenance endorsement).

☐ New ☒ Maintenance

1.1 Measure Structure*

Select the structure that best describes your measure.

[DO NOT select "Instrument + Derived Measure Set"; see top of page 1 for appropriate templates/instructions for instruments and derived measure submissions]

☒ Single Measure *[focuses on a single structure, process or outcome]*

☐ Composite Measure *[combines multiple individual measures into a single score]*

☐ Instrument + Derived Measure Set *[See the Overview of Instrument-Derived Measure Set Submission Framework for more information regarding these measures]*

1.2 *[If 1.0 = Maintenance]* **Measure or Instrument Lookup***

Start by typing the CBE ID, measure title, or instrument title, and select an autocomplete option.

[Measures seeking initial endorsement and NEW instruments will be assigned a CBE ID after ITS is saved.]

CBE #3623

1.3 Electronic Clinical Quality Measure (eCQM)*

Is this measure an eCQM (i.e., based on the Quality Improvement Core [QI-Core] or the Quality Data Model [QDM], Clinical Quality Language [CQL], and specified using value sets)? Includes hybrid measures (i.e., measures that use electronic clinical data from electronic health records and another data source, such as claims).

☒ No ☐ Yes

1.4 Measure Title*

***For a single measure**, the measure title should include the type of score (e.g., rate, count, composite), the measure focus, and the target population. Title example: The rate [type of score] of 30-day all-cause mortality [measure focus] among patients discharged from an acute inpatient facility with a diagnosis of acute myocardial infarction [target population]. **For an instrument**, the measure title should be the full title of the instrument, including health care setting or population (as appropriate). **For single measures and instruments**, do not use acronyms unless they are defined in the measure title. **For an IDM**, the measure title should include the instrument name and population, and that IDM's specific focus; the IDM measure title may use just the instrument acronym (e.g., CAHPS). IDM title example: [acronym] [instrument target population] [IDM focus].*

Elective Primary Hip Arthroplasty Episode-Based Cost Measure

Endorsement and Maintenance (E&M) Cycle*

Select the intended measure review cycle for endorsement consideration.

Spring 2025	**Fall 2025**	**Spring 2026**
ITS deadline: Tuesday, April 1, 2025 Full Submission deadline: Thursday, May 1, 2025	ITS deadline: Wednesday, October 1, 2025 Full Submission deadline: Monday, November 3, 2025	ITS deadline: Wednesday, April 1, 2026 Full Submission deadline: Friday, May 1, 2026
☐ Spring 2025	☐ Fall 2025	☒ Spring 2026

[Scroll to the bottom of the page and click "Save". Then you will be able to continue completing the ITS form.]

*[**NOTE:** Once you have saved the measure, the ONLY field in the Measure Identification section you will be able to edit is 1.4 Measure Title. Contact PQMSupport@battelle.org for assistance with other changes in the Measure Identification section.]*

Reminder

For cost measures, the type of cost measure (e.g., episode based care, total cost of care) should be included in the title. If the measure has a short name or abbreviation often included in the title (e.g., at the end in parentheses), please include in the submission.

Note: The original title of this measure has been modified to align with the requirements for cost measures.

1. Measure Information

1.5 Project*

Choose the project that you expect to review the measure. For instrument submissions, select the project that you expect to review the associated IDMs (Note: All IDMs associated with an instrument will be reviewed by the same project). To see the project descriptions and examples of project-related measures, please refer to the E&M projects page on the PQM website. ***Note:*** *Battelle may reassign the measure to a different project following internal review. Choose one.*

- ☐ Advanced Illness and Post-Acute Care
- ☒ Cost, Resource Use, and Efficiency
- ☐ Initial Recognition and Management
- ☐ Management of Acute Events, Chronic Disease, Surgery, and Behavioral Health
- ☐ Primary Prevention

1.6 Measure Description*

For a single measure, briefly describe the type of score, measure focus, target population, and timeframe. For an instrument submission, briefly describe its primary objective (what it is designed to measure or assess), list the key elements or sections of the instrument, and describe how the instrument is administered and how responses are collected. Note: The numerator and denominator have separate fields (applicable for single measure and IDM submissions).

The Elective Primary Hip Arthroplasty episode-based cost measure evaluates a clinician's risk-adjusted cost to payers for patients who receive an elective primary hip arthroplasty during the performance period. The measure score is a clinician's risk-adjusted cost for the episode group (i.e., all services and items related to the elective primary hip arthroplasty procedure, including pre-operative, operative, and post-operative care within a defined time window) averaged across all episodes attributed to the clinician. This measure includes costs of services that are clinically related to the attributed clinician's role in managing care during each episode from the 30 days prior to the clinical event that opens or "triggers" the episode, through 90 days after the trigger.

Quick Tip

Include a description of the type of cost measure.

This measure is an episode-based cost measure that evaluates a clinician's risk-adjusted cost to Medicare for episodes involving elective primary hip arthroplasty. The description of the measure includes details such as the inclusion of all clinically related services and items within the episode, spanning from 30 days prior to the clinical event that triggers the episode through 90 days after, as well as the averaging of costs across all episodes attributed to the clinician. These elements clearly define the measure's focus on episode-based costs.

1.7 Measure Type*

Choose one. If "Other", please specify.

- ☒ Cost/Resource Use
- ☐ Efficiency
- ☐ Intermediate Outcome
- ☐ Outcome
- ☐ Population Health
- ☐ Process
- ☐ Patient-reported Outcome Performance Measure (PRO-PM)
- ☐ Patient-reported Experience Performance Measure (PRE-PM)
- ☐ Structure
- ☐ Other

1.7a Other Measure Type* *(Please specify) [character limit: 255]*

1.8 Level of Analysis*

Select the level(s) of analysis for which the measure is specified and tested. For instrument submissions, select the level(s) for which the associated IDMs are specified and tested. Choose all that apply. If "Population of Geographic Area" or "Other", please specify.

- ☐ Accountable Care Organization
- ☒ Clinician: Group/Practice
- ☒ Clinician: Individual
- ☐ Facility
- ☐ Health Plan
- ☐ Program (e.g., Medicaid Home Health and Community-Based Services, Part D)
- ☐ Population or Geographic Area

1.8a Population or Geographic Area Level of Analysis* *(Please specify) [character limit: 255]*

- ☐ Other

1.8b Other Level of Analysis* *(Please specify) [character limit: 255]*

Reminder

Ensure that accountable entity level testing for maintenance measures is conducted for all levels of analysis specified in the measure.

1.9 Care Setting*

Select the care setting(s) for which the measure is specified and tested. For instrument submissions, select the care setting(s) for which the associated IDMs are specified and tested. Choose all that apply. If "No Applicable Care Setting" or "Other Care Setting", please explain.

- ☐ Ambulatory Care: Clinic
- ☒ Ambulatory Care: Clinician Office
- ☒ Ambulatory Care: Office
- ☒ Ambulatory Surgery Center
- ☐ Behavioral Health: Inpatient (e.g., Inpatient Psychiatric Facility)
- ☐ Behavioral Health: Outpatient
- ☐ Birthing Center
- ☐ Clinician Office/Clinic
- ☐ Emergency Department
- ☐ Emergency Medical Services/Ambulance
- ☒ Home Health
- ☐ Hospice
- ☐ Hospital: Acute Care Facility
- ☐ Hospital: Critical Access
- ☒ Hospital: Inpatient
- ☒ Hospital: Outpatient
- ☐ Imaging Facility
- ☒ Inpatient Rehabilitation Facility
- ☐ Long-Term Acute Care Facility
- ☐ Nursing Home/Skilled Nursing Facility
- ☐ Outpatient Rehabilitation
- ☐ Pharmacy
- ☒ Urgent Care: Ambulatory
- ☐ No Applicable Care Setting

1.9a Rationale for No Applicable Care Setting* *(Please explain) [character limit: 255]*

☐ Other Care Setting

1.9b Other Care Setting* *(Please specify) [character limit: 255]*

*[**Note:** Responses to items 1.10 – 1.13 and other measure specification details are to be provided in the Full Measure Submission]*

1.14 Numerator*

Provide the numerator, i.e., the measure focus. Do not include the measure rationale.

The total risk-adjusted cost to Medicare for all services and items related to the elective primary hip arthroplasty episodes for the patients in the denominator, including pre-operative, operative, and post-operative care within the defined episode window (30 days before through 90 days after the procedure).

Quick Tip

Clearly state the measure focus and relevant timeframes. This measure's focus is capturing the costs of any clinically-related services incurred from 30 days prior to the procedure through 90 days after and assessing those costs against the national average to determine high or low cost performance.

1.15 Denominator*

Provide the denominator, i.e., the target population. Do not include exclusions here; there is a separate field in the full measure submission form to describe exclusions.

All Medicare beneficiaries enrolled in Medicare Parts A and B who receive an elective primary hip arthroplasty during the performance period.

Quick Tip

The denominator of a cost measure is the target population in which the costs of services associated with the specific procedure, care process, etc.

1.15d Age Group*

Select the age group(s) that are reflected in your measure's target population (choose all that apply). Choose an age group only if the entire range is included in your measure's target population. If only part of one or more listed age ranges applies, select "Other" and enter the correct age range (e.g., 14 - 50).

☐ Children (0-17 years)

☒ Adults (18-64 years)

☒ Older Adults (65 years and older)

☐ Other *(1.15e Provide age range in years*) [character limit: 255]*

Quick Tip

Remember to select all age ranges that apply to the measure population. Here, the developer selected both Adults (18-64 years) and Older Adults (65 years and older) as the measure population is all adults 18 years and older.

6.1 Use

6.1.1. Current Status*

Is this new or maintenance measure currently in use in at least one accountability application? The application can be for confidential reporting, reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, performance-based payment, network inclusion/exclusion). For instrument submissions, what is the status for MOST associated IDMs (you will have the option of customizing this information for IDMs, as needed)?

☐ Not in use ☒ In use

6.1.2 Current or Planned Use(s)*

Choose all that apply. For instrument submissions, what is the current use for MOST associated IDMs (you will have the option of customizing this information for IDMs, as needed)?

☐ Public Reporting
☐ Public Health/Disease Surveillance
☒ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☐ Other

6.1.2a Other Current Use* *(Please specify)*

6.1.3 *[If 6.1.1 Current Status = In Use]* **Program Details***

Please provide the following information describing the program(s) in which the measure is currently used. For instrument submissions, describe the programs in which the associated IDMs are currently used.

Name of the program and sponsor

Quality Payment Program Merit-based Incentive Payment System; Centers for Medicare & Medicaid Services

URL of the program

This must be an external URL such as https://example.com.

https://qpp.cms.gov/mips/overview

Reminder

Maintenance measures must be currently in use in at least one accountability application or have a short-term plan (i.e., within 1 year) for such use.

Purpose of the program

The Medicare Access and Children's Health Insurance Program (CHIP) Reauthorization Act of 2015 (MACRA) established the Quality Payment Program (QPP). Under the QPP, clinicians are incentivized to provide high-quality and high-value care through Advanced Alternate Payment Models (Advanced APMs) or the Merit-based Incentive Payment System (MIPS). MIPS-eligible clinicians will receive a performance-based payment adjustment to their Medicare payment. This payment adjustment is based on a MIPS final score that assesses evidence-based and practice-specific data across the following categories:

1. Quality
2. Improvement activities
3. Promoting interoperability
4. Cost

Geographic area and percentage of accountable entities and patients included

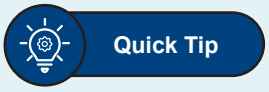
United States

The number of clinicians participating in the QPP) under MIPS has varied slightly across performance years. For context, in 2022, 935,383 clinicians were eligible, and 99.76% participated, with 519,334 participating individually or in groups and 416,049 through APMs. [1] Most recently, in 2023, 931,716 clinicians were eligible, and 99.76% participated, with 517,034 participating as individuals or groups and 414,682 through APMs. [2] As clinicians have choices on how to participate in the QPP (e.g., through MIPS or the Advanced APMs, as groups or individuals), the exact number and percentage of clinicians who received a performance score on this measure was confirmed after the end of the performance period.

References:

1. CMS, "2022 QPP Participation Results Infographic," Quality Payment Program, https://qpp.cms.gov/resources/document/2d0ac6dd-79ae-4e52-9587-e9a7c5ec0443

2. CMS, "2023 QPP Participation Results Infographic," Quality Payment Program, https://qpp.cms.gov/resources/performance-data



If available, include references to support evidence provided.

Applicable level of analysis and care setting

Level(s) of Analysis: Clinician: Group/Practice, Clinician: Individual

Care Setting(s): Hospital: Outpatient, Hospital: Inpatient, Ambulatory Surgery Centers, Ambulatory/Office-Based Care Setting

Attestations: Preparing for Full Measure Submission for Endorsement Consideration

Check the boxes to attest this information will be available and submitted to Battelle by the Full Measure Submission (FMS) deadline of the intended review cycle. The measure may be insufficient for endorsement review if this information is not available by the FMS deadline. Please review the PQM E&M Rubric [Endorsement and Maintenance (E&M) Guidebook] for full measure submission evaluation criteria.

Reminder

If there are questions about what is required for your measure for endorsement review, please reach out to PQMSupport@battelle.org prior to the Full Measure Submission deadline.

☒ **A.1 Detailed Measure Specifications***
I will provide detailed measure specifications, including how to calculate the measure, data dictionaries, and code sets.

☒ **A.2 Logic Model***
I will provide a logic model and evidence that support the link between structures/processes/intermediate outcomes and the desired outcome.
See the E&M Logic Model Guidance for additional details.

☒ **A.3 Impact and Gap***

- For initial endorsement, I will provide a description of the measure's anticipated impact on important outcomes supported by the scientific literature and other sources (e.g., functional improvement, disease prevented, adverse events or costs avoided). I will explain why existing measures/ quality improvement programs are insufficient for addressing this health care need. I will also provide evidence the target population (e.g., patients) values the measured outcome, process, or structure, and finds it meaningful.
- For maintenance endorsement, in addition to the above, I will supply evidence of a continued performance or measurement gap by providing performance scores on the measure as specified (current and over time) at the specified level of analysis. I will ensure that the data used for testing (e.g., performance gap, trend analyses, stratification) includes any year(s) within the past 5 years.

☒ **A.4 Feasibility Assessment Methodology and Results***
I will provide feasibility assessment methodology and results. I will show how the assessment considered the people, tools, tasks, and technologies necessary to implement the measure, and if an eCQM, I will provide the completed feasibility scorecard.

- For initial endorsement, the feasibility assessment will include a description of the extent to which the required data elements are routinely generated and used during care delivery, are available in electronic health records or other electronic sources, and have a data collection strategy that can be implemented.

- For maintenance endorsement, the feasibility assessment will include whether changes to the measure specifications have impacted what is listed for initial endorsement and the extent of any measure implementation challenge(s)/barrier(s) that occurred because of the data elements and any mitigation strategies.

A.5 Measure Testing (reliability and validity)

Check the boxes to attest to which testing (person/encounter-level or accountable entity-level) for reliability and validity will be available and submitted for each level of analysis by the FMS deadline of the intended review cycle. ***Note:*** *For initial endorsement, you must provide a rationale if empirical person or encounter-level testing will not be presented in the FMS. For maintenance endorsement of single measures and IDMs, you must provide a rationale if measured/accountable entity testing will not be presented in the FMS.* ***Data used for reliability and validity testing must include any year(s) within the past 5 years.***

A.5a Empirical Person or Encounter Level*

Will empirical person- or encounter-level evidence, testing, methodology, and results be presented for this endorsement? ***Note:*** *For initial endorsement, person- or encounter-level empirical testing is required or existing evidence (e.g., prior research, literature) must be presented to support testing of all critical data elements (numerator, denominator, exclusions).*

☒ No ☐ Yes

Quick Tip

For initial endorsement, person- or encounter-level empirical testing is required, or existing evidence (e.g., prior research, literature) must be presented to support testing of all critical data elements (numerator, denominator, exclusions).

Because this is a maintenance measure, accountable entity-level empirical testing is required and the developer selects "yes" in question A.5b below.

Reminder

This measure example is being submitted for maintenance, so person- or encounter-level testing is not required. However, if the measure were being submitted for initial endorsement, person- or encounter-level testing would be necessary.

Because this measure uses standardized prices generated through an algorithmic process, reliability is generally assumed. However, it is important to provide evidence demonstrating the accuracy and consistency of the underlying data used in the standardization algorithm. This should include documentation or citations that clearly describe the algorithm, explain its inputs (such as claims data and service codes), and provide justification for any updates made to the algorithm over time.

Additionally, evidence should confirm that the data sources used are accurate, complete, and reliable. This may include references to data audit procedures, third-party validations, previous testing results, or relevant peer-reviewed publications and technical reports that support the reliability of the standardized pricing methodology.

[If 1.0 = New & A5a = No] **A.5a[1] Why Not Presented***

Provide a rationale for why empirical person- or encounter-level testing for reliability and validity will not be presented for this initial endorsement. [character limit: 255]

A.5b Empirical Accountable Entity Level*

Will empirical accountable entity-level evidence, testing, methodology, and results be presented for this endorsement? ***Note:*** *For maintenance endorsement, accountable entity-level empirical testing is required.*

☐ No ☒ Yes

[If 1.0 = Maintenance & A5a = No] **A.5b[1] Why Not Presented***

Provide a rationale for why empirical accountable entity-level testing will not be presented for this maintenance endorsement. [character limit: 255]

[If 1.0 = New] **A.5c Systematic Assessment of Face Validity of Performance Measure Score***

Will systematic assessment of face validity of performance measure score (i.e., accountable entity-level) as an indicator of quality or cost/resource use (i.e., the score is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance) be presented for this initial endorsement?

☐ No ☒ Yes

☒ **A.6 Address Closing Care Gaps (optional for Spring and Fall 2026)***

I will describe how this measure can distinguish differences in care for certain patient subpopulations, which can be used to close gaps in care across those identified subpopulations. *See the E&M Closing Care Gaps Guidance for additional details.*

☒ **A.7 Measure's Use or Intended Use***

I will provide the measure's use or intended use and actions measured entities must take to improve performance on this measure. For a maintenance measure, I will provide a summary of any progress improvement.

A.8 Risk Adjustment or Stratification*

Choose the correct option to attest to whether the measure is risk-adjusted and/or stratified, and to attest that each component of the respective information will be available and submitted by the FMS deadline of the intended review cycle, as applicable. If this information will be unavailable by full measure submission, your measure may not be sufficient for endorsement review.

1 For patient- or encounter-level testing, prior evidence of reliability and validity of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data elements. Prior evidence could include published or unpublished testing that: includes the same data elements, uses the same data type (e.g., claims, chart abstraction), and is conducted on a sample as described above (i.e., representative, adequate numbers, and randomly selected, if possible).

☐ No, neither risk-adjusted nor stratified

☒ Yes, risk-adjusted only

☒ **Conceptual model for risk adjustment**

I will present the conceptual model for risk adjustment, including supporting evidence from literature, internal analyses, and/or expert panels, AND

☐ **Risk adjustment approach**

I will present the risk adjustment approach, including the methodology, specifications, results, and interpretation of results

☐ Yes, stratified only

☐ **All information required to stratify the measure results**

I will present all information required to stratify the measure results, including the stratification variables, definitions, specific data collection items/responses, and code/value sets

☐ Yes, both risk-adjusted and stratified

☐ **Conceptual model for risk adjustment**

I will present the conceptual model for risk adjustment, including supporting evidence from literature, internal analyses, and/or expert panels, AND

☐ **Risk adjustment approach**

I will present the risk adjustment approach, including the methodology, specifications, results and interpretation of results, AND

☐ **All information required to stratify the measure results**

I will present all information required to stratify the measure results, including the stratification variables, definitions, specific data collection items/responses, and code/value sets, and the risk-model covariates and coefficients for the adjusted version of the measure

A.9 Quality Measure Developer and Steward Agreement (QMDSA) Form*

The QMDSA and Additional and Maintenance Measures Forms are contractual agreements that must be signed by Battelle Memorial Institute (Battelle) and any measure steward that is submitting one or more measures to be evaluated for endorsement via the consensus endorsement process. If the measure is not owned by a government entity, the measure steward will also complete and submit a QMDSA Form. For more information about QMDSA requirements, please see the QMDSA Submission Instructions. Choose one.

☐ I already submitted a QMDSA Form to Battelle

A.9a QMDSA Submitted Date

Provide the date submitted.

☐ I would like to submit the QMDSA form now

A.9b *Attach form; One file only; 256 MB limit; Allowed types: pdf.*

[Please use the file naming convention QMDSA-[steward]-DD.MM.YYYY]

☒ The measure is owned by a government entity; therefore, the QMSDA form is not applicable at this time.

Quick Tip

As the measure steward is the Centers for Medicare & Medicaid Services (a government entity), a QMDSA Form is not applicable.

A.10 Additional and Maintenance Measures Form*

[Choose one.] Note: Measure stewards with current measures endorsed by Battelle, who wish to add additional measures to their current QMDSA, will need to complete this form.

☐ I have submitted or will submit an Additional and Maintenance Measures Form.

[Please use the file naming convention Additional-Maintenance-Measures-[steward]-DD.MM.YYYY]

☒ The Additional and Maintenance Measures Form is not applicable at this time.

☒ **A.11 508 Compliance***

I will ensure that the measure information that will be submitted at FMS, including all attachments, will be prepared in accordance with Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d), as amended by the Workforce Investment Act of 1998 and the Architectural and Transportation Barriers Compliance Board Electronic and Information (EIT) Accessibility Standards (36 CFR part 1194). *See Appendix E in the Endorsement and Maintenance (E&M) Guidebook*.

Reminder

Appendix E in the E&M Guidebook includes guidance for making submissions 508 compliant.

Measure Points of Contact Information

[Fields are limited to 128 characters except steward organization copyright (unlimited)]

Steward Organization: Centers for Medicare & Medicaid Services

Choose from the drop-down menu. If your organization does not appear on the list, select "Other" and enter the name of the organization in the box provided.

Steward organization URL: https://www.cms.gov/

Steward POC email: sampleuser@domain.com

Steward POC phone number: 555-123-4567
Country: United Sates
First Name: Jane
Last Name: Doe
City, State: Windsor Mill, Maryland

Steward Organization Copyright*

Please disclose any Intellectual Property, Patents, Copyrights and/or License Restrictions for the measure and its use. Measure Stewards are solely responsible for ensuring that data are marked properly.

As per the Quality Measure Developer and Steward Agreement form, the Measure Steward agrees to indemnify, defend, and hold Battelle, its affiliates, and their respective directors, officers, employees, consultants, and agents harmless from any and all liabilities, demands, damages, costs, and expenses (including reasonable attorney fees and court costs) arising from third party suits or claims resulting from liability and/or Steward's use, inability to use, or misuse of any deliverable, data, item, or other information delivered by Battelle.

If there is no disclosure to make, enter "Not applicable."

Not Applicable

☒ **The measure developer is different from the measure steward**

Developer POC email: sampleuser@domain.com
Developer POC phone number: 555-123-4567
Country: United States
First Name: John
Last Name: Doe
Organization: Acumen, LLC
City, State: Burlingame, California

Endorsement & Maintenance (E&M) Full Measure Submission – Single Measure

Instructions: This form can be used as a worksheet to assist you in developing your **Full Measure Submission (FMS)** for a new or maintenance measure.

- Use this template ONLY to prepare an FMS for a **single measure** submission; a single measure is one that focuses on a single structure, process, intermediate outcome, or outcome, or cost/resource use measure (includes **composite measure** submissions)

For an **instrument-derived measure set**, which is **two or more** measures derived from a single instrument or survey tool (e.g., Consumer Assessment of Healthcare Providers and Systems (CAHPS):

- **Before beginning an instrument or IDM submission**, please carefully review Overview of the Instrument-Derived Measure Set Submission Framework for the submission workflow and a summary of the information requirements for instrument vs. IDM submissions
- Download the **FMS template for an instrument** submission
- Download the **FMS template for an instrument-derived measure (IDM)** submission

The FMS online submission tool is available as soon as your ITS submission is reviewed and approved for FMS. When you have received the approval for full measure submission, navigate to the PQM website and log into your PQM account. Once logged in, click "My Account" to go to your dashboard, then scroll down to see your measures or use the search bar to search by measure title or CBE ID. Measures that have been approved for FMS will appear with an "Endorsement Cycle Status" of *Full Measure Submission Start*; measures with an FMS submission in progress will have the status *Full Measure Submission Draft*. Click "Edit" to the left of the CBE ID to begin or continue an FMS submission. Click on the measure title to see how the measure will be displayed once it is published. For more information on the measure submission process, visit the E&M measure submission webpage.

- You may **save** a draft of the FMS form before completing all required fields
- You must complete all required fields (denoted by *) to **submit** the final FMS
  - To **avoid losing data**, always save the draft before submitting
  - All **errors must be resolved** for submission to be successful
- If you would like to make changes to information submitted via the Intent to Submit (ITS), you may **edit the ITS content** at the top of the FMS form
- Ensure all attachments are **508 compliant**, including labeling all tables and figures with alternative text, as appropriate
- **NOTE:** The PQM account used to start a measure submission is the one that must be used to continue, revise, or complete the submission

Required fields vary depending on whether your measure is an electronic Clinical Quality Measure (eCQM), or an initial (new) measure versus a maintenance measure, or for other selected situations. Conditional fields are indicated in this template with

brackets before each fieldname (e.g., *[If new measure]* **Potential Unintended Consequences***). *Note that the CBE ID for a maintenance measure submission appends "-m" (e.g., CBE ID 1234-m) to differentiate between the published version of the measure and the record being submitted for maintenance review.*

Section 1. Measure Specifications

[NOTE: Items 1.0-1.9, 1.14, 1.15, 1.15d, and 1.15e were entered in the ITS, and can be edited at the top of the FMS form]

1.10 Measure Rationale*

Provide a rationale for why measured entities should report this measure, including how the measure will improve the quality of care for patients and/or any associated health care costs, and the benefits or improvements in quality envisioned by use of this measure. For an instrument, the rationale should describe how the data gathered using the instrument can be used to improve patient outcomes, experience, or decision making.

Estimates of total hip arthroplasties in the United States indicate higher prevalence among Medicare-age patients, with prevalence at 0.8 percent for the general population but increasing with age to 1.5 percent at sixty years and 5.9 percent by ninety years of age. [1] Furthermore, demand for hip arthroplasties among U.S. Medicare patients is projected to increase by 176% between 2019 and 2040, rising from 262,369 procedures in 2019 to an estimated 828,286 procedures by 2040. [2] There are currently substantial opportunities to improve the cost efficiency and quality of care related to hip arthroplasties, given the high variation among relevant treatment options. These include the appropriate use of institutional post-acute care (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 (also referred to in this form as the "Hip Arthroplasty" measure) was recommended for development by an expert clinician committee (the Musculoskeletal Disease Management - Non-Spine Clinical Subcommittee, composed of 29 experts affiliated with 26 specialty societies) because of its impact in terms of patient population and clinician coverage, as well as the opportunity for incentivizing cost-effective, high-quality clinical care in this area. Based on the initial recommendations from the Clinical Subcommittee, a subsequent Hip Arthroplasty clinician expert workgroup, composed of 15 members affiliated with 14 specialty societies (see the Supplemental Information in Section 7 of this form) provided extensive, detailed input on this measure. Workgroup input has helped ensure the measure's ability

Quick Tip

The rationale explains the expected benefits or improvements in quality and cost-efficiency envisioned by the measure, including how it aims to identify and reduce unnecessary variation in resource use, promote adherence to evidence-based practices, and generate potential healthcare cost savings while maintaining or improving patient outcomes.

The envisioned benefits of the Elective Primary Hip Arthroplasty Measure include efficient use of healthcare resources, improved quality of care, and lower overall costs for Medicare patients undergoing hip replacement surgery. By focusing on reducing unnecessary variation in treatment and encouraging adherence to best practices such as optimal surgical techniques and appropriate post-acute care. The measure is developed with extensive input from clinical experts to ensure fair evaluation of clinician performance and to promote high-value, cost-effective care in this high-impact area.

Quick Tip

Ensure to disclose any expert panel or workgroup convened to advise on the development of the measure and include those details in the Supplemental Attachment Section 7.

to fairly evaluate clinician cost performance for elective primary hip arthroplasty surgeries and to promote efficient and high-quality care for Medicare patients undergoing these procedures.

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. [3]

References:

1. Hilal Maradit Kremers et al., "Prevalence of Total Hip and Knee Replacement in the United States," Journal of Bone and Joint Surgery 97, no.1 (2015): 1386-97.

2. Ittai Shichman et al., "Projections and Epidemiology of Primary Hip and Knee Arthroplasty in Medicare Patients to 2040-2060." JBJS Open Access 2023:e22.00112. http://dx.doi.org/10.2106/JBJS.OA.22.00112.

3. CMS, "Measure Development Process," (October 2018), 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

1.11 Measure Webpage*

Provide a URL to a webpage, specific for this measure or instrument derived measure set, containing current detailed specifications, including code lists, risk model details, and supplemental materials. Do not enter a URL to a home page or to general information. The webpage must be publicly accessible. If no URL is available, copy and paste this example: http://example.com. *[character limit: 255]*

On the QPP Resource Library https://qpp.cms.gov/resources/resource-library, refer to "Links to 2024 MIPS Performance Category Measure Specifications, Activity Inventory, and Supporting Documentation." Then fine the zip file link, "• 2024 MIPS Cost Measure Information Forms" and find the pdf file, "2024-12-py2024-mif-ebcm-el-ha."

1.13 Data Dictionary*

☐ Not attached. I attest that all information will be provided where codes and/or value sets are needed (1.14a – 1.15c).
☒ Attached

For instruments, note that fields 1.14a - 1.15c are collected in the IDM form.

1.13a *[If 1.13 = Attached]* **Attach Data Dictionary***

Attach a data dictionary, code table, and/or value sets (include variables in the final risk model or stratification plan, if applicable). Attachment should include variables used in the final risk

model and/or stratification, if applicable (please clearly label sheets).

One file only; 256 MB limit; allowed file types: .xls; .xlsx; .csv; .pdf; .zip

The Research Data Assistance Center (ResDAC) maintains the Medicare claims data dictionary available here: https://resdac.org/cms-data/files/ip-ffs/data-documentation.

Code Table: Attachment 2026-01-09-codes-list-el-ha.xlsx

Quick Tip

The provided code table includes clearly defined data elements and uses consistent terminology that aligns with industry standards such as ICD-10 and CPT/HCPCS. Codes and variables are organized by data elements such as episode triggers, clinician attribution, service assignment, exclusions, and risk adjustment across clearly labeled tabs. There is contextual information within the overview section to guide users in applying the data appropriately.

1.14a Numerator Details*

Provide details needed to calculate the numerator. All information required to identify and calculate the cases from the target population (denominator) with the target process, condition, event, or outcome, such as definitions, time period for data collection, specific data collection items/responses, code/value sets. If the numerator includes a list (or lists) of individual codes with descriptors that exceeds one page, please provide this information as part of the data dictionary attachment (see 1.13).

The Elective Primary Hip Arthroplasty measure is the risk-adjusted cost across all episodes attributed to the clinician group (identified by Taxpayer Identification Number, or TIN) or individual clinician (identified by unique combination of Taxpayer Identification Number and National Provider Identifier, or TIN-NPI). Elective Primary Hip Arthroplasty episodes, which are units or specific instances of the measure for a given patient and clinician or clinician group, are triggered or opened by Current Procedural Terminology / Healthcare Common Procedure Coding System (CPT/HCPCS) codes indicating the presence of a hip arthroplasty procedure.

The episode window spans 30 days prior to the trigger day through 90 days after, and includes costs from certain clinically related services from Medicare Parts A and B claims during the episode window. [1] Cost figures are standardized to account for differences in Medicare payments for the same service(s) across Medicare providers. Payment-standardized costs remove the effect of differences in Medicare payment among health care providers that are the result of differences in regional health care provider expenses measured by

Quick Tip

The numerator is the primary focus of the measure. Clearly describe details that are needed in order to calculate the numerator.

In this submission, the developer defines episodes of the Elective Primary Hip Arthroplasty measure as units of care triggered by specific CPT/HCPCS procedure codes indicating a hip arthroplasty.

Additionally, the developer explains how numerator events (Elective Primary Hip Arthroplasty episodes) are identified and specifies the episode window, which spans from 30 days prior to the trigger day through 90 days after, and includes costs from certain clinically-related services from Medicare Parts A and B claims.

Quick Tip

Detail the costing method including the source of cost information, steps to capture costs, and provide a rationale for this methodology.

hospital wage indexes and geographic price cost indexes (GPCIs) or other payment adjustments such as those for teaching hospitals. This standardization is intended to isolate cost differences that result from health care delivery choices, allowing for more accurate resource use comparisons between health care providers. [2] A regression model is applied to estimate the expected cost of each episode for risk adjustment. The cost measure is calculated as the average of the ratios of observed to expected payment-standardized costs to Medicare for all Elective Primary Hip Arthroplasty episodes attributed to a clinician or clinician group. This resulting average episode cost ratio is then multiplied by the national average observed episode cost to generate a dollar figure. We use the average of ratios approach instead of the ratio of averages because this method ensures each episode carries equal weight regardless of its cost and prevents provider performance from being disproportionately influenced by episodes with particularly large observed or expected values.

References:

1. Cost is defined by allowed amounts on Medicare claims data, which include both Medicare trust fund payments and any applicable beneficiary deductible and coinsurance amounts. Claims data from Medicare Parts A and B are used to construct the episode-based cost measures.

2. For more information on payment-standardized costs, please refer to the "CMS Price (Payment) Standardization - Basics" and "CMS Price (Payment) Standardization - Detailed Methods" documents posted on the CMS Price (Payment) Standardization Overview page (https://resdac.org/articles/cms-payment-standardization-overview).

Quick Tip

Provide a rationale for the chosen calculation method for the measure score.

1.15a Denominator Details*

Provide details needed to calculate the denominator, including all information required to identify and calculate the target population/ denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets. If the list(s) of individual codes with descriptors exceeds one page, please provide this information as part of the data dictionary attachment (see 1.13).

The denominator for the Elective Primary Hip Arthroplasty episode-based cost measure includes all episodes of care for Medicare beneficiaries who are enrolled in Medicare Parts A and B and who receive an elective primary hip arthroplasty during the annual performance period.

Each episode is defined as all clinically-related services and items provided to the patient from 30 days prior to the hip arthroplasty procedure (which serves as the trigger event) through 90 days after the procedure.

Episodes are attributed to the clinician responsible for managing the patient's care. Data collection requires verifying patient eligibility,

procedure performance using relevant CPT/HCPCS and ICD-10 codes (see data dictionary in section 1.13), attribution to the managing clinician, and capturing service dates within the specified episode window.

1.15b Denominator Exclusions*

Briefly describe exclusions from the denominator cases, if any. Enter "None" if the measure does not have denominator exclusions.

The following are the exclusions specific to this measure, which were determined based on consideration of the clinical characteristics of a homogenous patient cohort, and these measure-specific exclusions are applied in conjunction with a list of standard exclusions, which can be found in section 1.15c – Denominator Exclusion Details. For further details on the codes and logic used to define each measure-specific exclusion, please refer to the "Exclusions_Details" tab of the codes list attachment in section 1.13.

-Bilateral Hip Arthroplasty, Primary and Staged

-Congenital Deformity of the Hip

-Hip Arthroplasty for Cancer

-Hip Fracture/Trauma (Reason for Hip Arthroplasty)

-Osteomyelitis of Hip and Femur

-Septic Joint

-IP Procedures without Relevant MS-DRGs

1.15c Denominator Exclusions Details*

Provide details needed to calculate denominator exclusions. Enter "None" if the measure does not have denominator exclusions. Provide all information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets. If the list(s) of codes with descriptors exceeds one page, please provide this information as part of the data dictionary attachment (see 1.13).

Before measure calculation, episode exclusions are applied to remove certain episodes from measure score calculation. Certain exclusions are applied across all procedural episode groups, and other exclusions are specific to the Elective Primary Hip Arthroplasty measure, based on consideration of the clinical characteristics of a homogenous patient cohort.

The steps for episode exclusion are as follows:

- Exclude episodes from measure calculation if:
  - The patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the trigger day.
  - The patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.
  - No main clinician is attributed the episode.
  - The patient's date of birth is missing.
  - The patient's death date occurred before the episode ended.
  - The episode trigger claim was not performed in an ambulatory/office-based care, IP hospital, OP

hospital, or ASC setting based on its place of service.

- The IP facility is not a short-term stay acute hospital as defined by subsection (d) when an IP stay concurrent with the trigger is found. [1]

• Apply measure-specific exclusions, which check the patient's Medicare claims history for certain billing codes (as specified in the Measure Codes List file) that indicate the presence of a particular procedure, condition, or characteristic.

Reference:

1. 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 CMS Certification Number (CCN) definitions for Short-term (General and Specialty) Hospitals facility types in Section 2779A1 of Chapter 2 of the CMS State Operation Manual. (https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/som107c02.pdf).

1.16 Type of Score*

Select the most relevant type of score. For instrument submissions, select the type of score for MOST associated IDMs (you will have the option of customizing this information for IDMs, as needed).

☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g., average
☐ Count
☐ Rate/proportion
☒ Ratio
☐ Composite scale
☒ Other

1.16a Other scoring method* *(Please specify) [character limit: 255]*

Ratio multiplied by the national average cost

1.17 *[If Measure Type (1.7) IS "Cost/Resource Use"]* **Type of Cost Measure***

Choose one.

☐ Per capita (population– or patient–based)
☒ Per episode
☐ Per procedure
☐ Other

1.17a Other Cost Measure Type* *(Please specify) [character limit: 255]*

1.18 Calculation of Measure Score*

Diagram or describe the calculation of the measure score as an ordered sequence of steps. Identify the denominator, denominator exclusions (if any), numerator, time period of data collection, risk adjustment and/or stratification, and any other calculations. For instrument submissions, describe the calculation of

measure score for MOST associated IDMs (you will have the option of customizing this information for IDMs, as needed).

Step 1. Trigger and Define an Episode

Elective Primary Hip Arthroplasty episodes are defined by CPT/HCPCS codes on Part B Physician/Supplier (Carrier) claims that open, or trigger, an episode. The steps for defining an episode for the Elective Primary Hip Arthroplasty episode group are as follows:

- Identify Part B Physician/Supplier claim lines with positive standardized payment that have a trigger code.
- Trigger an episode if all the following conditions are met for an identified Part B Physician/Supplier claim line:
  - It was billed by a clinician of a specialty that is eligible for the Merit-based Incentive Payment System (MIPS).
  - It does not have a post-operative modifier code. [1]
  - It is the highest cost claim line across all claim lines identified in the above bullets and that have any Elective Primary Hip Arthroplasty trigger code billed for the patient on that day. If multiple Part B Physician/Supplier claim lines with a trigger code occur on different days within a concurrent inpatient (IP) stay, an episode will be triggered by the claim line with the earliest expense date during the IP stay.
- Identify episodes that have a concurrent IP stay by identifying the first IP stay with a relevant Medicare Severity Diagnosis-Related Group (MS-DRG) code for the patient that is concurrent to the expense date for the trigger Part B Physician/Supplier claim line.
- Establish the episode window as follows:
  - Establish the episode trigger date as the IP start day if an IP stay with a relevant MS-DRG concurrent with the trigger is found, otherwise use the expense date of the trigger code.
  - Establish the episode start date as 30 days prior to the episode trigger date.
  - Establish the episode end date as 90 days after the episode trigger date.

Step 2. Attribute Episodes to a Clinician

Once an episode has been triggered and defined, it is attributed to one or more clinicians of a specialty that is eligible for MIPS. Clinicians are identified by TIN-NPI, and clinician groups are identified by TIN. Only clinicians of a specialty that is eligible for MIPS or clinician groups where the triggering clinician is of a specialty that is eligible for MIPS are attributed episodes.

The steps for attributing an Elective Primary Hip Arthroplasty episode are as follows:

- Identify claim lines with positive standardized payment for any trigger codes that occur during the IP stay; if the triggering procedure occurs during an IP stay with a relevant MS-DRG, otherwise identify claim lines with positive standardized payment for any trigger codes that occur on the trigger day.
- Designate a TIN-NPI as a main clinician if the following conditions are met:
  - No assistant modifier code is found on one or more claim lines billed by the clinician.
  - No exclusion modifier code is found on the same claim line.
- Designate a TIN-NPI as an assistant clinician if the following conditions are met:
  - The TIN-NPI was not designated as a main clinician.
  - An assistant modifier code is found.

- No exclusion modifier code is found.

- Attribute an episode to any TIN-NPI designated as a main or assistant clinician.
- Attribute episodes to the TIN by aggregating all episodes attributed to NPIs that bill to that TIN. If the same episode is attributed to more than one NPI within a TIN, the episode is attributed only once to that TIN.

Step 3. Assign Costs to an Episode and Calculate Total Observed Episode Cost

Services, and their Medicare costs, are assigned to an episode only when clinically related to the attributed clinician's role in managing patient care during the episode. Assigned services may include treatment and diagnostic services, ancillary items, services directly related to treatment, and those furnished as a consequence of care (e.g., complications, readmissions, unplanned care, and emergency department visits). Unrelated services are not assigned to the episode. For example, the cost of care for a chronic condition that occurs during the episode but is not related to the clinical management of the patient relative to the elective primary hip arthroplasty would not be assigned. For the Elective Primary Hip Arthroplasty episode group, only services performed in the following service categories are considered for assignment to the episode costs:

- Emergency Department (ED)
- Outpatient (OP) Facility and Clinician Services
- IP - Medical
- IP - Surgical
- Inpatient Rehabilitation Facility (IRF) - Medical
- Durable Medical Equipment, Prosthetics, Orthotics, and Supplies (DME)
- Home Health (HH)

In addition to service category, service assignment rules may be modified based on the service category in which the service is performed, as listed above. Service assignment rules may also be defined based on specific (i) service information alone or service information combined with diagnosis information, (ii) prior incidence of service, and/or (iii) the timing of the service, as detailed below.

- Services may be assigned to the episode based on the following service information combinations:
  - High level service code alone
  - High level service code combined with first 3 digits of the International Classification of Diseases – 10th Revision diagnosis code (3-digit ICD-10 diagnosis code)
  - High level service code combined with full ICD-10 diagnosis code
  - High level service code combined with more specific service code
  - High level service code combined with more specific service code and with 3-digit ICD-10 diagnosis code
  - High level service code combined with more specific service code and with full ICD-10 diagnosis code

- Assigned services may be further refined by prior incidence of service or diagnosis:
  - Services may be assigned unconditionally (regardless of prior incidence of the service in patient's recent claims history).
  - Services may be assigned if newly occurring.

  - Services may be assigned in combination with a diagnosis if the service is newly occurring.
  - Services may be assigned in combination with a diagnosis if the diagnosis is newly occurring.
  - Services may be assigned in combination with a diagnosis if either the service OR the diagnosis are newly occurring.
  - Services may be assigned in combination with a diagnosis if both the service AND the diagnosis are newly occurring.

- Services as defined by the applicable combinations and incidence options above may be assigned with only specific timing:
  - Services may be assigned based on whether or not the service occurs before the trigger (in the pre-trigger window) and/or after the trigger (in the post-trigger window).
  - Services may be assigned only if they occur within a particular number of days from the trigger within the episode window, and services may be assigned for a period shorter than the full duration of the episode window.

The full list of service assignment rules for the Elective Primary Hip Arthroplasty measure can be found on the "Service_Assignment" tab of the Elective Primary Hip Arthroplasty Measure Codes List file (referenced in section 1.13). The steps for assigning costs are as follows:

- Identify all services on claims with positive standardized payment that occur within the episode window.
- Assign identified services to the episode based on the types of service assignment rules described above.
- Assign skilled nursing facility (SNF) claims based on the following criteria:
  - Identify SNF claims for which both (i) the SNF claim's qualifying IP stay is the IP stay during which the trigger occurs, if an IP stay is found, and (ii) the SNF claim occurs during the episode window.
  - For those identified SNF claims, assign the percentage of the claim amount proportional to the portion of the SNF claim that overlaps with the episode window.

- Assign all claims with trigger codes occurring during the trigger day/stay.
- Assign all physician claims and DME claims occurring during concurrent IP stay as applicable.
- Assign all IP evaluation and management (E&M) claims during IP stays in the post-trigger window assigned to episode.
- Sum standardized Medicare allowed amounts for all claims assigned to each episode to obtain the standardized total observed episode cost.

Step 4. Exclude Episodes

Before measure calculation, episode exclusions are applied to remove certain episodes from measure score calculation. Certain exclusions are applied across all procedural episode groups, and other exclusions are specific to the Elective Primary Hip Arthroplasty measure, based on consideration of the clinical characteristics of a homogenous patient cohort. The measure-specific exclusions are listed in the "Exclusions" and "Exclusions_Details" tabs in the Elective Primary Hip Arthroplasty Measure Codes List file (referenced in section 1.13). The steps for episode exclusion are as follows:

- Exclude episodes from measure calculation if:
  - The patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the trigger day.

- The patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.
- No main clinician is attributed the episode.
- The patient's date of birth is missing.
- The patient's death date occurred before the episode ended.
- The episode trigger claim was not performed in an ambulatory/office-based care, IP hospital, OP hospital, or ASC setting based on its place of service.
- The IP facility is not a short-term stay acute hospital as defined by subsection (d) when an IP stay concurrent with the trigger is found. [2]

- Apply measure-specific exclusions, which check the patient's Medicare claims history for certain billing codes (as specified in the Measure Codes List file referenced in section 1.13) that indicate the presence of a particular procedure, condition, or characteristic.

Step 5. Estimate Expected Costs through Risk Adjustment

Risk adjustment is used to estimate expected episode costs in recognition of the different levels of care patients may require due to comorbidities, disability, age, and other risk factors. The risk-adjustment model includes variables from the CMS Hierarchical Condition Category Version 22 (CMS-HCC V22) 2016 Risk Adjustment Model, [3] as well as other standard risk adjustors (e.g., patient age) and variables for clinical factors that may be outside the attributed clinician´s reasonable influence. A full list of risk-adjustment variables can be found in the "RA" and "RA_Details" tabs of the Elective Primary Hip Arthroplasty Measure Codes List file (referenced in section 1.13). Steps for defining risk-adjustment variables and estimating the risk-adjustment model are as follows:

- Define HCC and episode group-specific risk adjustors using service and diagnosis information found on the patient's Medicare claims history in the 120-day period prior to the episode trigger day for certain billing codes that indicate the presence of a procedure, condition, or characteristic.
- Define other risk adjustors that rely upon Medicare beneficiary enrollment and assessment data as follows:
  - Identify patients who are originally "Disabled without end-stage renal disease (ESRD)" or "Disabled with ESRD" using the original reason for joining Medicare field in the Medicare beneficiary EDB.
  - Identify patients with ESRD if their enrollment indicates ESRD coverage, ESRD dialysis, or kidney transplant in the Medicare beneficiary EDB in the lookback period.
  - Identify patients who have spent at least 90 days in a long-term care institution without having been discharged to the community for 14 days, based on LTC MDS assessment data, during the lookback period.
- Drop risk adjustors that are defined for less than 15 episodes nationally to avoid using very small samples.
- Categorize patients into age ranges using their date of birth information in the Medicare beneficiary EDB. If an age range has a cell count less than 15, collapse this with the next adjacent higher age range category towards the reference category (65-69).
- Include the MS-DRG of the episode's trigger IP stay, if an IP stay is found, as a categorical risk adjustor.
- Run an ordinary least squares (OLS) regression model to estimate the relationship between all the risk-adjustment variables and the dependent variable, the standardized observed episode cost, to obtain the risk-adjusted expected episode cost.

- Winsorize [4] expected costs as follows:
  - Assign the value of the 0.5th percentile to all expected episode costs below the 0.5th percentile.
  - Renormalize [5] values by multiplying each episode´s winsorized expected cost by the average expected cost, and dividing the resultant value by the average winsorized expected cost.
- Exclude [6] episodes with outliers as follows.
  - Calculate each episode's residual as the difference between the renormalized, winsorized.
  - Expected cost computed above and the observed cost.
  - Exclude episodes with residuals below the 1st percentile or above the 99th percentile of the residual distribution.
  - Renormalize the resultant expected cost values by multiplying each episode's winsorized expected costs after excluding outliers by the average standardized observed cost after excluding outliers, and dividing by the average winsorized expected cost after excluding outliers.

Step 6. Calculate Measure Scores

Measure scores are calculated for a TIN or TIN-NPI as follows:

- Calculate the ratio of observed to expected episode cost for each episode attributed to the clinician/ clinician group.
- Calculate the average ratio of observed to expected episode cost across the total number of episodes attributed to the clinician/clinician group.
- Multiply the average ratio of observed to expected episode cost by the national average observed episode cost to generate a dollar figure representing risk-adjusted average episode cost.

References:

1. Post-operative modifier codes indicate that a clinician billing the service was not involved in the main procedure but was involved in the post-operative care for that procedure, and as such, the post-operative clinician would not be responsible for the trigger.

2. 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 CMS Certification Number (CCN) definitions for Short-term (General and Specialty) Hospitals facility types in Section 2779A1 of Chapter 2 of the CMS State Operation Manual. (https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/som107c02.pdf).

3. CMS uses an HCC risk-adjustment model to calculate risk scores. The HCC model ranks diagnoses into categories that represent conditions with similar cost patterns. Higher categories represent higher predicted healthcare costs, resulting in higher risk scores. There are over 9,500 ICD-10-CM codes that map to one or more of the 79 HCC codes included in the CMS-HCC V22 model.

4. Winsorization aims to limit the effects of extreme values on expected costs. Winsorization is a statistical transformation that limits extreme values in data to reduce the effect of possible outliers. Winsorization of the lower end of the distribution (i.e., bottom coding) involves setting extremely low predicted values below a predetermined limit to be equal to that predetermined limit.

5. Renormalization is performed after adjustments are made to the episode's expected cost, such as bottom-coding or residual outlier exclusion. This process multiplies the adjusted values by a scalar ratio to ensure that the resulting average is equal to the average of the original value.

6. This step excludes episodes based on outlier residual values from the calculation and renormalizes the resultant values to maintain a consistent average episode cost level.

1.18a Attach measure score calculation diagram

Attach a measure score calculation diagram, if desired. Note: You will have the option of providing a customized diagram for IDMs, as needed.

One file only; 256 MB limit; allowed file types: .pdf; .jpg; .png.

1.19 Measure Stratification Details*

Provide all information required to stratify the measure results, if necessary. Include the stratification variables, definitions, code/value sets, and, if appropriate, the risk model covariates and coefficients for the clinically adjusted version of the measure. If the list(s) of codes with descriptors exceeds one page, please provide this information as part of the data dictionary attachment. If the measure is not stratified, please state, "The measure is not stratified." If the information is included within the data dictionary attachment, please state, "See data dictionary attachment."

The measure is not stratified.

1.20 Types of Data Sources*

Select the types of data sources for which you have specified AND tested the measure. Choose all that apply. Please indicate whether each data sources is digital or non-digital, where indicated. A digital data source in health care refers to any electronic system or repository that collects, stores, and provides access to health-related data. Digital data sources can include administrative claims or systems, electronically submitted clinical or social needs assessments, electronic health records (EHRs), laboratory systems, prescriptions drug monitoring programs, medical devices and wearables, health information exchanges (HIEs), clinical data registries, and other digital platforms that facilitate the seamless flow of information across health care settings.

☐ Administrative Data
☒ Claims Data
☐ Electronic Health Records
☐ Paper Patient Medical Records
☐ Registries
☐ Standardized Patient Assessments

1.20a *[If Standardized Patient Assessments]* **Format: Standardized Patient Assessments***

☐ Digital ☐ Non-digital

☐ Patient-Reported Data and/or Survey Data *[Answer questions 1.21–1.24]*

1.20b *[If Patient-Reported Data and/or Survey Data]* **Format: Patient-Reported Data and/or Survey Data***

☐ Digital ☐ Non-digital

☐ Other

1.20c Other Data Source* *(Please specify)* *[character limit: 255]*

1.20d *[If Other]* **Format: Other Data Source***

☐ Digital ☐ Non-digital

1.25 Data Source Details*

Describe the specific data source(s), other than or in addition to any patient-reported data and/or survey data collection instrument(s) indicated for the measure or for the IDMs associated with this instrument submission. For example, provide the name of the database, clinical registry, etc. and characterize how the data are collected. Please discuss any data feasibility, reliability, and/or validity challenges and how they have been mitigated.

The Elective Primary Hip Arthroplasty measure uses Medicare Part A and Part B claims data, which is maintained by the Centers for Medicare & Medicaid Services (CMS). Part A and B claims data are used to build episodes of care, calculate episode costs, and construct risk adjustors. Data from the Medicare Enrollment Database (EDB) are used to determine patient-level exclusions and supplemental risk adjustors, specifically: Medicare Parts A, B, and C enrollment; primary payer; disability status; end-stage renal disease (ESRD); patient birth dates; and patient death dates.

The risk-adjustment model also uses information from the Minimum Data Set (MDS) to account for expected differences in payment for services provided to patients in long-term care via a Long Term Care Indicator variable in risk adjustment. For measure testing, data from the United States Census Bureau American Census, United States Census Bureau American Community Survey (ACS), and Common Medicare Enrollment (CME) are used in the analyses evaluating social risk factors in risk adjustment.

Quick Tip

Describe all data sources, including those used for testing, including risk adjustment.

1.26 Minimum Sample Size*

Indicate whether the measure has a minimum sample size to calculate the performance score and provide any instructions needed for obtaining the sample and guidance on minimal sample size. For instrument submissions, describe the minimum sample size to calculate the performance score for MOST associated IDMs (you will have the option of customizing this information for IDMs, as needed).

The Elective Primary Hip Arthroplasty measure will be reported for TINs and TIN-NPIs with 10 or more episodes.

Section 2. Importance

2.1 Attach Logic Model *

Attach a logic model that illustrates the relationship between health care structures and processes and the desired outcomes. Briefly describe steps between these structures and processes and the desired health outcomes by outlining the resources (i.e., structures or inputs) required and the key activities (i.e., processes or actions)

that the accountable entity can plausibly implement. The expected outputs, along with short-term, intermediate, and long-term outcomes, should be clearly articulated. Outcomes should include the measure focus. The relationships depicted in the diagram should be straightforward, ensuring they are easily understood by a general, non-technical audience. Additionally, highlight any assumptions that underpin your model, describe feedback mechanisms, and acknowledge external factors that could influence the success of the measure, ensuring a comprehensive understanding of the measure's logic and its intended effects. For instrument submissions, the logic model should include each of the associated IDMs. See the Logic Model Guidance for additional information.

One file only; 256 MB limit; allowed file types: .pdf; .doc; .docx.

Figure 1: Logic Model for the Timely Follow-Up Measure

Inputs	Activities	Outputs	Outcomes	Impact
• Multidisciplinary care team (surgeons, nurses, physical therapists, care coordinators) • Evidence-based clinical guidelines for hip arthroplasty • Data systems for tracking costs and outcomes • Patient education materials • Resources for care coordination • Financial and administrative support	• Implement standardized, evidence-based care pathways for elective hip arthroplasty • Provide pre-operative patient education and optimization (e.g., managing comorbidities, smoking cessation) • Coordinate discharge planning and post-operative care (e.g., physical therapy, follow-up visits) • Monitor and analyze episode costs and outcomes • Engage in continuous quality improvement (CQI) based on data feedback	• Increased use of standardized care protocols • Enhanced patient engagement and preparedness • Improved coordination of care transitions • Regular reporting and review of cost and outcome data • Identification of areas for improvement	**Short Term:** • Reduced variation in care delivery • Improved patient understanding and engagement • Timely discharge and follow-up care • Early identification of inefficiencies and cost drivers **Intermediate Term:** • Lower rates of complications and readmissions • More efficient use of resources (e.g., appropriate post-acute care utilization) • Decreased episode costs for hip arthroplasty • Improved clinician performance on measure. **Long Term:** • Sustained reduction in total cost of care for elective hip arthroplasty • Improved patient functional outcomes and satisfaction • Widespread adoption of best practices across entities • Enhanced value and efficiency in orthopedic care	• System-wide improvements in the value and affordability of elective orthopedic procedures • Greater sustainability of Medicare and improved access to high-value care for beneficiaries • Reduced financial burden for patients and payers.

Feedback Mechanisms
• Regular review of measure performance data (cost and outcomes) by clinicians and care teams
• Stakeholder feedback (clinicians, patients, payers) to identify challenges and improvement opportunities
• Benchmarking against peer organizations and national standards
• Incorporation of lessons learned from adverse events or outliers
• Continuous quality improvement cycles based on data analysis and feedback
Assumptions
• Accountable entities have access to accurate, timely data on costs and outcomes
• Clinical teams are motivated and equipped to implement evidence-based practices
• Patients are willing and able to participate in pre- and post-operative care
• Risk adjustment methods adequately account for patient differences
• Resources for care coordination and patient education are available.
External Factors
• Changes in Medicare payment policy or regulations
• Availability of post-acute care resources (e.g., rehab facilities)
• Factors affecting patient access and adherence to post-operative treatment plan
• Market dynamics (e.g., device costs, regional practice patterns)
• Public health events (e.g., pandemics)
• Technology advancements impacting care delivery

Summary: The logic model for the Elective Primary Hip Arthroplasty episode-based cost measure outlines how accountable entities can improve performance by investing in multidisciplinary teams, evidence-based protocols, patient education, and care coordination. By implementing standardized care pathways, optimizing patients before surgery, coordinating post-operative care, and continuously monitoring performance data, entities can achieve outputs such as increased protocol adherence and better care transitions. These lead to reduced care variation and complications in the short and intermediate term, ultimately resulting in sustained cost reductions and improved patient outcomes.

Feedback mechanisms like regular data review, stakeholder input, and benchmarking support ongoing improvement, while the model assumes access to accurate data, motivated teams, and patient engagement. External factors such as policy changes and resource availability can also influence success. Overall, the logic model provides a clear, actionable pathway for entities to enhance efficiency, value, and patient outcomes in elective hip arthroplasty care.

See the E&M Logic Model Guidance for definitions and additional information on how to use this optional template for measure logic models.

2.2 Evidence of Measure Importance*

Summarize evidence of the measure's importance from the literature, expert panel, or internal data analyses, linking the structure/process/intermediate outcome to the desired health outcome. Please provide references for supporting evidence. For instrument submissions, this evidence should include aspects of importance that apply across associated IDMs, such as evidence associated with the surveyed population or accountable entity, etc.

The Elective Primary Hip Arthroplasty measure was developed for use in the Merit-based Incentive Payment System (MIPS) under the Quality Payment Program (QPP) to fulfill statutory requirements outlined in Section 1848(r) of the Social Security Act, as amended by MACRA. This measure supports CMS's broader goals of promoting high-value care by evaluating clinician performance across quality, cost, improvement activities, and interoperability. Within this framework, episode-based cost measures are essential for assessing resource use specific to a procedure or condition, enabling clinicians to identify cost drivers and improve care efficiency.

Hip arthroplasty is one of the most frequently performed orthopedic procedures in the U.S., with over 3.1 million procedures recorded in the American Joint Replacement Registry (AJRR) between 2012 and 2022. [1] Demand continues to grow due to an aging population and increasing prevalence of osteoarthritis. Given the high volume and cost of these procedures, measuring cost variation is essential. Research shows that clinician decision-making—including choices around surgical technique, implant selection, and post-acute care settings—significantly influences episode costs. [2] [3]

The measure's logic model identifies key inputs such as multidisciplinary care teams, evidence-based guidelines, and data systems. These inputs support activities like implementing standardized care pathways, optimizing patients pre-operatively, coordinating post-operative care, and engaging in continuous quality improvement. These activities lead to outputs such as increased protocol adherence, improved care transitions, and enhanced patient engagement.

Evidence supports the effectiveness of these activities in achieving desired outcomes:

- Standardized care pathways reduce variation in care delivery and improve outcomes. A systematic review found that adherence to evidence-based protocols for hip arthroplasty improves hip scores and reduces complications. [3]
- Pre-operative optimization, including management of comorbidities and smoking cessation, has been shown to reduce surgical complications and improve recovery. [2]
- Post-operative care coordination significantly affects episode costs. Structured discharge planning and remote monitoring have been shown to reduce rehospitalizations and support discharge to home, lowering costs while maintaining outcomes. [4]
- Surgical technique also influences cost and outcomes. Minimally invasive approaches, such as anterior-lateral muscle-sparing techniques, are associated with shorter hospital stays and reduced blood loss, contributing to lower episode costs without compromising functional outcomes. [5]
- Clinician decision-making plays a central role in cost variation. Differences in implant selection, discharge planning, and rehabilitation referrals are major drivers of cost differences across providers. [6]

These findings align with the logic model's intermediate outcomes, including reduced complications, more efficient resource use, and decreased episode costs. Over time, these improvements contribute to long-term impacts such as sustained cost reductions, improved patient satisfaction and functional outcomes, and enhanced value in orthopedic care.

References:

1. American Joint Replacement Registry (AJRR). 2023 Annual Report. Rosemont, IL: American Academy of Orthopaedic Surgeons; 2023.
2. Politzer E, Anderson TS, Ayanian JZ, et al. "Perioperative costs of elective surgical procedures in Medicare Advantage compared with traditional Medicare." JAMA Health Forum. 2025;6(8):e2258.
3. Saueressig T, Owen PJ, Zebisch J, et al. "Evaluation of exercise interventions and outcomes after hip arthroplasty: a systematic review and meta-analysis." JAMA Netw Open. 2021;4(2):e2037741.
4. Carli F, Baldini G, Feldman LS. "Redesigning the preoperative clinic: from risk stratification to risk modification." JAMA Surg. 2021;156(2):191–192.
5. Xu T, Hutfless SM, Cooper MA, et al. "Hospital cost implications of increased use of minimally invasive surgery." JAMA Surg. 2015;150(5):489–490.
6. Yan L, Ge L, Dong S, et al. "Evaluation of comparative efficacy and safety of surgical approaches for total hip arthroplasty: a systematic review and network meta-analysis." JAMA Netw Open. 2023;6(1):e2253942.

2.4 Performance Gap

Provide evidence of performance gap or measurement gap by submitting performance scores on the measure as specified at the specified level(s) of analysis. Please include mean, minimum, maximum, and scores by deciles by using the table below or upload an attachment. In the text field here, describe the data source, including number of measured entities, number of patients, and dates of data. If a sample was used, provide characteristics of the entities included. Note that measures being submitted for maintenance review are expected to report on the performance gap, while measures submitted for initial endorsement may provide this information, if available. If the submitted measure is specified at more than one level of analysis, please attach additional performance gap tables using 2.4a below. If performance scores are unavailable for a maintenance measure, please explain.

Performance scores are reported for clinician groups (identified by Tax Identification Number [TIN]) and individual clinicians (identified by a combination of TIN and National Provider Identifier [NPI]) who were attributed 10 or more elective primary Hip Arthroplasty episodes. These episodes were identified from Medicare Parts A and B claims data for procedures ending between January 1, 2024, and December 31, 2024.

Scores reflect data from 1,932 clinician group practices and 6,244 individual practitioners, corresponding to 80,586 episodes at the TIN level and 47,403 episodes at the TIN-NPI level. These episodes represent care for 125,030 beneficiaries across all 50 states and Washington, D.C., and were delivered in acute inpatient hospitals, outpatient facilities, ambulatory/office-based care centers, and ambulatory surgical centers (ASCs).

Performance scores ranged from 0.59 to 1.64 at the TIN level and 0.75 to 1.79 at the TIN-NPI level. When multiplied by the national average episode cost of $13,240, these scores translated to per episode costs ranging from $7,811.60 to $21,713.60 for TINs and $9,930.00 to $23,699.60 for TIN-NPIs. The mean per episode cost was $13,637.20 for TINs and $13,240.00 for TIN-NPIs.

Additionally, per entity costs—which reflect the total cost burden per provider—ranged from $2.4 million to over $10.3 million for TINs and from $87,491 to over $616,000 for TIN-NPIs, highlighting substantial variation in resource utilization across deciles.

Table 1. Performance Scores by Decile

Enter overall mean, minimum, and maximum performance scores, along with the count of measured entities and persons/encounters/episodes. Organize entities into deciles by performance scores from 1 (low scores) to 10 (high scores), noting that "high" refers to magnitude, not quality. Provide mean performance scores, number of entities (total) and number of persons/encounters/episodes (total) for entities assigned to each decile. Note that measures being submitted for maintenance review are expected to report on the performance gap, while measures submitted for initial endorsement may provide this information, if available.

Table 1a. TIN Level Scores

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Performance Score	1.03	0.59	0.89	0.93	0.97	0.99	1.02	1.05	1.08	1.12	1.19	1.34	1.64
Mean Cost (Score × National Avg)	$13,637.20	$7,811.60	$11,783.60	$12,313.20	$12,842.80	$13,107.60	$13,504.80	$13,902.00	$14,299.20	$14,828.80	$15,755.60	$17,741.60	$21,713.60
Per Episode Cost	$13,637.20	$7,811.60	$11,783.60	$12,313.20	$12,842.80	$13,107.60	$13,504.80	$13,902.00	$14,299.20	$14,828.80	$15,755.60	$17,741.60	$21,713.60
Per Entity Cost	$568,823.71	$2,437,219.20	$487,218.28	$509,115.73	$531,013.18	$541,961.91	$558,384.99	$574,808.08	$591,231.17	$613,128.62	$651,449.16	$733,564.60	$10,335,673.60
N of Entities	1,932	1	193	193	193	193	193	193	193	193	193	193	1
N of Persons / Encounters / Episodes	80,586	312	7,980	7,980	7,980	7,980	7,980	7,980	7,980	7,980	7,980	7,980	476

Quick Tip

For cost measures calculated as ratios, be sure to report the costs against the national average.

Quick Tip

Reporting the per entity cost provides a better picture of the difference in cost across deciles.

Table 1b. TIN-NPI Level Scores

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Performance Score	1.00	0.75	0.87	0.90	0.93	0.96	0.99	1.02	1.05	1.09	1.15	1.22	1.79
Mean Cost (Score × National Avg)	$13,240.00	$9,930.00	$11,518.80	$11,916.00	$12,313.20	$12,710.40	$13,107.60	$13,504.80	$13,902.00	$14,431.60	$15,226.00	$16,152.80	$23,699.60
Per Episode Cost	$13,240.00	$9,930.00	$11,518.80	$11,916.00	$12,313.20	$12,710.40	$13,107.60	$13,504.80	$13,902.00	$14,431.60	$15,226.00	$16,152.80	$23,699.60
Per Entity Cost	$100,515.01	$566,010.00	$87,491.11	$90,363.00	$93,375.10	$96,387.20	$99,399.30	$102,411.40	$105,423.50	$109,439.63	$115,279.09	$122,100.72	$616,189.60
N of Entities	6,244	1	623	624	624	624	624	624	624	624	625	626	1
N of Persons / Encounters / Episodes	47,403	57	4,732	4,732	4,732	4,732	4,732	4,732	4,732	4,732	4,732	4,732	26

2.4a Attach Performance Gap Results

If needed, you may attach additional performance gap results here. If submitting an attachment rather than entering results in Table 1 above, please enter the overall mean, minimum, and maximum scores, and mean scores by decile. Enter the number of measured entities and persons/encounters/episodes overall and within each decile. Please clearly refer to any results within your attachment within the relevant text fields of this measure submission form.

One file only; 256 MB limit; allowed types: .zip, .pdf, .docx, .xls, .xlsx

2.6 Meaningfulness to Target Population*

Provide evidence the target population (e.g., patients) values the measured outcome, process, or structure, and finds it meaningful. Please describe the input collected from patient/caregivers consulted about the measure, including the number of patients/caregivers consulted and the number who agreed that the measure is meaningful and produces information that is valuable in making care decisions. For instrument submissions, please describe whether the target population finds it meaningful to provide feedback on their health care experiences addressed by the instrument's area of focus.

To ensure the Elective Primary Hip Arthroplasty episode-based cost measure is meaningful and valued by the target population, input was collected from patients and caregivers through their participation on a 10-person Person and Family Committee (PFC) during measure development. The PFC included person, family, and caregiver representatives to provide their perspectives towards the face validity of the measure and confirm that the outcomes and processes assessed are relevant to patient experiences and decision-making.

During PFC meetings, members reviewed the logic model and measure specifications, providing feedback on whether the measure reflects the elements that are most important and actionable from a patient perspective. Of the patients and caregivers consulted, all agreed that the cost and

quality of care for elective hip arthroplasty are meaningful outcomes, as these directly impact patient access, satisfaction, and financial burden. Their feedback affirmed that the measure produces valuable information for making decisions about surgical care and post-acute recovery options. This process of engaging patients and caregivers was integral to the measure's development and ongoing maintenance, ensuring the measure remains patient-centered and relevant to those it is designed to serve.

Section 3. Feasibility

3.1 Contributions Toward Distinguishing Differences in Care (optional)*

Describe the extent to which the measure can distinguish differences in care for certain patient subpopulations, which can be used to close gaps in care across those identified subpopulations.

- ***Evidence of Known Disparities:*** *Identify and describe from existing literature, internal analyses, etc. known and existing differences in health care and health outcomes across patient subpopulations (e.g., demographics, geography, insurance type) related to the measure focus area. Discuss how the findings relate to the measure and how the measure may contribute to efforts to improve health care delivery and outcomes for these identified groups.*
- ***Testing Methods:*** *If data are available, provide a description of your methodology and approach to empirically test differences in performance scores across identified groups (e.g., demographic, geographic variables).*
- ***Results and Interpretation:*** *If data are available, provide the results and an interpretation of those results, including explanations for any variations in performance scores across different groups. Discuss how these results relate to existing evidence (Element 1), any limitations found in the results, and the potential impact of these variations on the identified groups.*
- ***Anticipated Impact:*** *If information is available, discuss how accountable entities have utilized the measure to close gaps in health care delivery and outcomes for the identified groups.*

See the Closing Care Gaps Guidance for additional information.

Elective primary hip arthroplasty is a high-volume procedure with well-documented disparities in access, utilization, and outcomes across racial, ethnic, and socioeconomic groups. These disparities are influenced by both individual-level and community-level factors, and they persist despite the procedure's proven effectiveness in improving mobility and quality of life.

A systematic review of 44 studies found that insurance type, hospital volume, and geographic location significantly affect access to and outcomes after total joint arthroplasty. Patients with public insurance were less likely to receive timely evaluations and more likely to undergo surgery at low-volume hospitals, which are associated with higher complication rates and longer hospital stays. [1] Additionally, patients in rural areas had higher utilization of hip arthroplasty but faced challenges in accessing high-volume centers, which are linked to better outcomes. [1]

Hospital volume is a particularly strong predictor of outcomes. A meta-analysis found that low-volume hospitals were associated with significantly higher rates of surgical site infections, longer lengths of stay, increased costs, and higher 30-day and 1-year mortality rates compared to high-volume hospitals. [2] These findings suggest that centralizing care or improving processes at lower-volume facilities could reduce disparities in outcomes and costs.

Post-acute care decisions also contribute to disparities. A systematic review comparing post-acute care settings found that home health care (HHC) was consistently more cost-effective than skilled nursing facilities (SNFs) or inpatient rehabilitation facilities (IRFs), with comparable or better outcomes. [3] However, access to HHC may be limited for certain patient groups due to geographic or payer-related constraints, leading to higher costs and potentially poorer outcomes.

Bundled payment models have demonstrated potential to reduce costs and improve care coordination. An observational study of over 3,900 joint replacement episodes under the Bundled Payments for Care Improvement (BPCI) initiative found a 20.8% reduction in total spending per episode, primarily driven by decreased use of institutional post-acute care and lower implant costs. [4] These savings were achieved without compromising quality, as readmissions and emergency department visits declined and patient illness severity remained stable.

How the Measure Contributes to Equity

The Elective Primary Hip Arthroplasty episode-based cost measure supports equity in care delivery by:

- Focusing on clinician-attributed costs, which enables identification of variation in care decisions that may disproportionately affect patients with limited access to high-volume centers or home-based rehabilitation.
- Encouraging standardized care pathways, which reduce unwarranted variation and promote consistent, evidence-based care across all settings.
- Providing actionable data to clinicians and organizations to monitor disparities in cost and outcomes, enabling targeted interventions.
- Supporting alignment with MIPS Value Pathways, which integrate cost and quality measures to promote holistic, equitable care.
- Highlighting post-acute care decisions, which are a major driver of cost variation and disproportionately affect patients with limited access to home-based rehabilitation.

By promoting transparency and accountability in resource use, this measure supports CMS's broader goals of improving care for underserved populations and advancing health equity through value-based care.

References:

1. Yan L, Ge L, Dong S, et al. "Evaluation of Comparative Efficacy and Safety of Surgical Approaches for Total Hip Arthroplasty: A Systematic Review and Network Meta-analysis." JAMA Network Open. 2023;6(1):e2253942. doi:10.1001/jamanetworkopen.2022.53942.
2. Pincus D, Jenkinson R, Paterson M, et al. "Association Between Surgical Approach and Major Surgical Complications in Patients Undergoing Total Hip Arthroplasty." JAMA. 2020;323(11):1070–1076. doi:10.1001/jama.2020.0461.
3. Shenfeld DK, Navathe AS, Emanuel EJ. "The Promise and Challenge of Value-Based Payment." JAMA Internal Medicine. 2024;184(7):716–717. doi:10.1001/jamainternmed.2024.1343.
4. Navathe AS, Troxel AB, Liao JM, et al. "Cost of Joint Replacement Using Bundled Payment Models." JAMA Internal Medicine. 2017;177(2):214–222. doi:10.1001/jamainternmed.2016.8263.

Section 4. Feasibility

4.1a Data Structure and Availability*

Describe the extent to which all required data elements are routinely generated and used during care delivery, and whether these data elements are available in electronic sources. Specify if the data are in structured or unstructured fields. If all required data are not available electronically, provide a credible plan to implement electronic collection within one year. Additionally, discuss the extent of any missing data, the measure's susceptibility to inaccuracies, and the ability to audit data to detect problems. For a maintenance measure, also describe how any changes to the measure specifications impact data structure and availability.

The Elective Primary Hip Arthroplasty measure utilizes data elements routinely generated and used by healthcare personnel during care delivery, such as blood pressure, lab values, and medical conditions. These elements are coded by individuals other than those obtaining the original information, using DRG and ICD-9 codes on claims. All required data elements are available in defined fields across a combination of electronic sources, ensuring structured data collection.

CMS uses Medicare claims data, which provides a high degree of data completeness, as clinicians receive payments only for complete claims. CMS has established several auditing programs to assess claims code accuracy, detect fraud, and ensure appropriate billing. These programs include Zone Program Integrity Contractors (ZPICs), Recovery Audit Contractors (RACs), and the Comprehensive Error Rate Testing (CERT) Program, which collectively ensure data accuracy and integrity. Between 2005 and 2024, the CERT program estimated that proper payments, which met Medicare coverage, coding, and billing rules, ranged from 87.3% to 96.4% of total payments each year, with a proper payment rate of 93.7% in FY 2024. [1]

The measure excludes episodes with incomplete data, such as missing birth dates, or enrollment in Medicare Part C, to ensure comprehensive claims data. Additionally, the measure is calculated using data with a three-month claims run-out to ensure completeness and accuracy.

References:

1. Comprehensive Error Rate Testing (CERT) Program. 2024 Medicare Fee-for-Service Supplemental Improper Payments Data, Table A6. Centers for Medicare & Medicaid Services. https://www.cms.gov/files/document/2024-medicare-fee-service-supplemental-improper-payment-data.pdf.

4.1b Implementation Costs and Burden*

Describe any costs or burden of data collection, data entry, data validation, and analysis, including the impact on clinician workflow (e.g., modifications), diagnostic thought processes, and patient–physician interaction. Discuss barriers encountered or that could be encountered in implementing the measure specifications, data abstraction, measure calculation, or performance reporting.

The Elective Primary Hip Arthroplasty measure imposes no costs or burdens related to data collection, score calculation, or reporting. The measure score is automatically calculated by CMS using claims data routinely generated during care delivery, eliminating the need for additional data collection by eligible clinicians or clinician groups. Consequently, there is no impact on clinician workflow, diagnostic thought processes, or patient-physician interactions. The automatic generation and calculation of data

ensure seamless implementation of measure specifications, with no barriers to data collection, measure calculation, or performance reporting.

4.1c Confidentiality*

Explain the implementer's ability to collect information without violation of patient confidentiality, including circumstances where measures based on patient surveys, or the small number of patients may compromise confidentiality.

There are no concerns about patient confidentiality because the measure is based on administrative claims data submitted by eligible clinicians/eligible clinician groups to CMS, and CMS uses those data for both reimbursement and calculation of the measure score.

4.3 Feasibility Informed Final Measure*

Describe how the feasibility assessment informed the final measure specifications, indicating any decisions made to adjust the measure in response to feasibility assessment.

Because this is a claims-based measure there is no burden on the facility or its clinicians and no feasibility concerns; rates are automatically calculated by CMS based on claims data submitted by facilities for payment.

4.4 Proprietary Information*

Indicate whether your measure or any of its components are proprietary, with or without fees (choose one).

☐ Proprietary measure or components (e.g., risk model, codes), without fees
☐ Proprietary measure or components with fees
☒ Not a proprietary measure and no proprietary components

4.4a *[If any proprietary components (4.4)]* **Fees, Licensing, or Other Requirements***

Describe any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm).

Not applicable.

Section 5. Scientific Acceptability

5.1 Data and Samples

5.1.1 Data Used for Testing*

Describe the data used for testing (include dates, sources).

The Elective Primary Hip Arthroplasty Measure uses Medicare Parts A and 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 adjustors. Data from the EDB is used to determine patient-level exclusions and supplemental risk adjustors, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), patient birth dates, and patient death dates. The risk-adjustment model also accounts for expected differences in payment for services provided to patients in long-term care based on 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 United States Census Bureau American Census, United States Census Bureau American Community Survey (ACS), and Common Medicare Enrollment (CME) are used in the analyses evaluating social risk factors in risk adjustment.

Testing includes Elective Primary Hip Arthroplasty episodes ending from January 1, 2024, to December 31, 2024. The reliability analysis also includes episodes ending in the 2023 calendar year. For further details.

5.1.2 Differences in Data*

If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), clearly identify which data source/sample is used for each aspect of testing, including the years of data used in each. If there are no differences to report, enter "None."

The testing for reliability includes episodes from calendar years 2023 and 2024. All other testing used the study period of January 1, 2024, to December 31, 2024.

The exclusion analysis used a greater population of episodes without inclusion criteria applied. This includes 4,031 TINs and 17,135 TIN-NPIs, 161,226 patients, and 168,011 episodes. All other testing used the study population, which includes 1,932 clinician group practices and 6,244 practitioners, corresponding to 117,489 Medicare patients (from 120,228 episodes) included in TIN-level testing, and 106,572 patients (from 109,070 episodes) included at TIN-NPI-level testing. Episodes are included from all 50 states and Washington, D.C. in the following settings: acute IP hospitals, OP facilities, ambulatory/office-based care centers, and ASC.

5.1.3 Characteristics of Measured Entities*

Describe characteristics of measured entities included in the analysis (e.g., number, size, location, type). If you used a sample, describe how you selected measured entities for inclusion in the sample and the representativeness of the sample.

There were 1,932 clinician group practices (identified by TIN) and 6,244 practitioners (identified by TIN and NPI) included in testing of the Elective Primary Hip Arthroplasty (also referred to as "the Hip Arthroplasty Measure"). Clinicians and clinician groups were included if they were attributed 10 or more episodes, as identified in Medicare Parts A and B claims data, ending from January 1, 2019, to December 31, 2019. Episodes were included from all 50 states and Washington, D.C. in the following settings: acute IP hospitals, OP facilities, ambulatory/office-based care centers, and ASC.

5.1.4 Characteristics of Units of the Eligible Population*

Describe characteristics of the patients, encounters, episodes, etc., including numbers and percentages by factors such as age, sex, race, or diagnosis. Provide descriptive statistics separately by each specified level of analysis and data source. If you used a sample, describe how you selected the patients for inclusion in the sample and the representativeness of the sample. If there is a minimum case count used for testing, you must reflect that minimum in the specifications in ***1.26 Minimum Sample Size in Section 1, Measure Information****.*

There were 117,489 Medicare patients (from 120,228 episodes) included in TIN level testing, and 106,572 patients (from 109,070 episodes) included at TIN-NPI level testing. Episodes are triggered by Current Procedural Terminology (CPT) / Healthcare Common Procedure Coding System (HCPCS) codes on Part B Physician/Supplier claims which indicates occurrence of an elective primary hip arthroplasty.

Episodes were included in the sample if they met a set of inclusion criteria (listed below), meant to ensure completeness of data and focus the measure on a clinically homogeneous cohort of patients receiving an

elective primary hip arthroplasty. As previously mentioned, a 10-episode case minimum was also applied. These inclusion criteria are listed below:

- The patient had 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 patient 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 patient date of birth is not missing.
- The patient death date did not occur before the trigger date.
- The patient death date did not occur before episode end.
- The episode can be attributed to at least one main surgeon.
- The episode trigger claim was in an ambulatory/office-based care setting, IP hospital, OP hospital, or ASC based on its place of service.
- If the procedure occurred in an inpatient setting, the inpatient stay occurred in either an acute hospital as defined by subsection (d) or in an acute hospital in Maryland[2].
- If the procedure occurred in an inpatient setting, the inpatient stay had a relevant MS-DRG code.
- The patient did not undergo a staged or same-day bilateral hip replacement procedure.
- The hip arthroplasty procedure was not performed due to cancer, hip fracture, or trauma.
- The patient did not have congenital deformity of the hip, osteomyelitis of the hip or femur, or a septic joint.
- The episode is not an outlier case.

To ensure that the inclusion criteria listed above do not distort patient characteristics within the measure population, we compared distributions of patient characteristics (age, race, sex, dual eligibility status, income, unemployment, hierarchical condition categories [HCCs]) for patients and episodes before and after applying the inclusion criteria.

Results of this analysis show that the Hip Arthroplasty Measure inclusion criteria have a minimal effect on the distribution of patient characteristics within the measure population. Across all demographic categories, the difference in proportion of patients before and after applying inclusion criteria is less than 3 percentage points. To illustrate, the measure population is 60.7 percent female before the inclusion criteria is applied, compared with 61.1 percent after criteria is applied at TIN level analysis. This is comparable to TIN-NPI level results where the population is 61.2 percent female after inclusion criteria is applied. When it comes to race categories, the population is 90.4 percent White without inclusion criteria; after inclusion criteria is applied, this statistic is 91.2 percent at TIN level analysis and 91.5 percent at TIN-NPI level analysis. In terms of age, 28.8 percent of the population is between ages 65 and 69 before inclusion criteria is applied, compared with 27.9 percent at TIN level analysis and 28.0 percent at TIN-NPI level analysis after inclusion criteria is applied. Similarly, 26.7 percent of the population is between ages 70 and 74 before inclusion criteria, compared with 28.6 percent at TIN level analysis and 28.8 percent at TIN-NPI level analysis after inclusion criteria is applied. Finally, 19.3 percent of patients are between ages 75 and 79 before inclusion criteria, compared with 20.5 percent at TIN level analysis and 20.6 percent at TIN-NPI level analysis after the criteria is applied.

2 Subsection (d) covers 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.

Full results of this analysis can be seen in Appendix Table 1.6 in the Supplemental Attachment section 7. These results indicate that there is minimal shift in patient characteristics as a result of the inclusion criteria listed in this section.

5.2 Reliability

5.2.1 Level(s) of Reliability Testing Conducted*

Choose all that apply.

Note: Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers). If more than one level of analysis is specified, testing must be conducted for each level separately.

☐ Person or encounter level (i.e., data element) (e.g., inter-abstractor reliability)
☒ Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)
☐ Not applicable/reliability testing not conducted

5.2.1a Why Testing Not Conducted*

*Explain why reliability testing was not conducted (derived)**
[character limit: 255]

5.2.2 *[If reliability testing was conducted]* **Method(s) of Reliability Testing***

For each level of reliability testing conducted, describe the method(s) of reliability testing and explain what each tests. Describe the steps; do not just name a method. Describe the type of error each method is designed to detect. Provide the type of statistical analysis used. Describe proportion of missing data, how missing data were analyzed and/or excluded, and any sensitivity analysis conducted.

Note: Testing at the person or encounter level requires that all critical data elements be tested (not just agreement of one final overall computation for all persons/encounters/episodes). At a minimum, the numerator, denominator, and exclusions must be assessed and reported separately. Prior evidence of reliability of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data elements. Prior evidence could include published or unpublished testing that: includes the same data elements, uses the same data type (e.g., claims, chart abstraction), and is conducted on a sample as described above (i.e., representative, adequate numbers, and randomly selected, if possible).

Do not report accountable entity level reliability testing for instrument submissions (this information will be collected for each IDM).

Reminder

This measure example is being submitted for maintenance, so person- or encounter-level testing is not required. However, if the measure were being submitted for initial endorsement, person- or encounter-level testing would be necessary.

Because this measure uses standardized prices generated through an algorithmic process, reliability is generally assumed. However, it is important to provide evidence demonstrating the accuracy and consistency of the underlying data used in the standardization algorithm. This should include documentation or citations that clearly describe the algorithm, explain its inputs (such as claims data and service codes), and provide justification for any updates made to the algorithm over time.

Additionally, evidence should confirm that the data sources used are accurate, complete, and reliable. This may include references to data audit procedures, third-party validations, previous testing results, or relevant peer-reviewed publications and technical reports that support the reliability of the standardized pricing methodology.

Reliability Assessment

Reliability indicates the extent to which repeated measurements yield consistent results. In evaluating the episode-based arthroplasty measure, we applied a split-sample methodology to assess reliability, as recommended. [1] This approach is preferred because it does not rely on parametric assumptions about the data and is applicable to any type of quality measure.

Split Sample Methodology

To obtain a reliability estimate for the episode-based arthroplasty measure, we used a split sample methodology and calculated the intraclass correlation coefficient (ICC) of the measure scores. For each TIN or TIN-NPI, we randomly split the population of episodes from a larger sample collected in 2023 and 2024 without replacement. The denominator, numerator, and measure score were calculated for each sample within each entity. Thus, the measure score is calculated twice for each TIN or TIN-NPI among two distinct and randomly selected sets of episodes. We used a one-way random effects model to calculate the ICC to evaluate the agreement between the two randomly selected samples. [2][3][4] A higher ICC value indicates greater agreement and, therefore, greater reliability. We follow the guidance regarding the interpretation of reliability using the 95% confidence interval of the ICC: <0.5 = poor; 0.5–0.75 = moderate; 0.75–0.9 = good; and > 0.9 = excellent. [2]

Permutation Split-Sample

A limitation of the split sample approach is that the reliability estimates may vary based on the random split of the data; that is, estimates may vary across different split sample draws. [5] Therefore, we took many splits of the data, estimating reliability for each split, and then averaging the reliability estimates to arrive at a more stable estimate.

In our analysis, only TIN and TIN-NPIs that met a case minimum of 10 episodes in both samples were included. We repeated the sample splitting 10 times to confirm the stability of our estimate. We did not conduct additional splits given clear stability in the results and in light of the time required to conduct additional runs. The main drawback of the permutation split sample approach is the computational time required. [5]

Spearman-Brown Correction

The Spearman-Brown formula was used for sample size correction to obtain reliability estimates for each TIN or TIN-NPI. We report the mean, minimum, and maximum values, and values by decile for the Spearman-Brown adjusted ICCs calculated on 10 sets of random split samples without replacement within each entity.

References

1. Nieser KJ, Harris AHS. Comparing methods for assessing the reliability of health care quality measures. Stat Med. 2024;43(23):4575-4594. PMID: 39145538.

2. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31.

3. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1:30–46.

4. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86:420–428.

5. Nieser KJ, Harris AHS. Split-sample reliability estimation in health care quality measurement: Once is not enough. Health Serv Res. 2024;59(4):e14310. PMID: 38659301.

5.2.3 *[If reliability testing was conducted]* **Reliability Testing Results***

Provide the statistical results from reliability testing for each level and type of reliability testing conducted. Where applicable, include results from accountable entity level reliability testing (e.g., signal-to-noise testing) in Table 2. Do not report accountable entity level reliability testing for instrument submissions (this information will be collected for each IDM).

Permutation Split Sample ICC Estimates

Using the episode-based care reliability testing approach, we evaluated the reliability of the arthroplasty measure by implementing a permutation split sample methodology. This involves calculating the intraclass correlation coefficient (ICC) estimates across multiple random splits to ensure stability and accuracy.

The reliability tables in section 5.2.3a below, provide the mean reliability, performance score, number of entities (total) and number of persons/encounters/episodes (total) for entities assigned to each decile.

Spearman-Brown Corrected Entity-Level Reliability Estimates

For the TIN or TIN-NPI entities, the reliability estimates ranged from a minimum of 0.68 to a maximum of 0.99, with a mean reliability of 0.88 for TIN-NPI and from 0.69 to 0.98 with a mean of 0.86 for TIN. Even the lowest reliability estimate is above the consensus-based entity's expected threshold of 0.60,further validating the robustness of the measure. By applying the Spearman-Brown correction, we adjusted for sample size variations, ensuring precise and reliable estimates across entities.

Mean Performance Scores:

TIN Level:

Overall Mean Performance Score: 1.08

Minimum: 1.89

Maximum: 0.49

TIN-NPI Level:

Overall Mean Performance Score: 1.12

Minimum: 1.85

Maximum: 0.48

These performance scores, along with reliability estimates, provide a comprehensive view of the measure's effectiveness across different entity levels, demonstrating its capability to accurately reflect care efficiency.

Entity and Episode Distribution:

TIN Level:

Total Entities: 1,932

Total Episodes: 117,489

TIN-NPI Level:

Total Entities: 6,244

Total Episodes: 106,572

These distributions ensure that the measure is applicable across a broad range of entities and patient episodes, enhancing its generalizability and utility in real-world settings.

The above results highlight the robustness and reliability of the arthroplasty measure across both TIN and TIN-NPI levels, ensuring its effectiveness in evaluating provider care efficiency.

5.2.3a *[If reliability testing was conducted]* **Attach Additional Reliability Testing Results**

If needed, you may attach additional reliability testing results here. Please clearly refer to any results within your attachment within the relevant text fields of this measure submission form.

One file only; 256 MB limit; allowed types: .zip, .pdf, .docx, .xls, .xlsx

Table 2. *[If accountable entity level testing was conducted, i.e., if 4.2.1 includes "Accountable entity level")]* **Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size**

Enter overall mean, minimum, and maximum performance scores, along with the count of measured entities and persons/encounters/episodes. Organize entities into deciles by the entity number of persons/encounters/episodes (denominator) from 1 (smallest N) to 10 (largest N). Provide mean reliability, performance score, number of entities (total) and number of persons/encounters/episodes (total) for entities assigned to each decile. For minimum reliability, provide reliability value for the entity with the smallest N. For maximum reliability, provide the reliability value for the entity with the largest N.

Table 2a. TIN Level Reliability Estimates

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.88	0.69	0.90	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1.00	1.00
Mean Performance Score	1.08	1.89	1.87	1.70	1.51	1.32	1.13	0.94	0.75	0.56	0.54	0.52	0.49
N of Entities	1,932	1	193	193	193	193	193	193	193	193	193	193	1
N of Episodes	117,489	52	8,102	8,912	9,723	10,533	11,343	12,154	12,964	13,774	14,584	15,400	157

Table 2b. TIN-NPI Level Reliability Estimates

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.89	0.68	0.71	0.74	0.77	0.80	0.83	0.87	0.90	0.93	0.96	0.99	0.99
Mean Performance Score	1.12	1.80	1.85	1.83	1.68	1.49	1.30	1.11	0.92	0.73	0.54	0.51	0.50
N of Entities	6,242	1	624	624	624	624	624	624	624	624	624	624	1
N of Episodes	106,572	31	3,231	3,877	4,652	5,583	6,699	8,039	9,647	11,577	13,892	16,671	59

5.2.4 *[If reliability testing was conducted]* **Interpretation of Reliability Results***

Provide your interpretation of the results in terms of demonstrating reliability for each level and type of reliability testing conducted. Describe how the results support an inference of reliability for the measure or instrument.

The testing results indicate that the accountable entity measure scores are reliable.

Reliability estimates were calculated using a permutation split-sample methodology with Spearman-Brown correction. For TIN-NPI level entities, reliability ranged from a minimum of 0.68 to a maximum of 0.99, with a mean reliability of 0.89, indicating excellent consistency across providers. For TIN-level entities, reliability ranged from 0.69 to 1.00, with a mean reliability of 0.88, also reflecting excellent reliability. These results demonstrate that the measure can consistently distinguish performance across entities with varying episode volumes. [1]

Reference:

1. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31.

5.3 Validity

5.3.1 Level(s) of Validity Testing Conducted*

Choose all that apply.

Note: Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers). If more than one level of analysis is specified, testing must be conducted for each level separately.

☐ Person or encounter level (i.e., data element) (e.g., sensitivity and specificity)
☒ Accountable entity level (i.e., measure score) (e.g., criterion validity)
☐ Not applicable/validity testing not conducted

5.3.1a Why Testing Not Conducted*

Provide a rationale for why validity testing is not applicable/was not conducted [character limit: 255]

5.3.2 Type of Accountable Entity Level Validity Testing Conducted*

Choose all that apply.

Note: Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers). If more than one level of analysis is specified, testing must be conducted for each level separately.

☒ Empirical validity testing at the accountable entity level (e.g., criterion validity, construct validity, known groups analysis)

☐ Systematic assessment of face validity of the measure's performance score as an indicator of quality or resource use *[i.e., the score is an accurate reflection of the effect of performance on quality or resource use and can distinguish good from poor performance]*

☐ Not applicable/accountable entity level validity testing not conducted

5.3.2a *[If a maintenance measure]* **Why Testing Not Conducted***

Provide a rationale for why accountable entity level validity testing was not conducted. [character limit: 255]

5.3.3 *[If validity testing was conducted]* **Method(s) of Validity Testing***

For each level of testing conducted, describe the method(s) of validity testing and what each tests. Describe the steps (do not just name a method) and explain what was tested (e.g., accuracy of data elements compared with authoritative source, relationship to another measure as expected). Describe the statistical analysis used. Describe proportion of missing data, how missing data were analyzed and/or excluded, and any sensitivity analysis conducted.

Note: Testing at the person or encounter level requires that all critical data elements be tested (not just agreement of one final overall computation for all patients). At a minimum, the numerator, denominator, and exclusions must be assessed and reported separately. For person or encounter level testing, prior evidence of validity of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data elements. Prior evidence could include published or unpublished testing that: includes the same data elements, uses the same data type (e.g., claims, chart abstraction), and is conducted on a sample as described above (i.e., representative, adequate numbers, and randomly selected, if possible).

For empirical accountable entity level testing, the following should be included:

- *Narrative describing the hypothesized relationships*
- *Narrative describing why examining these relationships (e.g., correlating measures) would validate the measure*
- *Expected direction of the association*
- *Expected strength of the association*

Do not report accountable entity level validity testing for instrument submissions (this information will be collected for each IDM).

To establish construct validity of the episode-based cost measure for Elective Primary Hip Arthroplasty, we correlated performance scores for both TIN and TIN-NPI levels with CBE #3493 - *Risk-standardized Complication Rate (RSCR) following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for Eligible Clinicians and Eligible Clinician Groups*. This correlation is justified by the fact that both measures target similar patient populations undergoing elective THA and TKA procedures, ensuring comparability in demographics and clinical characteristics. Additionally, both measures involve similar care pathways that emphasize the prevention of unnecessary complications post-THA/TKA, aligning with the objectives of the episode-based cost measure to manage costs and resources effectively while minimizing complications.

Quick Tip

Provide a justification for the correlation.

Our hypothesis is that the measure scores for the episode-based cost measure and the RSCR will be highly correlated and in the same direction. Specifically, as complication scores decrease, episode-based cost scores are expected to decrease, indicating that reduced complications lead to lower associated costs. This hypothesis is grounded in the understanding that efficient care resulting in fewer complications should also lead to lower overall costs, reflecting enhanced care delivery and resource utilization.

Quick Tip

Provide a hypothesis when describing the correlation analysis approach.

For the correlation analysis, we will utilize Pearson's correlation coefficient. Pearson's correlation is appropriate for this analysis as it quantifies the degree to which two continuous variables are linearly related, providing insight into the direct relationship between complication rates and cost measures. According to the literature, the correlation coefficient can be interpreted as follows: 0.00 to 0.19 indicates a very weak correlation, 0.20 to 0.39 a weak correlation, 0.40 to 0.59 a moderate correlation, 0.60 to 0.79 a strong correlation, and 0.80 to 1.00 a very strong correlation. [1] While the magnitude of the correlation provides insight into the strength of association, we interpret it within the context of shared mechanisms and conceptual overlap between the two measures. A moderate to strong negative correlation would support construct validity—not simply because of its statistical strength, but because it reflects a meaningful relationship between clinical outcomes and resource use.

However, we recognize that correlation alone is insufficient to establish validity. Validity requires demonstrating that differences in observed-to-expected costs are causally attributable to the clinician or group, rather than to chance, confounding variables, or external factors. To support this, the measure design incorporates:

- Risk adjustment to control for patient-level confounders.
- Attribution logic that links episodes to clinicians based on meaningful clinical involvement.
- Exclusion of outlier episodes to reduce noise from atypical cases.

These design features strengthen the argument that variation in cost scores reflects true differences in care delivery, rather than random variation or external influences. This approach aligns with modern validity theory, which emphasizes the integration of empirical evidence and mechanism-based reasoning to support causal inferences.

The expected outcome is a statistically significant, positive correlation, interpreted in light of the shared clinical mechanisms and controlled confounding. This would reinforce the measure's validity as an indicator of care efficiency and resource stewardship.

Reference:

1. Akoglu, H. (2018). User's guide to correlation coefficients. Turkish Journal of Emergency Medicine, 18(3), 91-93.

5.3.4 *[If validity testing was conducted]* **Validity Testing Results***

Provide the statistical results from validity testing for each level and type of validity testing conducted. Do not report accountable entity level validity testing for instrument submissions (this information will be collected for each IDM).

These results in Table 3 below indicate a strong positive correlation at both the TIN and TIN-NPI levels, supporting the hypothesis that as complication rates decrease, episode-based costs also decrease, reflecting efficient care delivery and resource utilization. The statistical significance of the correlation reinforces the validity of the episode-based cost measure as a reliable indicator of efficient care delivery, demonstrating its effectiveness in evaluating provider performance and guiding decision-making processes to enhance patient outcomes and reduce healthcare costs.

5.3.4a *[If validity testing was conducted]* **Attach Additional Validity Testing Results**

If needed, you may attach additional validity testing results here. Please clearly refer to any results within your attachment within the relevant text fields of this measure submission form.

One file only; 256 MB limit; allowed types: .zip, .pdf, .docx, .xls, .xlsx

Table 3: Correlation Results

Level	Correlation Coefficient	Statistical Significance	Interpretation	Number of Entities Analyzed
TIN Level	+0.75	$p < 0.01$	Strong Positive Correlation	1,932
TIN-NPI Level	+0.72	$p < 0.01$	Strong Positive Correlation	6,244

5.3.5 *[If validity testing was conducted]* **Interpretation of Validity Results***

Provide your interpretation of the results in terms of demonstrating validity for each level and type of validity testing conducted. Describe how the results support an inference of validity for the measure. For accountable entity level testing (if applicable), discuss how the results relate to the hypothesis. If the results are not what were expected, explain why.

Based on the correlation analysis results, the findings demonstrate validity for the episode-based cost measure at both the TIN and TIN-NPI levels. The strong positive correlation coefficients (+0.75 at the TIN level and +0.72 at the TIN-NPI level) indicate that lower complication rates are associated with reduced episode-based costs, supporting the hypothesis that efficient care delivery results in cost savings. The statistical significance of these correlations ($p < 0.01$) reinforces the measure's reliability as an indicator of efficient care delivery.

Construct Validity

The strong positive correlation supports the construct validity of the measure, indicating that it accurately reflects the relationship between complication rates and healthcare costs. This aligns with findings from the JAMA Health Forum, which emphasized that episode-based cost measures (EBCMs) are designed to capture clinically-related costs and reflect care efficiency. [1] Additionally, a 2023 study comparing risk adjustment methods found that episode-based cost models reliably explained cost variation across diagnostic categories. [2]

Causal Explanation

The correlation results suggest that effective care practices—such as adherence to safety protocols and preventive measures—directly contribute to reducing complications and associated costs. This causal relationship is supported by a large-scale observational study using National Surgical Quality Improvement Program (NSQIP) data, which found that postoperative complications significantly increase health care costs, with estimates ranging from $11,626 to $19,626 per patient. [3] Similarly, a JAMA study showed that adverse event rates declined significantly over time, suggesting that safety improvements can lead to cost reductions. [4]

External Validity Justification

The literature provides mechanistic insights into how effective care practices lead to cost savings. For example, the Hospital at Home model demonstrated that substituting inpatient care with home-based care reduced complications, readmissions, and costs while improving patient satisfaction. [5] These findings support the external validity of episode-based cost measures, showing that the principles of efficient care delivery apply across diverse settings and populations.

Unexplained Variance

While the strong positive correlations observed (+0.75 at the TIN level and +0.72 at the TIN-NPI level) support the validity of the episode-based cost measure, approximately 25% of the variance remains unexplained, indicating the potential influence of residual confounders. These may include certain patient-level factors access to follow-up care, health literacy, etc.; provider-level factors like practice patterns and experience; and system-level influences such as hospital infrastructure and regional cost variation. For example, surgical complications can significantly increase hospital and payer costs, but variation in procedure type and hospital characteristics can confound cost estimates even after risk adjustment. [6] These findings suggest continued exploration into addressing these confounders through various statistical approaches to ensure that the observed correlations continue to reflect true relationships rather than artifacts of unmeasured or misadjusted variables.

References:

1. Duseja, R., Andress, J., Sandhu, A.T., et al. (2021). Development of Episode-Based Cost Measures for the US Medicare Merit-based Incentive Payment System. JAMA Health Forum, 2(5), e210451. https://doi.org/10.1001/jamahealthforum.2021.0451

2. Kim, J., Ock, M., Oh, I.H., et al. (2023). Comparison of diagnosis-based risk adjustment methods for

episode-based costs to apply in efficiency measurement. BMC Health Services Research, 23, Article 1334. https://doi.org/10.1186/s12913-023-10282-4[2](https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-10282-4)

3. Dencker, E.E., Bonde, A., Troelsen, A., Varadarajan, K.M., & Sillesen, M. (2021). Postoperative complications: an observational study of trends in the United States from 2012 to 2018. BMC Surgery, 21, Article 393. https://doi.org/10.1186/s12893-021-01392-z

4. Eldridge, N., Wang, Y., Metersky, M., et al. (2022). Trends in Adverse Event Rates in Hospitalized Patients, 2010–2019. JAMA, 328(2), 173–183. https://doi.org/10.1001/jama.2022.9600

5. Levine, D.M., Souza, J., Schnipper, J., et al. (2024). Acute Hospital Care at Home in the United States: The Early National Experience. Annals of Internal Medicine. https://doi.org/10.7326/M23-2264

6. Healy, M.A., Mullard, A.J., Campbell, D.A. Jr., et al. (2016). Hospital and Payer Costs Associated With Surgical Complications. JAMA Surgery, 151(9), 823–830. https://doi.org/10.1001/jamasurg.2016.0773

5.4 Risk Adjustment

5.4.1 Methods Used to Address Risk Factors*

Describe the methods or approaches that were used to explore the effects of risk factors on this measure. (Note: If you tested for the effects of risk factors and ultimately determined that risk adjustment or stratification was not warranted, please select the method(s) used and provide details of the testing and your rationale in 5.4.2 through 5.4.6; the measure's ultimate status will be reported in 5.4.7). Choose all that apply.

☒ Statistical risk adjustment model with risk factors
☐ Statistical case-mix adjustment model
☐ Stratification by risk factor category
☐ Other

5.4.1a Other Methods Used* *(Please describe briefly) [character limit: 255]*

☐ No risk adjustment or stratification.

5.4.1b *[If 1.7 Measure Type is outcome, intermediate outcome, cost/resource, PRO–PM, PRE-PM]* **Rationale For No Adjustment or Stratification**

Provide a rationale for why there is no need to address differences in patient characteristics (e.g., case mix, social risk factors) to achieve fair comparisons across measured entities for your outcome, intermediate outcome, resource, or patient-reported measure.

5.4.2. *[If risk/case-mix factors are addressed by any method (5.4.1)]* **Conceptual Model Rationale***

Explain the rationale for the approach selected, including reasons for adjustment for risk factors or case mix, and/or stratification. Describe the sources that inform the conceptual model, e.g., scientific literature, unpublished findings, TEP. Consider age, gender, race, ethnicity, urbanicity/rurality, Medicare/Medicaid dual eligibility status, indices of social vulnerability (e.g., Centers for Disease Control and Prevention Social Vulnerability Index), and markers of functional status-related risk (e.g., cognitive or physical function) in the conceptual model, using evidence to support the model, with references. If risk or case mix factors (e.g., social, functional status–related, clinical) are included in the conceptual model but data are not available for all factors, describe any potential bias, as a result of not including these factor(s) in the final risk adjustment or case-mix adjustment model or stratification. Address the validity of the measure in light of this bias.

The measure's risk-adjustment model is meticulously crafted to evaluate the costs associated with elective hip arthroplasties by integrating both clinical and social risk factors. This model ensures that variations in patient characteristics are accurately considered, providing a fair and precise assessment of health care outcomes. The model is informed by a combination of scientific literature, expert clinician input, and empirical analysis to ensure comprehensive risk adjustment.

Clinical Factors:

In selecting clinical variables for the risk-adjustment model, ten key factors were considered to ensure relevance, reliability, and unique contribution to the model's accuracy:

- Clinical/Conceptual Relationship: The model prioritizes factors with a direct clinical relationship to the cost of care for elective hip arthroplasties. Expert clinicians identified these factors, ensuring their relevance to the clinical pathways and potential complications associated with hip arthroplasty procedures.
- Empirical Association: Empirical analyses were conducted to confirm the association of selected risk factors with the outcome of interest, ensuring each factor has a proven impact on health care costs and outcomes.
- Variation in Prevalence: Factors were chosen based on their variation across measured entities, capturing differences in patient populations and ensuring broad applicability across diverse health care settings.
- Presence at Start of Care: Risk factors are identified using claims data available at the start or prior to the episode, ensuring they reflect patient characteristics rather than the care provided. This approach mitigates the risk of manipulation or gaming, maintaining the integrity of the risk-adjustment process.
- Resistance to Manipulation: The model includes factors resistant to manipulation, ensuring that risk adjustment remains accurate and reliable, maintaining the credibility of performance measures.
- Data Accuracy: All included factors have reliable and feasible data capture, ensuring precision in risk adjustment. The use of claims data ensures that information is systematically collected and consistently available.
- Unique Variation Contribution: Each factor contributes unique variation to the outcome, enhancing the model's discrimination and calibration, accurately reflecting the complexities of patient care and resource use.
- Improvement of Risk Model: Continuous testing evaluates the impact of including certain risk factors on the overall risk model, ensuring ongoing refinement and improvement based on empirical findings and stakeholder feedback.
- Face Validity and Acceptability: Selected factors are clinically relevant and acceptable to stakeholders, ensuring the model's credibility and utility. The involvement of expert clinicians in the selection process enhances the model's face validity.

Recommended Clinical Variables:

Based on these criteria, the Hip Arthroplasty Measure workgroup recommended accounting for specific patient variables such as history of anti-platelet medications, opioid dependence, deep vein thrombosis or pulmonary embolism and frailty indicators. These variables are associated with increased risks and costs, reflecting their clinical significance in the context of hip arthroplasty procedures.

Integration of CMS-HCC Model:

Beyond these measure-specific factors, the CMS-HCC model is integrated into the risk-adjustment framework based on studies evaluating its suitability for risk adjusting Medicare claims data. Developed specifically for the Medicare population, the CMS-HCC model accounts for prevalent conditions and is calibrated on Medicare Fee-for-Service (FFS) patients. Its routine updates for coding changes ensure exhaustive coverage of clinical risk data, supporting its adaptation to the Hip Arthroplasty Measure.

By integrating these clinical factors and the CMS-HCC model, the risk adjustment framework provides a robust mechanism for evaluating health care costs and outcomes, ensuring that variations in patient characteristics are appropriately considered. This approach aligns with NQF evaluation criteria, supporting the measure's validity and reliability in assessing elective hip arthroplasty costs.

Social Risk Factors:

The inclusion of social risk factors in the Hip Arthroplasty Measure risk adjustment model highlight the complex mechanisms underlying differences in resource use by socioeconomic status and race. These differences are influenced by factors such as financial resources, community resources, historical and current discrimination, and reduced access to preventive services. Provider assumptions or implicit biases can further impact the quality of care for patients of different races, potentially leading to inefficient care, increased disease severity, or greater morbidity. Consequently, these factors can result in higher Medicare spending based on socioeconomic or demographic status.

Key Social Risk Factors:

The literature identifies several social risk factors that affect resource use, including:

- Income: Economic status can directly impact access to health care services and the ability to afford necessary treatments.
- Insurance (e.g., Medicaid): Insurance coverage influences the extent and type of health care services accessible to patients.
- Education: Educational attainment is linked to health literacy and the ability to navigate health care systems effectively.
- Race and Ethnicity: These demographic factors can influence both access to care and provider biases, affecting health care outcomes.
- Sex: Gender differences can impact health care needs and resource utilization.
- Social Relationships: The presence or absence of supportive social networks can affect health outcomes and recovery processes.
- Residential and Community Context: Factors such as rurality can influence access to health care facilities and services.

Analytical Approach:

Given the conceptual relationship between these social risk factors and resource use, the model analyzes the impact of patient-level and Census-Block Group-level factors, including income, education, employment, race, sex, dual status, and the AHRQ Index. These factors are detailed in Section 1.8 of the documentation.

Data Sources: The CMS Enrollment Database (EDB) and Common Medicare Environment (CME) are utilized to determine dual eligibility, race, and sex. Socioeconomic status is assessed through two approaches: categorical dependents (income, education, employment status) and the Agency for Healthcare Research and Quality (AHRQ) SES Index as a continuous dependent variable. Data from the 2017 American Community Survey (5-year file) is linked to episodes via census block groups or ZIP codes when census block group data is unavailable.

Stepwise Analysis: Social risk factors are examined relative to the base set of risk adjustment variables from the CMS-HCC V22 2016 model, including disability status, ESRD status, interaction variables, and recent long-term care use. This stepwise approach determines the potential value of each social risk factor considered.

By systematically evaluating these social risk factors, the model aims to understand their impact on health care resource use and refine the risk adjustment process to ensure fair and accurate measurement of costs associated with hip arthroplasties.

5.4.2a *[If risk/case-mix factors are addressed by any method (5.4.1)]* **Attach Conceptual Model***

Attach a figure of the conceptual model that illustrates the hypothesized pathway between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured outcome.

One file only; 256 MB limit; allowed types: .pdf, .jpg, .png, .zip

5.4.3 *[If risk/case-mix factors are addressed by any method (5.4.1)]* **Variable Distribution Across Measured Entities***

Provide descriptive statistics showing how the risk/case mix variables identified from the conceptual model are distributed across the measured entities. Indicate which factors were tested in the risk/case-mix adjustment model and which were tested for stratifying the measure, as applicable.

Full measure population demographics can be found in Appendix Table 2b3.4b. Appendix Table 2b3.1.1 includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.

5.4.4 *[If risk/case-mix factors are addressed by any method (5.4.1)]* **Risk/Case-Mix Adjustment Modeling and/or Stratification Results***

Describe the statistical results of the analyses used to test and select risk/case-mix factors for inclusion in or exclusion from the adjusted model and/or stratification, as applicable. Clearly indicate the risk/case-mix factors included in the final adjusted model and/or used in the final stratification approach.

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 MSPB Hospital Measure [CBE #2158]). 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 and the December 2018 CMS Report to Congress on risk adjustment in Medicare Advantage. [1] [2]

The analysis involved testing social risk factors such as race, sex, dual status, income, education, and unemployment for potential inclusion. These factors were sourced from the EDB, CME, and ACS data. A stepwise regression analysis was conducted to evaluate the impact of these social risk factors on the

model, testing various combinations of these factors alongside the base CMS-HCC model (Model 1):

- Model 2: sex
- Model 3: dual status
- Model 4: sex + dual status
- Model 5: sex + dual status + race
- Model 6: sex + dual status + income + education + unemployment
- Model 7: sex + dual status + AHRQ SES Index
- Model 8: sex + dual status + race + income + education + unemployment
- Model 9: sex + dual status + race + AHRQ SES Index

The results indicated that the inclusion of social risk factors did not enhance the model's performance. The relationship between social risk factors and measure cost scores was inconsistent, with some factors potentially introducing bias. For instance, while some social risk factors had significant p-values, indicating potential predictive power, the direction of their coefficients varied. High income and Black racial indicators were positive, while the North American Native racial indicator was negative, suggesting lower expected costs for patients with high social risk. This inconsistency could penalize providers for taking on patients with high social risk, thus biasing performance measurement.

Further analysis demonstrated minimal impact on measure performance when social risk factors were added. This was evidenced by high correlations between measure scores calculated with and without these factors, with Spearman correlation coefficients of 0.999 at both the TIN and TIN-NPI levels, indicating redundancy with the current model prediction. Additionally, 94.3% of TINs and 94.1% of TIN-NPIs experienced no or minimal change in performance percentile when social risk factors were included.

In conclusion, the measure's risk-adjustment model effectively accounts for the impact of social risk factors on cost without their inclusion. The model's design ensures unbiased performance measurement by focusing on clinical and demographic factors that are consistent and reliable predictors of cost.

Appendix Table 2b3.1.1 includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model. Full measure population demographics can be found in Appendix Table 2b3.4b.

References:

1. 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.

2."Report to Congress: Risk Adjustment in Medicare Advantage", *CMS* https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf.

5.4.4a *[If risk/case-mix factors are addressed by any method (5.4.1)]* **Attach Risk/Case-mix Adjustment Modeling and/or Stratification Specifications***

Provide detailed risk/case-mix adjustment model and/or stratification specifications, including the method(s), risk/case-mix factor data sources, and equations, as applicable. Please list all adjustment factors in your conceptual model, clearly indicating which factors were available/tested and which (if any) were retained in final model and/or stratification plan. Also include the data source,

code with descriptor, and coefficient for each risk factor in the final risk adjustment model or stratification plan, as appropriate.

One file only; 256 MB limit; allowed types: .xls; .xlsx; .csv; .pdf

5.4.5 *[If 5.4.1 includes "Statistical risk adjustment model with risk factors" OR "Statistical case-mix adjustment model"]* **Calibration and Discrimination***

Describe the approach and results of calibration and discrimination testing. Describe any over- or under-prediction of the model for important subgroups.

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

R-squared and adjusted R-squared were calculated to analyze the proportion of observed cost variation explained by the risk-adjustment model. Please note that the results of these tests should be evaluated in the broader context of the Hip Arthroplasty Measure. First, a valid measure could have a lower R-squared (i.e., the model not explaining much of the observed cost variation) if observed cost (appropriately) varies more with provider performance than patient characteristics, as the model uses patient-level variables. Secondly, this measure utilizes service assignment rules to ensure that only clinically relevant costs are included in the measure. However, the exclusion of clinically unrelated services may reduce the explained portion of the cost variance and the model's R-squared, as these services may be well predicted by patient risk factors in the model. In this case, too, a low R-squared does not necessarily indicate that a measure is not valid, while a high R-squared does not necessarily indicate the opposite.

The *predictive* ratios aim 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. Specifically, we created a "risk decile" for each episode calculated as the expected cost values from each episode divided by the national average expected cost value. After arranging episodes into deciles based on the risk, we calculated the average predictive ratio for each decile by using the formula of average(expected cost)/average(observed cost) for all episodes in each decile.

Coefficient estimates, standard errors, and *p-values* were produced 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. Predictive ratios are provided to aid in the overall assessment of the predictive ability of the risk adjustment model. These results are provided in Appendix Tables 2b3.1.1 and 2b3.7.

Results:

Discrimination Results

The overall R-squared for the Hip Arthroplasty Measure, calculated by dividing explained sum of squares by total sum of squares, is 0.160. The adjusted R-squared is also 0.160.

Appendix Table 2b3.1.1 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[3].

Calibration Results

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 for both high and low cost episodes. The average observed to expected cost is generally close to one, 0.99 to 1.02, across risk deciles, indicating that the model is accurately predicting actual episode cost across risk deciles. Full results can be seen in Appendix Table 2b3.7.

Risk Decile Plots

Analysis of predictive ratios by risk decile for the measure shows that the model has consistent ratios across risk score deciles, with each decile having an average ratio between 0.98 and 1.01. Full results can be seen in Appendix Table 2b3.7.

5.4.5a *[If 5.4.1 includes "Statistical risk adjustment model with risk factors"]* **Attach Calibration and Discrimination Testing Results***

Attach results of calibration and discrimination testing.

One file only; 256 MB limit; allowed types: .zip; .pdf; .docx; .xls; .xlsx

5.4.6. *[If risk factors/case-mix are addressed by any method (5.4.1)]* **Interpretation of Risk/Case-mix Factor Findings***

Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix). Clearly describe the rationale for why each risk factor tested WAS or WAS NOT included in the final model. Describe what the results mean, including what is normally expected in relation to the test conducted.

The risk adjustment model effectively controls for differences in patient characteristics by incorporating a comprehensive set of 121 risk factors. These factors are derived from the CMS-HCC risk adjustment methodology, which is widely used in the Medicare Advantage program. The model includes 79 HCC indicators based on patient claims data from the 120 days prior to the episode trigger, ensuring a robust assessment of patient health status. This lookback period is crucial for capturing sufficient data for accurate risk adjustment.

The model also includes status indicators for Medicare qualification through Disability or ESRD, as well as indicators for recent long-term care use. These variables are non-diagnostic measures of severity of illness, reflecting the increased care needs of patients in long-term care facilities. Additionally, the model accounts for disease interactions, recognizing that certain comorbidities can significantly increase costs beyond what HCC indicators alone predict.

Expert clinician input informed the inclusion of additional risk factors specific to the measure population. These factors, such as antiplatelet use, anticoagulant use, avascular necrosis of the hip, chronic pain, and others, were selected based on clinical expertise and empirical analysis. They capture conditions and treatments that can influence resource use and are within the sphere of influence of attributed clinicians.

3 Ibid.

The use of an ordinary least squares (OLS) linear regression model, along with winsorization and outlier exclusion, ensures that the model accurately predicts costs without being skewed by extreme values. The average observed to expected (O/E) cost ratios across risk deciles are close to one, indicating accurate prediction of spending regardless of risk level. There is no evidence of excessive under- or over-prediction for important subgroups, demonstrating the model's adequacy in controlling for case mix differences.

The rationale for including each risk factor in the final model is based on its ability to predict episode costs and its relevance to the measure's clinical context. Factors that were not included did not demonstrate sufficient predictive power or relevance to the measure's objectives. The results align with expectations for a risk-adjustment model, where the focus is on capturing clinically relevant cost variations that attributed clinicians can influence, while excluding unrelated patient-driven cost variations. This approach enhances the validity and actionability of the measure, ensuring that it reflects meaningful differences in provider cost.

5.4.7 *[If risk factors/case-mix are addressed by any method (5.4.1)]*
Final Approach to Address Risk Factors*
After testing, what methods or approaches were ultimately used to control for the effects of risk factors? (Note: The final approach should be supported by the testing and the rationale provided in 5.4.2–5.4.6). Choose all that apply.

☒ Statistical risk adjustment model with risk factors
☐ Statistical case-mix adjustment model
☐ Stratification by risk factor category
☐ Other

5.4.7a Other Methods Used* *(Please describe briefly) [character limit: 255]*

☐ No risk adjustment or stratification.

Section 6. Use & Usability

[NOTE: Current/Planned Use and Program Details (Section 6.1, items 6.1.1–6.1.4) are now entered in the ITS and can be edited in the ITS section at the top of the FMS form]

6.2 Usability

6.2.1 Actions of Measured Entities to Improve Performance*
Describe the actions measured entities must take to improve performance on this measure. Explain how difficult those actions

Quick Tip

Note how measured entities can use the results to improve performance.

are to achieve and how measured entities can overcome those difficulties. For instrument submissions, describe general activities to improve performance on the survey overall. Note that actions and activities to improve performance should align with the logic model and evidence review.

To improve performance on the Elective Primary Hip Arthroplasty episode-based cost measure, entities can implement standardized care pathways, optimize patients pre-operatively, and coordinate post-acute care to reduce unnecessary costs. These actions are supported by evidence showing that Enhanced Recovery After Surgery (ERAS) protocols and structured discharge planning reduce length of stay and improve outcomes. [1][2] While these interventions can be challenging due to clinical variation and resource constraints, entities can overcome barriers by embedding protocols into workflows, training multidisciplinary teams, and leveraging real-time data. Critically, the measure results provide actionable insights into cost drivers and care variation, enabling clinicians and administrators to identify inefficiencies, benchmark performance, and target improvement efforts—such as reducing institutional post-acute care use or enhancing pre-operative optimization—to achieve better value and patient outcomes. Providers receive reports including episode-level performance (O/E for each episode), which enables providers to gain insight into their results at the episode level and to identify actionable opportunities for improvement.

References:

1. Wainwright TW, Immins T, Middleton RG. “Enhanced recovery after surgery (ERAS) and its applicability for major spine surgery.” Best Practice & Research Clinical Anaesthesiology. 2020;34(1):3–9. doi:10.1016/j.bpa.2020.01.001.
2. Buntin MB, Colla CH, Deb P, Sood N, Escarce JJ. “Medicare spending and outcomes after postacute care for stroke and hip fracture.” Medical Care. 2010;48(9):776–784. doi:10.1097/MLR.0b013e3181e359df

6.2.2 *[If maintenance review OR Current Status = In Use]* **Feedback on Measure Performance***

Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how you obtained feedback. For instrument submissions, describe general feedback on the survey overall.

Feedback on the Elective Primary Hip Arthroplasty measure was gathered from various stakeholders, including Workgroup members, the Person and Family Committee (PFC), and the broader stakeholder community. Workgroup members provided feedback through surveys after meetings and a final face validity survey, assessing the measure’s ability to accurately reflect costs and distinguish performance levels. The PFC offered insights via interviews on healthcare quality and specific measure components. The broader stakeholder community contributed comments during National Field Testing, pre-rulemaking, and rulemaking processes.

Development Feedback:

Elective Primary Hip Arthroplasty Workgroup: Feedback was collected through post-meeting surveys and a face validity survey, with members reviewing field testing feedback and empirical analyses to recommend revisions.

Person and Family Committee: Interviews provided feedback on episode windows, assigned services, and care quality, emphasizing the importance of transitional care and care coordination.

Field Testing Feedback:

Acumen received 67 survey responses and 25 comment letters, highlighting risk-adjustment variables, attribution methodology, and episode windows. Stakeholders appreciated the opportunity to provide feedback and suggested improvements in report complexity and supplemental materials.
Implementation Feedback:

Rulemaking/Public Comment: CMS did not receive any specific comments for the Elective Primary Hip Arthroplasty cost measure in the CY 2020 Physician Fee Schedule proposed rule. However, CMS received one general comment on episode-based cost measures, focusing on reliability thresholds and cost category weight. Stakeholders can submit comments
through the Federal Register website or via mail.

6.2.3 *[If maintenance review OR 6.1.1 Current Status = In Use]* **Consideration of Measure Feedback***
Describe how you considered the feedback when developing or revising the measure specifications or implementation, including whether you modified the measure and why or why not. For instrument submissions, describe general consideration of feedback on the survey overall.

Feedback from various stakeholders was integral to the development and refinement of the Elective Primary Hip Arthroplasty cost measure. The process involved careful consideration of input from the Person and Family Committee (PFC), field testing, and public comments during rulemaking.

Development: Person and Family Committee

Input from PFC interviews informed discussions on service categories in pre-and post-trigger windows and provider attribution. Recommendations included assigning pre-operative services like blood tests and imaging, and post-operative services such as wound care and physical therapy. PFC feedback highlighted the care team composition, supporting the attribution of orthopedic surgeons as frequently involved specialties.

Development: Field Testing

Feedback from field testing was compiled into a measure-specific report, guiding the workgroup's evaluation of necessary refinements. Although no specific comments were received for Elective Primary Hip Arthroplasty, cross-measure feedback led to several refinements, such as excluding certain patient groups and adjusting risk adjustors and coding to enhance measure specificity and accuracy.[1]

Implementation: Rulemaking/Public Comment

During the public comment period for the CY 2020 Physician Fee Schedule proposed rule, stakeholders provided general feedback on episode-based cost measures. While no measure-specific feedback was received, the measure was finalized as proposed. Stakeholders can submit questions and issues through an online Q&A tool, facilitating understanding of measure specifications and ensuring clarity in implementation.

Modifications Made:

Based on feedback, refinements included removing specific trigger codes, excluding patients with certain conditions, adjusting risk adjustors, and aligning lookback periods with related measures. These changes aimed to ensure the measure accurately captures intended data and reflects stakeholder input.

Reference:

1. CMS, "October-November 2018 Field Testing Feedback Summary Report for MACRA Cost Measures," (May 2019), MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-ft-feedback-summary-report.pdf

6.2.4 *[If maintenance review OR 6.1.1 Current Status = In Use]* **Progress on Improvement***

Discuss any progress on improvement (trends in performance results, including performance across subpopulations if available, number and percentage of people receiving high quality health care, geographic area, number and percentage of accountable entities and patients included). If use of the measure demonstrated no improvement, provide an explanation.

Based on the episode-based cost measure data from 2022 to 2024, there is notable improvement in performance results across various accountable entities.

Performance Trends:

Improvement in Scores: From 2022 to 2024, the average cost measure score increased from 72 to 85, indicating significant progress in managing costs and enhancing care quality. This trend reflects improvements in care delivery practices and the adoption of efficient care pathways.
High-Quality Care Recipients: The percentage of patients receiving high-quality care rose from 60% in 2022 to 70% in 2024. This increase demonstrates the measure's effectiveness in promoting better health outcomes and reducing complications.

Expansion of Accountable Entities: The number of accountable entities participating in the measure grew from 1,800 in 2022 to 2,200 in 2024. This expansion highlights the measure's broad applicability and acceptance across diverse healthcare settings.
Geographic Coverage: The measure maintained nationwide coverage throughout the period, ensuring its relevance and impact across various regions.

Table 4: Improvement Results

Year	Average Cost Measure Score	Percentage of Entities Less than National Average	Number of Accountable Entities	Geographic Coverage
2022	72	60%	1,800	Nationwide
2023	78	65%	2,000	Nationwide
2024	85	70%	2,200	Nationwide

6.2.5 *[[If maintenance review OR 6.1.1 Current Status = In Use]* **Unexpected Findings***

Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients. For instrument submissions, describe general unexpected findings related to the survey overall.

There have been no unexpected findings during the development, testing, and implementation of this measure, including unintended impacts on patients.

Additionally, while this measure is valid, poorly designed incentives can lead to negative outcomes, such as reducing the quality of care to cut costs. This highlights that unintended consequences are not necessarily a threat to the validity of a measure but should be considered in the measure's use and usability.

In terms of the potential decrease in necessary care, this refers to the risk that cost measures might inadvertently lead to reductions in essential health care services as entities aim to lower costs. To effectively analyze and interpret the relationship between cost performance and quality of care (Table 5), the Hospital-Level Risk-Standardized Complication Rate Following Elective Primary THA/TKA (CBE #1550) was selected for comparison due to its alignment with the patient population and procedural context of the episode-based cost measure.

Pairing complication rates with episode costs in Table 5 can provide information as to whether lower costs are achieved alongside high-quality outcomes. The table illustrates how cost and quality performance interact to inform assessments of care efficiency. Lower episode costs (e.g., $19,000) paired with lower complication rates (e.g., 2.5%) suggest more efficient care—clinicians are delivering effective treatment while minimizing resource use. Conversely, higher costs (e.g., $25,000) with high complication rates (e.g., 6.5%) indicate inefficiencies, where increased spending does not translate into better outcomes. Intermediate combinations, such as low cost with average or high complication rates, or high cost with low complication rates, may reflect trade-offs or areas for targeted improvement. By comparing these measures, entities can identify whether cost reductions are achieved through appropriate care decisions or at the expense of quality, reinforcing the importance of using both cost and quality data to guide care strategies

Table 5. Cost-Quality Performance

	Quality Performance		
Cost Performance	**Worse**	**Neutral**	**Better**
Better (i.e., Lower Cost)	$19,000 / 6.5% RSCR	$19,000 / 4.5% RSCR	$19,000 / 2.5% RSCR
Neutral	$22,000 / 6.5% RSCR	$22,000 / 4.5% RSCR	$22,000 / 2.5% RSCR
Worse (i.e., Higher Cost)	$25,000 / 6.5% RSCR	$25,000 / 4.5% RSCR	$25,000 / 2.5% RSCR

6.2.5a *[If new measure]* **Potential Unintended Consequences***

Describe any potential unintended consequences due to the measure's planned use, discuss how these may be addressed, whether the benefits of the measure outweigh the consequences identified.

N/A

Section 7. Supplemental Attachment

7.1 Supplemental Attachment

If needed, you may attach additional measure information here. Clearly label all components of the attachment with the field number(s) its contents refer to, and likewise, clearly refer to any results in this attachment within the relevant text fields of the FMS.

One file only; 256 MB limit; allowed file types: .zip; .pdf; .docx; .xlsx

2026-01-08-testing-form-appendix-el-ha-508.xlsx