Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 3474

Measure Title: Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)

Measure Steward: Centers for Medicare & Medicaid Services

sp.18. Brief Description of Construction Logic :

This measure estimates hospital-level, risk-standardized payments for a 90-day episode of care for an elective primary THA/TKA. To this end, we construct a cohort of patients who underwent elective primary THA/TKA based on principal discharge diagnosis in Medicare administrative claims data. Specifically, we include Medicare FFS patients age 65 or older with a principal discharge diagnosis of elective primary THA/TKA procedure. We then apply six exclusion criteria as detailed in section sp.35. Once our cohort is finalized, we examine all payments for these patients (including co-pays, co-insurance, and deductibles) for the first 30 days after admission and THA/TKA-related claims for days 31-90 (Kim et al. 2014). We include payments for all care settings, supplies, and services, except Part D. We standardize payments across providers by removing geographic and policy adjustments that are unrelated to clinical care. These standardized payments were then assigned to the initial admitting hospital. As part of our model, we risk adjust these payments for patient comorbidities identified from outpatient and inpatient claims in the 12 months prior to the index admission as well as from the secondary diagnoses included in the index admission; we use a hierarchical generalized linear regression model to calculate a risk-standardized payment for each hospital included in the measure.

Reference

Kim N, Ott L, Lin Z, Zhou S, Keshawarz A, Spivack S, Xu X, George E, Parisi M, Reilly E, Zribi R, Suter L, Krumholz HM. Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 1.0) 2014 Measure Methodology Report. December 2014; Centers for Medicare & Medicaid Services (CMS). Available at: https://www.qualitynet.org/dcs/ContentServer?cid=1228774267858&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page

1b.01. Developer Rationale:

Previous Submission

This measure is intended to align with current quality measures to facilitate profiling hospital value (payments and quality). Given that THA/TKA is a procedure with substantial variability in costs of care aligning this payment measure with quality measures (e.g., RSCRS) will allow the assessment of hospital value. By evaluating their RSPs and RSCRs for THA/TKA, hospitals have an opportunity to consider actionable improvements and efficiencies on a

broader scale to impact the value of care. This measure provides transparency on the payments made for Medicare beneficiaries undergoing THA/TKA. Hospitals receive detailed information on how they compare with other institutions regarding the amount and venues of resources expended on patients. As such, the measure provides insight into hospitals that is not otherwise possible.

Current Submission

The rationale for this measure remains as stated above.

Measure Type: Cost and Resource use

sp.16. Data Source:

Claims

Other (specify)

Enrollment data

Enrollment Data

sp.08. Level of Analysis:

Facility

IF Endorsement Maintenance – Original Endorsement Date: 2019-06-10 01:14 PM

Most Recent Endorsement Date: 6/10/2019 1:14:51 PM

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Describe intent of the measure and its components, including the rationale (note any citations) for analyzing variation in resource use in this way.

[Response Begins]

Previous Submission

THA and TKA are common procedures among elderly patients with a substantial range in costs of care, likely due to different practice patterns (Sood et al. 2011). A hospital-level, episode-of-care payment measure for THA and TKA is informative for a number of reasons. First, it provides transparency into the differences in costs to Medicare for the same procedures across hospitals. Second, it allows hospitals to assess the payments for patients admitted to their institution relative to other hospitals and thus may incentivize hospitals to examine their own practices and coordinate with post-discharge providers to seek new efficiencies. Finally, when paired with existing outcome measures for THA/TKA patients, it identifies institutions that, after removing the effect of geography, policy adjustments, case mix, and dual-eligible status, demonstrate good patient outcomes at a low cost. Such hospitals may provide important examples of positive deviance from which other hospitals can learn.

The THA/TKA Payment measure is aligned with the THA/TKA Complication measure (NQF #1550). Other measures of quality include THA/TKA Readmission (NQF #1551) and the Hip/Knee Functional Status (in development) measures. Although other payment measures, such as Payment-Standardized Medicare Spending per Beneficiary (NQF #2158) and Episode Treatment Groups (ETG)-based Hip/Knee Replacement cost of care measure (NQF #1609), are endorsed by NQF, the THA/TKA Payment measure would have the benefit of being specific to THA/TKA and aligned with publicly reported THA/TKA outcome measures.

Reference

Sood N, Huckfeldt PJ, Escarce JJ, Grabowski DC, Newhouse JP. Medicare's bundled payment pilot for acute and postacute care: analysis and recommendations on where to begin. Health Aff (Millwood). Sep 2011;30(9):1708-1717.

Current submission

The intent of the THA/TKA Payment measure remains the same as described above.

[Response Ends]

1a.02. Provide data and/or summarize relevant literature to demonstrate the measure focus addresses a high-impact aspect of healthcare.

For example, affects large numbers, is a leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality.

[Response Begins]

Current Submission

In 2020, total national health expenditures were $4.1 trillion, representing 19.7% of the Gross Domestic Product (GDP). In 2020, total Medicare spending was $829.5 billion, or 20 percent of total national healthcare spending. Total healthcare spending is projected to hit $6.2 trillion by 2028, and Medicare spending is expected to increase by 7.6 percent each year through 2028.

Continued growth in Medicare spending underscores the need to focus on high-value care, which requires transparency of costs of care that is comparable across providers. The THA/TKA payment measure provides insight into the costs of practice patterns that occur during inpatient admission and immediately thereafter. The measure specifications are aligned with current quality-of-care measures so that the costs of care can be interpreted in the context of the health outcomes they deliver. In this way, the measure can facilitate the profiling of hospital value and encourage the most efficient delivery of high-quality care.

More than one million total joint arthroplasties are performed each year, with an 18.9% increase in annual volume between 2012 and 2017 and a 2% increase in the number of procedures per Medicare Beneficiary (Lopez et al., 2020). While CMS recently began allowing THA and TKA procedures in the outpatient setting, a substantial number of procedures remain in the inpatient setting, performed among older patients of higher acuity (Debbi et al. 2021).

Reference

Lopez CD, Boddapati V, Neuwirth AL, Shah RP, Cooper HJ, Geller JA. Hospital and surgeon Medicare reimbursement trends for total joint arthroplasty. Arthroplasty Today. 2020 Sept;6(3):437-444. https://doi.org/10.1016/j.artd.2020.04.013

Debbi EM, Mosich GM, Bendich I, Kapadia M, Ast MP, Westrich GH. Same-Day Discharge Total Hip and Knee Arthroplasty: Trends, Complications, and Readmission Rates. J Arthroplasty. 2022 Mar;37(3):444-448.e1. doi: 10.1016/j.arth.2021.11.023. Epub 2021 Nov 19. PMID: 34808278.

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

Previous Submission

This measure is intended to align with current quality measures to facilitate profiling hospital value (payments and quality). Given that THA/TKA is a procedure with substantial variability in costs of care aligning this payment measure with quality measures (e.g., RSCRS) will allow the assessment of hospital value. By evaluating their RSPs and RSCRs for THA/TKA, hospitals have an opportunity to consider actionable improvements and efficiencies on a broader scale to impact the value of care. This measure provides transparency on the payments made for Medicare beneficiaries undergoing THA/TKA. Hospitals receive detailed information on how they compare with other institutions regarding the amount and venues of resources expended on patients. As such, the measure provides insight into hospitals that is not otherwise possible.

Current Submission

The rationale for this measure remains as stated above.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

Previous Submission

We examine the distribution of hospital payment scores to demonstrate the variation, current and over time, in payment among measured hospitals. The results below indicate that the mean RSP decreased over the three-year period, from $23,248 between April 2012 and March 2013 to $22,840 between April 2014 and March 2015. The median hospital RSP in the combined three-year dataset was $22,408 (IQR $21,134 - $24,174).

Distribution of Hospital THA/TKA RSPs over Different Time Periods ($2014)

Results for each data year

//04/2012-03/2013//04/2013-03/2014//04/2014-03/2015//02-2012-03/2015

Characteristic

Number of Hospitals// 2,614 // 3,312 // 3,298 // 3,285 // 3,452

Number of Admissions// 142,361 // 295,222 // 305,983 // 892,455

Mean (SD)// 23,248 (2,535) // 23,454 (2,431) // 22,840 (2,356) // 21,733 (2,330) // 22,686 (2,655)

Range (min. – max.)// 16,421 – 35,123 //16,965 – 46,407 //14,660 – 49,154 //15,545 – 40,604 // 15,481 – 49,496

25th percentile// 21,473// 21,821 // 21,240 // 20,134 // 20,847

50th percentile// 23,120 // 23,248 // 22,660 // 21,529 // 22,408

75th percentile// 24,885// 24,880 // 24,185 // 23,106 // 24,174

Current submission

Below we provide the performance of the THA/TKA measure over time, each year within the three-year performance period.

Performance Period	04/2016-03/2017	04/2017-03/2018	04/2018-03/2019
Number of Hospitals	3,268	3,266	3,249
Number of Admissions	331,895	325,570	290,992
Mean (SD)	21,907 (2,406)	20,957 (2,205)	20,314 (1,999)
Range (Min-Max)	13,031-43,567	15,748-38,555	12,5967-37,324
Minimum	13,031	15,748	12,596
10th percentile	19,007	18,346	17,957
20th percentile	19,888	19,097	18,715
30th percentile	20,600	19,751	19,233
40th percentile	21,151	20,290	19,685
50th percentile	21,695	20,725	20,071
60th percentile	22,266	21,226	20,567

Performance Period	04/2016-03/2017	04/2017-03/2018	04/2018-03/2019
70th percentile	22,838	21,794	21,079
80th percentile	23,697	22,595	21,805
90th percentile	24,911	23,708	22,827
Maximum	43,567	38,555	37,324

Table A: Performance of the THA/TKA measure over a three-year performance period.

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

Previous Submission

Distribution of THA/TKA RSPs by Proportion of Dual Eligible Patients (for Hospitals with 25 or More Cases):

Dates of Data: April 2012 through March 2015

Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (≤3.8%) Dual Eligible patients//Hospitals with a high proportion (≥11.5%) Dual Eligible patients

Number of Measured Hospitals// 698// 697

Number of Patients// 324,481 patients in low-proportion hospitals// 103,705 patients in high-proportion hospitals

Maximum// 33,678// 45,741

90th percentile// 25,044// 28,277

75th percentile// 23,485//26,041

Median (50th percentile)// 21,925// 23,974

25th percentile// 20,729//22,341

10th percentile// 19,600//21,078

Minimum //16,037//16,889

Distribution of THA/TKA RSPs by Proportion of Patients with AHRQ SES Index Scores (for Hospitals with 25 or More Cases):

Dates of Data: April 2012 through March 2015

Data Source: Medicare FFS claims and the American Community Survey (2009-2013) data

Characteristic//Hospitals with a low proportion of patients below AHRQ SES index score of 42.7 (≤6.2%)// Hospitals with a high proportion of patients below AHRQ SES index score of 42.7 (≥23.6%)

Number of Measures Hospitals// 699// 697

Number of Patients// 262,511 patients in hospitals with a low proportion of patients below the AHRQ SES index score of 42.7// 130,235 patients in hospitals with a high proportion of patients below the AHRQ SES index score of 42.7

Maximum// 33,678//44,663

90th percentile// 25,703// 27,281

75th percentile// 23,618// 25,358

Median (50th percentile)// 22,110// 23,501

25th percentile// 20,710// 21,975

10th percentile// 19,499// 20,529

Minimum // 16,373// 16,889

Current submission

Below we provide measure scores for each social risk factor (DE, low AHRQ SES, stratified by quartiles of the facility-proportion of patients with each social risk factor.

Dual Eligibility

Quartile	Q1	Q2	Q3	Q4
Proportion of patients with low AHRQ SES	(0-1.7)	(1.8-3.4)	(3.4-6.9)	(6.9-95.5)[GB1] [GB2] [PD3]
# of Hospitals	687	689	687	688
Minimum	15,964	16,495	15,918	16,290
10th percentile	17,872	17,967	18,163	18,825
25th percentile	18,789	18,860	19,263	20,037
50th percentile	19,968	20,209	20,688	21,525
75th percentile	21,351	21,606	22,010	23,561
90th percentile	22,949	22,973	23,550	25,529
Maximum	30,738	38,518	37,765	40,898

Table B: Measure scores for the dual eligibility social risk factor.

Low AHRQ SES

Quartile	Q1	Q2	Q3	Q4
Proportion of patients with low AHRQ SES	(0-5.4)	(5.4-11.1)	(11.2-21.2)	(21.3-79.2)
# of Hospitals	686	689	684	685
Minimum	16,246	16,657	15,918	16,290
10th percentile	17,977	18,070	17,910	18,838

Quartile	Q1	Q2	Q3	Q4
25th percentile	18,725	19,054	19,154	20,095
50th percentile	19,951	20,255	20,478	21,612
75th percentile	21,335	21,606	21,935	23,550
90th percentile	22,816	22,959	23,790	25,698
Maximum	29,115	31,724	36,526	40,898

Table C: Measure scores for the low AHRQ SES social risk factor.

Race (Black)

Quartile	Q1	Q2	Q3	Q4
Proportion of patients with low AHRQ SES	(0-0.2)	(0.2-2.1)	(2.1-6.3)	(6.3-93.7)
# of Hospitals	687	688	688	688
Minimum	15,918	16,525	15,964	16,398
10th percentile	18,029	17,783	18,344	18,400
25th percentile	19,054	18,718	19,441	19,839
50th percentile	20,325	19,989	20,712	21,409
75th percentile	21,730	21,453	22,006	23,173
90th percentile	23,422	22,964	23,754	25,081
Maximum	32,366	37,765	31,724	40,898

Table D: Measure scores for the Race social risk factor.

Rurality

Quartile	Q1	Q2	Q3	Q4
Proportion of patients with low AHRQ SES	(0-0.5)	(0.5-2.4)	(2.5-9.2)	(9.2-100)
# of Hospitals	687	688	688	688
Minimum	16,398	15,964	16,477	15,918
10th percentile	18,685	18,169	17,833	17,977
25th percentile	19,934	19,153	18,897	18,992
50th percentile	21,374	20,427	20,130	20,340
75th percentile	23,238	21,932	21,790	21,603
90th percentile	25,158	23,334	23,551	23,219
Maximum	31,256	29,583	40,898	38,518

Table E: Measure scores for the rurality social risk factor.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

N/A

[Response Ends]

Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eCQM should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

---

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]

Yes

[Yes Please Explain]

Please see section spma.02 for detailed updates to the measure specifications.

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]

The THA/TKA Payment measure was initially NQF endorsed in June of 2019. The following measure changes have been implemented since its endorsement:

- When CMS implemented the measure in 2017, CMS did not include adjustment for dual eligibility in the measure specifications. Therefore, the version of the measure that is undergoing endorsement maintenance in the Fall 2022 cycle does not include adjustment for dual eligibility.
    - Rationale: The THA/TKA Payment measure is publicly reported together with the THA/TKA complications measure on Care Compare. The THA/TKA Complications measure is NQF-endorsed without adjustment for dual eligibility and therefore the paired payment measure needed to align with the complications measure.
- Updated the ICD-10 code-based specifications used in the measures. Specifically, we:
    - incorporated the code changes that occurred in the ICD-10-CM/PCS code set releases into the cohort definitions the risk models, and the complication definitions used by the THA/TKA payment measure
    - applied a modified version of the V22 CMS-HCC crosswalk that is maintained by RTI International to the risk models;
    - made additional code specification changes prompted by clinical expert review including code frequency monitoring, and neighboring code searches. For example, ICD-10-CM code I21.9, Acute myocardial infarction, unspecified, was identified through a “neighboring code search”

(found near existing code I21.4, Non-ST elevation (N-STEMI) myocardial infarction) and determined through clinical review to be a code which meets measure intent. As a result, it was added to the AMI cohort inclusion list.

- Rationale: Updated versions of the ICD-10-CM/PCS and CMS-HCC crosswalk were released. Revisions to the measure specifications were warranted to accommodate these updates.

- the addition of ICD-10-PCS codes to the specifications that define 'Periprosthetic Joint Infection/Wound Infection and Other Wound Complications' THA/TKA-related payments. Please see the codes listed in the data dictionary, tab "6. HKPay Outcome Inclusion"
  - Rationale: During routine measure maintenance, our analyses showed the addition of these clinically relevant codes contributed to an increase in the THA/TKA national observed complication rate. Findings demonstrated an increase of approximately 0.5 percent (from 2.42 percent to 2.93 percent) in the THA/TKA national observed complication rate when evaluated for the FY 2021 performance period (April 1, 2016 through March 30, 2019). These findings suggest that the expanded outcome will allow the updated THA/TKA Complication measure to capture a more complete outcome.
- Description of the complication category 'Periprosthetic Joint Infection/Wound Infection' was changed to 'Periprosthetic Joint Infection/Wound Infection and Other Wound Complications'.
  - Rationale: Description was revised to reflect that conditions beyond periprosthetic joint infection/wound infection, such as wound disruption, are captured under this category; conditions that our clinical experts consider to be relevant and consistent with the intent of the THA/TKA payment measure.

[Response Ends]

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

This measure estimates hospital-level, risk-standardized payments for an elective primary total THA/TKA episode of care, starting with an inpatient admission to a short-term acute care facility and extending 90 days post admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older.

[Response Ends]

sp.04. Type of resource use measure (Select the most relevant).

[Response Begins]

Per episode

[Response Ends]

sp.05. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Musculoskeletal: Joint Surgery

Musculoskeletal: Osteoarthritis

Musculoskeletal: Rheumatoid Arthritis

Surgery: Orthopedic

[Response Ends]

sp.06. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Care Coordination: Readmissions

Care Coordination: Transitions of Care

Safety: Complications

[Response Ends]

sp.07. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Elderly (Age >= 65)

[Response Ends]

sp.08. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*

- *Population: Population*

[Response Begins]

Facility

[Response Ends]

sp.09. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Inpatient/Hospital

[Response Ends]

sp.10. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]

https://qualitynet.org/inpatient/measures/payment

[Response Ends]

sp.13. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3474_3474_2020 THA_TKA Payment_final-508.xlsx

sp.17. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Enrollment data

[Response Ends]

sp.18. Briefly describe the measure's construction logic.

If applicable, summarize the general approach or methodology to the measure construction. This is most relevant to measures that are part of or rely on the execution of a measure system or applies to multiple measures.

[Response Begins]

This measure estimates hospital-level, risk-standardized payments for a 90-day episode of care for an elective primary THA/TKA. To this end, we construct a cohort of patients who underwent elective primary THA/TKA based on principal discharge diagnosis in Medicare administrative claims data. Specifically, we include Medicare FFS patients age 65 or older with a principal discharge diagnosis of elective primary THA/TKA procedure. We then apply six exclusion criteria as detailed in section sp.35. Once our cohort is finalized, we examine all payments for these patients (including co-pays, co-insurance, and deductibles) for the first 30 days after admission and THA/TKA-related claims for days 31-90 (Kim et al. 2014). We include payments for all care settings, supplies, and services, except Part D. We standardize payments across providers by removing geographic and policy adjustments that are unrelated to clinical care. These standardized payments were then assigned to the initial admitting hospital. As part of our model, we risk adjust these payments for patient comorbidities identified from outpatient and inpatient claims in the 12 months prior to the index admission as well as from the secondary diagnoses included in the index admission; we use a hierarchical generalized linear regression model to calculate a risk-standardized payment for each hospital included in the measure.

Reference

Kim N, Ott L, Lin Z, Zhou S, Keshawarz A, Spivack S, Xu X, George E, Parisi M, Reilly E, Zribi R, Suter L, Krumholz HM. Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 1.0) 2014 Measure Methodology Report. December 2014; Centers for Medicare & Medicaid Services (CMS). Available at: https://www.qualitynet.org/dcs/ContentServer?cid=1228774267858&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page

[Response Ends]

sp.19. Detail logic steps used to cluster, group or assign claims beyond those associated with the measure's clinical logic.

[Response Begins]

To construct the measure, we use Medicare administrative claims data. These data contain claims for all care settings, supplies, and services as outlined in Section sp.25 (except for Part D). Claim payment data are organized by the setting, supply, or service in which they were rendered. Standard Medicare payment rates were assigned to each service based on claim type, facility type, and place of service codes. These payments are then summed by individual patients. To create a hospital-level measure, we aggregate the payments for all eligible patients at each hospital.

[Response Ends]

sp.20. Provide additional information about the construction logic, if needed.

Attach supplemental documentation (Save file as: Construction_Logic). All fields of the submission form that are supplemented within the attachment must include a summary of important information included in the attachment and its intended purpose, including any references to page numbers, tables, text, etc.

[Response Begins]

Attachment

[Response Ends]

Attachment: 3474_3474_HK Payment Original Methodology Report_(1)-508.pdf

sp.21. Indicate how the measure is specified to handle concurrency of clinical events, measure redundancy or overlap, and disease interactions.

[Response Begins]

Detail the method used for identifying concurrent clinical events, how to manage them, and provide the rationale for this methodology.

[Detail the method used for identifying concurrent clinical events, how to manage them, and provide the rationale for this methodology. Please Explain]

This measure examines payments for a 90-day episode of care beginning with an admission for elective THA/TKA and extending 90-days post admission. We determine if a patient has an elective THA/TKA by identifying the primary procedure code in the administrative data without indication of pathological or traumatic fracture. If a patient has any other procedure code or has a code for traumatic or pathological fracture of the lower extremity, this admission is not considered an index admission for this measure. Therefore, the concurrency of clinical events is not an issue when determining what triggers the episode of care. Once an episode is triggered, however, we include payments for all care settings, except Part D. The model risk adjusts for comorbidities listed in outpatient and inpatient claims in the 12 months prior to the index admission as well as the secondary diagnoses included in the index admission that are not considered complications of care.

[Response Ends]

sp.22. Indicate how the measure is specified to handle complementary services.

Complementary services are those associated with, but ancillary to, the primary health services (as applied to a population, an admission, an encounter, or a procedure) that are the measure's focus. One common example of a complementary service is a diagnostic test as a follow-up to a primary care visit. In your response, describe how complementary services are accounted for in the construction logic and cost totals of your measure.

[Response Begins]

Detail how complementary services have been linked to the measure and provide rationale for this methodology.

[Detail how complementary services have been linked to the measure and provide rationale for this methodology. Please Explain]

The measure includes payments for all care settings, except Part D, that occur during the 90-day window. If a claim for a complimentary service was filed in the study window, then it would be included in the measure.

[Response Ends]

sp.23. Clinical hierarchies

[Response Begins]

Detail the hierarchy of codes or condition groups used and provide rationale for this methodology.

[Detail the hierarchy of codes or condition groups used and provide rationale for this methodology. Please Explain]

The measure uses a risk-adjustment model based on Condition Categories (CCs) as opposed to Medicare Advantage Hierarchical Condition Categories (HCCs). We used CCs because they provide detailed descriptions about comorbidities that may influence care decisions that affect payment for THA/TKA without assigning hierarchy. This allows conditions that would be ranked lower in the hierarchy to be considered for risk adjustment if they are medically and statistically relevant.

[Response Ends]

sp.24. Indicate how the measure is specified to handle missing data.

[Response Begins]

We do not provide measure specifications for missing data. Detail your rationale

[We do not provide measure specifications for missing data. Detail your rationale Please Explain]

We do not impute missing data for any of the variables included in the measure. However, if a hospitalization is missing a DRG or DRG weight we exclude it as an index admission.

[Response Ends]

sp.25. Select the resource use service categories (units).

Select all categories that apply.

[Response Begins]

Inpatient services: Inpatient facility services

Inpatient services: Evaluation and management

Inpatient services: Procedures and surgeries

Inpatient services: Imaging and diagnostic

Inpatient services: Lab services

Inpatient services: Admissions/discharges

Inpatient services: Labor (hours, FTE, etc.)

Other inpatient services

Ambulatory services: Outpatient facility services

Ambulatory services: Emergency Department

Ambulatory services: Pharmacy

Ambulatory services: Evaluation and management

Ambulatory services: Procedures and surgeries

Ambulatory services: Imaging and diagnostic

Ambulatory services: Lab services

Ambulatory services: Labor (hours, FTE, etc.)

Other ambulatory services

Durable Medical Equipment (DME)

Other services not listed

[Response Ends]

sp.26. For each of the resource use service categories selected above, provide the rationale for their selection and detail the method or algorithms to identify resource units, including codes, logic and definitions.

[Response Begins]

To estimate payments for a 90-day episode of care for THA/TKA we included payments for all care settings, services, and supplies, except Part D (for more details, see Kim et al. 2014, p.19-31). We did not include Part D since a substantial proportion of Medicare beneficiaries are not enrolled in Part D and there is variation in

enrollment status across and within states. Including payments for Part D services would thus bias payments upwards for hospitals with high Part D enrollment. By following patients through an episode of care for THA/TKA, CMS and hospitals can gain key insights into the drivers of payments and how practice patterns vary across providers.

Specifically, for day 0 through day 30 (where day 0 = day of admission for the index hospitalization), we include payments for the following care settings in the measure:

- Inpatient hospital facility and physician
- Outpatient hospital facility and physician
- Skilled nursing facility and physician
- Hospice facility and physician
- Home health facility and physician
- Inpatient psychiatric facility and physician
- Inpatient rehab facility and physician
- Long-term care hospital facility
- Clinical labs facility and physician
- Comprehensive outpatient rehab facility and physician Outpatient rehab facility and physician
- Renal dialysis facility and physician
- Community mental health centers facility and physician DME/POS/PEN
- Observation stay facility
- Part B drugs
- Ambulance and ambulance physician
- Emergency department facility and physician office
- Federally qualified health centers facility and physician Rural health clinics facility and physician
- Ambulatory surgical centers facility and physician

We also include physician payments for the following care settings:

- Indian health service free-stand facility
- Indian health service provider facility
- Tribal free-standing facility
- Tribal facility
- Military treatment facility
- Independent clinic
- State or local health clinic
- Mass immunization center
- Walk-in retail health clinic
- Urgent care facility
- Unassigned
- Pharmacy
- School
- Homeless Shelter
- Prison

- Group Home
- Mobile Unit
- Temporary Lodging
- Birthing Center
- Intermediary Care/Mentally Retarded
- Residential Substance Abuse
- Psychiatric Residential Facility
- Non-Residential Substance Abuse
- Other Physician
- Other carrier claims with HCPCS codes P9603 or P9604

For days 31 through 90, we include payments for the following care settings or services, which we have defined as THA/TKA-related payments:

- Durable Medical Equipment (DME)
- Inpatient rehabilitation
- Outpatient rehabilitation
- Skilled Nursing Facilities (SNFs)
- Home health
- Outpatient hospital (joint manipulation procedures under anesthesia)
- Staged or repeat admission for single-site surgeries within 90 days of index admission
- Readmissions for complications as defined in the CMS THA/TKA Complication measure (wound/joint infection or mechanical complication) (Suter et al., 2014).

In order to assign claims to care settings, we examine the place of service code for physician claims and a combination of claim type and facility type codes to determine the facility in which care was provided. Depending on the specific facility and physician codes we standardize payments differently. Information on how we standardize claims can be found in section S.9.6.

References

Kim N, Ott L, Lin Z, Zhou S, Keshawarz A, Spivack S, Xu X, George E, Parisi M, Reilly E, Zribi R, Suter L, Krumholz HM. Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 1.0) 2014 Measure Methodology Report. December 2014; Centers for Medicare & Medicaid Services (CMS). Available at: https://www.qualitynet.org/dcs/ContentServer?cid=1228774267858&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page

Suter LG, Parzynski CS, et al. 2016 Measure Updates and Specifications: Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) Risk-Standardized Complication Measure (Version 2.0). March 2016. Available at: https://www.qualitynet.org/dcs/ContentServer?cid=1228774789978&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page

[Response Ends]

sp.27. If needed, provide supplemental resource use service category specifications in either URL (preferred) or as an attachment (Save file as S.7.8a_RU_Service_Categories):

[Response Begins]

URL

[URL Please Explain]

Kim N, Ott L, Lin Z, Zhou S, Keshawarz A, Spivack S, Xu X, George E, Parisi M, Reilly E, Zribi R, Suter L, Krumholz HM. Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 1.0) 2014 Measure Methodology Report. December 2014; Centers for Medicare & Medicaid Services (CMS). Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228774267858

[Response Ends]

sp.28. Briefly describe your clinical logic approach.

Indicate the clinical topic area, whether or not you account for comorbid and interactions, clinical hierarchies, clinical severity levels and concurrency of clinical events.

[Response Begins]

THA and TKA are common elective procedures among the elderly with substantial variability in payments due to different practice patterns (Sood et al. 2011). Quality measures for THA/TKA, such as the 90-day risk-standardized complication rate (RSCR) following THA/TKA, are already publicly reported. In the context of its publicly reported quality measures, THA/TKA is an ideal procedure in which to assess payments for Medicare patients and relative hospital value. Therefore, we created a measure of payments for a 90-day episode of care for THA/TKA that could be aligned with CMS's 90-day THA/TKA complication measure. This will allow CMS to assess the value of care provided for these episodes.

The measure uses Condition Categories (CCs) to adjust for patient case mix across hospitals. Details of our risk-adjustment strategy can be found in our technical report at *https://www.qualitynet.org/dcs/ContentServer?cid=1228774267858&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page*

This measure is for patients who are admitted for an elective primary THA/TKA. We identify these patients by examining the procedure codes in the administrative data. If a patient has a procedure code of any other procedure, this admission is not considered as an index admission. Therefore, the concurrency of clinical events is not applicable for this measure. However, the model does risk adjust for comorbidities listed in outpatient and inpatient claims in the 12 months prior to the index admission as well as the secondary diagnoses included in the index admission that are not considered complications of care.

Reference

Sood N, Huckfeldt PJ, Escarce JJ, Grabowski DC, Newhouse JP. Medicare's bundled payment pilot for acute and postacute care: analysis and recommendations on where to begin. Health Aff (Millwood). Sep 2011;30(9):1708-1717.

[Response Ends]

sp.29. Detail the clinical logic of the measure.

Detail any clustering and the assignment of codes, including the grouping methodology, the assignment algorithm, and relevant codes for these methodologies.

[Response Begins]

We focused on a 90-day episode of care triggered by admission for an elective primary THA/TKA as identified using ICD-10 procedure codes described in the data dictionary. The measure includes admissions for Medicare FFS beneficiaries aged 65 years and older not transferred from another acute care facility, undergoing elective primary THA or TKA. The cohort does not include admissions for primary THA or TKA if the patients had fractures, partial replacements, revisions, resurfacing, mechanical complications, malignant neoplasms, or device removals since procedures with these conditions have distinctly different risks and outcomes. A full list of codes used to identify these conditions is provided in the Measure Methodology Report.

Elective primary THA/TKA procedures are defined as those THA/TKA procedures without any of the following:

- Fracture of the femur, hip, or pelvic fractures coded in the principal or secondary discharge diagnosis fields of the index admission;
- A concurrent partial hip or knee arthroplasty procedure;
- A concurrent revision, resurfacing, or implanted device/prosthesis removal procedure;
- Mechanical complication coded in the principal discharge diagnosis field of the index admission; or,
- Malignant neoplasm of the pelvis, sacrum, coccyx, lower limbs, or bone/bone marrow or a disseminated malignant neoplasm coded in the principal discharge diagnosis field;
- Transfer from another acute care facility for the THA/TKA .

We assigned all payments for the episode of care to the hospital that originally admitted the patient

[Response Ends]

sp.30. Provide evidence to support the clinical logic described above.

Describe the rationale, citing evidence to support the grouping of clinical conditions in the measurement population(s) and the intent of the measure (as described in above).

[Response Begins]

The intent of the measure is to estimate payments for a 90-day episode of care for an elective primary THA/TKA in order to gain insight into drivers of payment within and across hospitals. To profile hospital payments fairly, the measure fulfills the following criteria. First, we standardize payments to remove geography and policy adjustments to isolate payment differences related to the clinical care of patients undergoing THA/TKA. Second, we adjust for hospital case mix. Third, we align the THA/TKA payment measure specifications with the nationally reported 90-day THA/TKA risk-standardized complication measure to identify practice patterns that may be expensive without conferring a quality benefit across an episode of care for THA/TKA. Lastly, we focus on specific procedures to provide the most meaningful feedback to hospitals and incentivize targeted improvements in care (Kim et al. 2014).

Reference

Kim N, Ott L, Lin Z, Zhou S, Keshawarz A, Spivack S, Xu X, George E, Parisi M, Reilly E, Zribi R, Suter L, Krumholz HM. Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 1.0) 2014 Measure Methodology Report. December 2014; Centers for Medicare & Medicaid Services (CMS). Available at:

https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228774267858

[Response Ends]

sp.31. Attach supplemental documentation of the clinical logic (Save file as: Clinical_Logic), if needed.

All fields of the submission form that are supplemented within the attachment must include a summary of important information included in the attachment and its intended purpose, including any references to page numbers, tables, text, etc.

[Response Begins]

No supplemental documentation provided.

[Response Ends]

sp.32. Detail the measure's trigger and end mechanisms, and provide a rationale for this methodology.

[Response Begins]

When considering hospital payments, we focused on an "episode of care" triggered by an admission for elective primary THA/TKA for several key reasons. First, THA and TKA procedures require ongoing post-discharge care. Second, a fixed 90-day timeframe incentivizes hospitals to optimize post-discharge care. Third, mechanical complications and wound or joint infections may present after 30 days. Fourth, the 90-day post-admission timeframe is consistent with CMS's THA/TKA complication measure, which captures specific complications up to 90 days after admission. Finally, a 90-day window was consistent with the timeframe recommended by members of our Technical Expert Panel (TEP). Based on these factors, we chose a follow-up period of 90 days that includes all payments for the initial 30 days of the episode, and payments defined as "related" to the index procedure for days 31 through 90. Related payments are defined in detail in section S.7.8.

[Response Ends]

sp.33. Describe how clinical severity levels are handled in your measure's specifications.

[Response Begins]

Detail the method used for assigning severity level and provide rationale for this methodology

[Detail the method used for assigning severity level and provide rationale for this methodology Please Explain]

The measure uses administrative claims data to risk adjust for patient comorbidities but does not include adjustments for clinical severity. Instead, the measure inclusion criteria eliminate the most severe cases. For example, the measure does not include procedures on patients with: femur, hip, or pelvic fractures; a concurrent revision, resurfacing, or implanted device/prosthesis removal procedure, or patients with malignant neoplasm of the pelvis, sacrum, coccyx, lower limbs, or bone/bone marrow or a disseminated malignant neoplasm (see section sp.35 for additional non-inclusions). Furthermore, our team has demonstrated the validity of claims-based measures for profiling hospitals for a number of prior measures by comparing either the measure results or the individual data elements to medical records. CMS validated the six NQF-endorsed claims-based measures currently in public reporting (i.e., mortality and readmission measures for AMI, HF, and pneumonia) with models that used medical record-abstracted data for risk adjustment. Specifically, claims model validation was conducted by building comparable models using abstracted medical record data for risk adjustment for heart failure patients (National Heart Failure data), AMI patients (Cooperative Cardiovascular Project data) and pneumonia patients (National Pneumonia Project dataset). When both models were applied to the same patient population, the hospital risk-standardized mortality and readmission rates estimated using the claims-based risk-adjustment models had a high level of agreement with the results based on the medical record model. In addition, a nationally convened TEP supported the face validity of the NQF-endorsed THA/TKA complication measure, which uses a similar risk-adjustment approach. Together, these factors support the use of claims-based models for public reporting.

[Response Ends]

sp.34. Describe how co-morbidities and interactions are handled in your measure's specifications.

[Response Begins]

Detail the treatment of co-morbidities and disease interactions and provide rationale for this methodology

[Detail the treatment of co-morbidities and disease interactions and provide rationale for this methodology Please Explain]

The goal of risk adjustment for this measure is to account for patient age, gender, and comorbid conditions that are clinically relevant and have strong relationships with the outcome, while illuminating important payment differences between hospitals.

Comorbidities that are included in risk adjustment are identified in administrative claims during the 12 months prior to and including the index admission. To assemble the more than 70,000 ICD-10 codes into clinically coherent variables for risk adjustment, the measure employs publicly available CMS condition categories (CCs) to group ICD-10 codes into CCs, and selects comorbidities on the basis of both clinical relevance and statistical significance [Pope, 2000].

Reference

Pope G, Ellis R, Ash A, et al. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review. 2000;21(3):26.

[Response Ends]

sp.35. Indicate whether you use initial inclusion and exclusion criteria.

[Response Begins]

Detail initial inclusion/exclusion criteria and data preparation steps (related to clinical exclusions, claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim, exclusion of ESRD patients)

[Detail initial inclusion/exclusion criteria and data preparation steps (related to clinical exclusions, claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim, exclusion of ESRD patients) Please Explain]

Inclusion Criteria

1. **Having a qualifying elective primary THA/TKA procedure during the index admission:**

Rationale: Elective primary THA or TKA is the procedure targeted for measurement (Table D.4.1). Elective primary THA/TKA procedures are defined as those THA/TKA procedures *without* any of the following:

- **Femur, hip, or pelvic fractures coded in the principal or secondary discharge diagnosis fields of the index admission**

  Rationale: Patients with fractures have higher mortality, complication, and readmission rates and the procedures are not elective.

- **A concurrent partial hip arthroplasty procedure**

  Rationale: Partial arthroplasty procedures are primarily done for hip fractures and are typically performed on patients who are older, frailer, and have more comorbid conditions. Partial knee arthroplasty procedures are not distinguished by ICD-9-CM codes and are therefore currently captured by the THA/TKA payment measure.

- **A concurrent revision, resurfacing, or implanted device/prosthesis removal procedure**

  Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication, and readmission rates. Resurfacing procedures are a different type of procedure involving only the joint's articular surface. Resurfacing

procedures are typically performed on younger, healthier patients. Elective procedures performed on patients undergoing removal of implanted device/prostheses procedures may be more complicated.

- **Mechanical complication coded in the principal discharge diagnosis field of the index admission**

  Rationale: A complication coded as the principal discharge diagnosis suggests the procedure was more likely the result of a previous procedure. These patients may require more technically complex arthroplasty procedures and may be at increased risk for complications, particularly mechanical complications, and readmission.

- **Malignant neoplasm of the pelvis, sacrum, coccyx, lower limbs, or bone/bone marrow or a disseminated malignant neoplasm coded in the principal discharge diagnosis field**

  Rationale**:** Patients with these malignant neoplasms are at increased risk for complications and readmission, and the procedure may not be elective.

2. Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of admission, and enrolled in Part A and Part B during the index hospitalization

Rationale: Claims data are consistently available only for Medicare FFS beneficiaries. The 12-month prior enrollment criterion ensures that patients were Medicare FFS beneficiaries and that their comorbidities are captured from claims for risk adjustment. Additionally, Medicare Part A is required at the time of admission to ensure that no Medicare Advantage patients are included in the measure. Medicare Part B is required to ensure coverage across all care settings.

3. Aged 65 or over

Rationale: Medicare patients younger than 65 usually qualify for the program due to severe disability. They are not included in the measure because they are considered to be too clinically distinct from Medicare patients 65 and over.

4. Not transferred from another acute care facility

Rationale: Hospitalizations in which a patient was transferred in from another acute care facility are not included because it is the hospital where the patient was initially admitted that initiates patient management and is responsible for making critical acute care decisions (including the decision to transfer and where to transfer). In addition, the measure is designed to capture elective procedures, which would be unlikely for a transferred patient.

Exclusion Criteria for THA/TKA Measure

1. **Discharged against medical advice (AMA)**

   Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge

2. **Incomplete administrative data in the 90 days following the index admission if discharged alive.**

   Rationale: This is necessary in order to identify the outcome (payments) in the sample over our analytic period.

3. **Transferred to a federal hospital**

   Rationale: We do not have claims data for these hospitals; therefore, including these patients would systematically underestimate payments.

4. **With more than two THA/TKA procedure codes during the index admission**

   Rationale: Although clinically possible, it is highly unlikely that patients would receive more than two elective THA/TKA procedures in one hospitalization, which may reflect a coding error.

5. **Not matched to admission in the THA/TKA complication measure**

   Rationale: As part of the current data processing, we match our index THA/TKA admissions to the THA/TKA complication cohort to obtain the risk-adjustment variables. Patients are excluded if they cannot be matched between the THA/TKA payment and THA/TKA complication cohorts.

6. **Missing index DRG weight where provider received no payment**

   Rationale: With neither DRG weight or payment data, we cannot calculate a payment for the patient's index admission; this would make the entire episode of care appear significantly less expensive.

For patients with more than one eligible admission for a THA/TKA in a given year, only one admission is randomly selected to include in the cohort as an index hip/knee hospitalization. After exclusions #1-6 are applied, the measure randomly selects one hospitalization per patient per year for inclusion in the cohort so that each episode of care is mutually independent. Additional admissions within that year are excluded. Similarly, for the three-year combined data, when index admissions occur during the transition between measure reporting periods (March and April-June of each year) and both are randomly selected for inclusion in the measure, the measure includes only the March admission. April-June admissions within the 90-day outcome window of the March admission are excluded to avoid assigning payments for the same claims to two admissions.

CMS FFS beneficiaries with an index hospitalization to an acute care non-federal hospital are included in the measure if they have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission to ensure a full year of administrative data for risk adjustment.

The episode of care begins with an admission for an elective primary THA or TKA to a short-term acute care hospital. The hospital that initially admits the patient is assigned all payments that occur during the episode of care. This includes payments for patients who are subsequently transferred to another hospital for further care of the index THA or TKA. Claims from an emergency department do not trigger the episode of care because CMS does not classify emergency department care as an inpatient admission. If a patient is transferred from an emergency department to another hospital and then subsequently admitted, the episode of care begins with the inpatient admission at the receiving hospital.

ICD-10-CM procedure codes are listed in the attached data dictionary

[Response Ends]

sp.36. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

Statistical risk model

[Response Ends]

sp.37. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.38. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

Not applicable. This measure is not stratified.

[Response Ends]

sp.39. Select a costing method.

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

[Response Begins]

Standardized pricing

[Standardized pricing Please Explain]

Medicare pays for health care services using a number of different payment systems that are generally organized by delivery setting. These payment systems consider not only the products the Medicare patient is buying in each setting, but also the characteristics of the care provider, the extent to which the same product may be furnished in different settings, and the market circumstances that affect providers' costs. Payment amounts within each payment system are usually updated annually (for example, the IPPS) with some fee schedules having quarterly updates (for example, Durable Medical Equipment/Prosthetics Orthotics and Supplies [DME/POS]). Information on CMS reimbursement rates for each care setting are made publicly available through either Final Rules published in the Federal Register or fee schedules provided on the CMS website. A summary of Medicare's reimbursement system for most care settings is publicly available at the Medicare Payment Advisory Committee (MedPAC) website. Below, we describe the key features of these payment systems and how we used these CMS payment algorithms to determine an episode-of-care payment for THA/TKA that isolates clinical care decisions. Please see Appendix C in the technical report for a full description of how we standardize payments for each care setting: https://www.qualitynet.org/inpatient/measures/payment/methodology.

[Response Ends]

sp.40. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Continuous variable, e.g. average

[Response Ends]

sp.41. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Score within a defined interval

[Response Ends]

sp.42. Detail steps to estimate measure score.

[Response Begins]

The RSP is calculated as the ratio of "predicted" payment to "expected" payment, multiplied by the national unadjusted average payment for the episode of care. The expected payment for each hospital is estimated using its patient mix and the average of the hospital-specific intercepts. The predicted payment for each hospital is estimated given the same patient mix but an estimated hospital-specific intercept. Operationally, the expected

payment for each hospital is obtained by summing the expected payments for all patients in the hospital. The expected payment for each patient is calculated via the hierarchical model by applying the subsequent estimated regression coefficients to the observed patient characteristics and adding the average of the hospital-specific intercepts. The predicted payment for each hospital is calculated by summing the predicted payments for all patients in the hospital. The predicted payment for each patient is calculated through the hierarchical model by applying the estimated regression coefficients to the patient characteristics observed and adding the hospital-specific intercept.

[Response Ends]

sp.43. Describe your approach to discriminating results.

Detail methods for discriminating differences (reporting with descriptive statistics--e.g., distribution, confidence intervals).

[Response Begins]

To categorize hospital payments, CMS estimates each hospital's RSP and the corresponding 95% interval estimate. CMS assigns hospitals to a payment category by comparing each hospital's RSP interval estimate to the national mean payment. Comparative payments for hospitals with 25 or more eligible cases are classified as follows:

- "No Different than the National Payment" if the 95% interval estimate surrounding the hospital's RSP includes the national mean payment.
- "Greater than the National Payment" if the entire 95% interval estimate surrounding the hospital's RSP is higher than the national mean payment.
- "Less than the National Payment" if the entire 95% interval estimate surrounding the hospital's RSP is lower than the national mean payment.

If a hospital has fewer than 25 eligible cases for a measure, CMS assigns the hospital to a separate category: "Number of Cases Too Small." This category is used when the number of cases is too small (fewer than 25) to reliably estimate the hospital's RSP. If a hospital has fewer than 25 eligible cases, the hospital's RSP and interval estimate will not be reported for the measure.

[Response Ends]

sp.44. Detail attribution approach.

Detail the attribution rules used for attributing resources/costs to providers (e.g., a proportion of total measure cost or frequency of visits during the measure's measurement period) and provide rationale for this methodology.

[Response Begins]

The measure attributes payments incurred during the 90-day episode to the original admitting hospital. We assign these payments to the admitting hospital because decisions made at the admitting hospital affect payments for care in the inpatient setting as well as the post-discharge and recovery periods for THA/TKA arthroplasty. Furthermore, attributing payments for a continuous episode of care to admitting hospitals may reveal practice variations in the full care of the illness that can result in increased payments. For patients who are admitted and then transferred to another hospital during the original index admission, we assign all payments to the original admitting hospital since this hospital is responsible for the initial care decisions and the decision to transfer the patient.

[Response Ends]

sp.45. Identify and define the peer group.

Detail how the peer group is identified and provide a rationale for this methodology.

[Response Begins]

As part of the measure methodology we compare payments for a hospital with the expected payment amounts for an average hospital with the same case mix. While we include all hospitals when estimating the risk-adjustment model, we do not calculate RSPs for hospitals with fewer than 25 THA/TKA procedures, since estimates for hospitals with fewer procedures are less reliable and CMS's past approach to public reporting has been not to report these results.

[Response Ends]

sp.46. Detail the sample size requirements for reporting measure results.

[Response Begins]

In order for hospitals to be publicly reported, they must have at least 25 index THA/TKA admissions during the measurement period.

[Response Ends]

sp.47. Define benchmarking and comparative estimates.

Detail steps to produce benchmarking and comparative estimates and provide rationale for this methodology.

[Response Begins]

Comparative estimates are provided by classifying hospitals as less than average, no different than average, or greater than average payment depending on the span of their confidence interval in comparison with the national average payment amount (i.e., the benchmark). To categorize hospital payments, we estimate each hospital's RSP and the corresponding 95% interval estimate. As with all estimates, there is a degree of uncertainty associated with the RSP. The interval estimate is a range of probable values around the RSP that characterizes the amount of uncertainty associated with the estimate. A 95% interval estimate indicates that there is 95% probability that the true value of the RSP lies between the lower limit and the upper limit of the interval. In an effort to provide fair comparisons, we provide three categories (less than, no different than, or greater than the national average payment amount), which allows for conservative discrimination of hospital RSPs.

[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.03. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]

Yes - Additional risk adjustment analysis is included

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.

- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Medicare Enrollment Data; The American Community Survey, Rural Urban Commuting Codes Dataset

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

The datasets/data sources we used in testing include Medicare administrative claims data, Medicare enrollment database (EDB), Medicare fee schedules, Federal Register Final Rules for Medicare PPS systems and payment policies, and CMS published wage index data.

To assess socioeconomic factors, we used census as well as Medicare enrollment data. Dual eligibility was obtained through enrollment data. The Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score was obtained using the American Community Survey (ACS), 2013-2017. Data on rurality was obtained from The Rural-Urban Commuting Area Codes 2019 dataset,

The datasets used varies by testing type; see Section 2a.07 for details.

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

Dates vary by dataset; see section 2a.07 for details.

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Facility

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years and older are included.

The number of measured entities (hospitals) varies by testing type; see Section 2a.07 for details.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

The number of admissions/patients varies by testing type; see Section 2a.07 for details.

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

The datasets, dates, number of measured hospitals, and number of admissions used for each type of testing are shown in Table 1 (**for the current and previous submissions**).

Current submission

For updated analytical testing for this measure, we used three years of Medicare administrative claims data (July 2016 – June 2019). The dataset also included administrative data on each patient for the 12 months prior to the index admission and the 30 days following it. The dataset contained inpatient and facility outpatient claims and Medicare enrollment database (EDB) data. The datasets also contain price-standardized payments for Medicare patients across all Medicare settings, services, and supplies (that is, inpatient, outpatient, SNF, home health agency, hospice, physician/clinical laboratory/ambulance services, and durable medical equipment, prosthetics/orthotics, and supplies). The CMS Standardization Methodology for Allowed Amount for 2006 through 2019 was applied to the claims to calculate the measures.

Refer to the original methodology report for further descriptions of these data sources.

Federal Register Final Rules for Medicare Prospective Payment Systems and Payment Policies: Certain data necessary to calculate payments (e.g. annual base payments and conversion factors, DRG weights, wage indexes, and average length of stay) were taken from applicable Federal Register Final Rules. CMS-published Wage Index Data Wage index data not published in Federal Register Final Rules (such as the wage index data for Renal Dialysis Facilities) were obtained via the CMS website.

CMS-published Wage Index Data Wage index data not published in Federal Register Final Rules (such as the wage index data for Renal Dialysis Facilities) were obtained via the CMS website.

Dataset	Applicable Section	Description of Dataset
Dataset 1: Original Development and Validation Datasets (Medicare Fee-For-Service Administrative Claims Data)	Reliability Testing (Section 2a.01) Validity Testing (Section 2b.01) Risk Adjustment/Stratification (Section 2b.13) Statistical Risk Model Discrimination Statistics (Section 2b.17) Statistical Risk Model Calibration Statistics (Section 2b.18)	Chronic Conditions Data Warehouse (CCW) data Dates of Data: July 1, 2010 – June 30, 2012 Sample A1: random 50% sample of July 2011-June 2012 Sample A2: remaining 50% of July 2011-June 2012 Sample B: full July 2010-June 2011 sample Number of Index Admissions: Sample A1: 142,361 Sample A2: 142,360 Sample B: 286,750 Patient Descriptive Characteristics: Sample A1: average age= 74.5, %male= 36.1 Sample A2: average age= 74.4, %male= 35.9 Sample B: average age= 74.5, %male= 36.0 Number of Measured Entities: Sample A1: 3,257 Sample A2: 3,246 Sample B: 3,318
Dataset 2: 2016 Public Reporting Cohort	Reliability Testing (Section 2a.01) Validity Testing (Section 2b.01) Testing of Measure Exclusion (Section 2b.07) Statistical Risk Model Discrimination Statistics (Section 2b.17) Statistical Risk Model Calibration Statistics (Section 2b.18) Meaningful Differences (Section 2b.23)	Dates of Data: April 1, 2012 – March 31, 2015 Number of Index Admissions: 887,061 Patient Descriptive Characteristics: average age= 74.1, %male= 36.9 Number of Measured Entities: 3,481

Dataset	Applicable Section	Description of Dataset
Fall 2022 EM Testing Dataset (Medicare Fee-For-Service Administrative Claims Data) (July 1, 2016 – June 30, 2019)	Reliability Testing (Section 2a.01) Validity Testing (Section 2b.01) Testing of Measure Exclusion (Section 2b.07) Risk Adjustment/Stratification (Section 2b.13) Statistical Risk Model Discrimination Statistics (Section 2b.17) Meaningful Differences (Section 2b.23)	Dates of Data: July 2016-June 2019. Number of admissions = 987,227. Number of measured hospitals: 3.417 This cohort was randomly split into two halves. First half of the split sample: -Number of Admissions: 473,371 -Number of measured hospitals: 3,362 Patient Descriptive Characteristics: -Mean age = 73.9 -%Male = 37.1 Second half of split sample: -Number of Admissions: 475,086 -Number of measured hospitals: 3,417 Patient Descriptive Characteristics: -Mean age = 73.9 -%Male = 37.2
The American Community Survey (ACS)	Risk adjustment/Stratification for Outcome or Resource Use Measures (Section 2b.13)	Dates of Data: 2013-2017 We used the AHRQ SES Index score derived from the American Community Survey (2013-2017) to study the association between the payment outcome and social risk factors. The AHRQ SES index score is based on beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
Rural Urban Commuting Area Codes (new for 2022 submission)	Risk adjustment/Stratification for Outcome or Resource Use Measures (Section 2b.13)	Dates of Data: 2010 We used information on patients' rurality to study the association between rural geographic location and 30-day measure outcomes.
Master Beneficiary Summary File (MBSF)	Risk adjustment/Stratification for Outcome or Resource Use Measures (Section 2b.13)	Dates of Data: July 2016 – June 2019 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the 30-day measure outcome and dual-eligible status.

Table 1: Dataset descriptions

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

Please see section 2a.11 for the conceptual model for social risk factors' potential impact on the outcome. For testing, we were limited to social risk factors that that are available and can be linked to claims data. The NQF-convened Technical Expert Panel that considered risk-adjustment for social risk factors recognized that testing and risk adjustment for social risk may be constrained by data limitations and data collection burden (NQF, 2021).

Below we list the variables that are available within, or that can be linked directly, to Medicare administrative claims data used for this measure. In selecting variables for analysis, our intent was to be responsive to the National Quality Forum (NQF) guidelines for measure developers and the findings of work funded by the IMPACT Act (NQF, 2021, ASPE 2016, ASPE 2020). Our approach was to examine patient-level indicators that are reliably available for all Medicare beneficiaries and linkable to claims data and to select those that have established validity.

Potential pathways for socioeconomic status and race variables' effects are described in Section 2a.11. This section is limited to a description of the variables.

The socioeconomic status and race variables that we examined, further described below, are:

- Dual-eligible status (updated analyses provided for endorsement maintenance)
- Black race (newly provided for Fall 2022 endorsement maintenance)
- Rurality (newly provided for Fall 2022 endorsement maintenance)
- AHRQ-validated SES Index score (updated analyses provided for endorsement maintenance)

Medicaid dual-eligible status (Medicaid-Medicare dual, Medicare only)

Data source: Medicare Enrollment Database

The dual-status patient-level variable provides a reliably-obtained indication of patients with low income/assets and high health care spending. Following guidance from ASPE and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable (ASPE 2016, ASPE 2020). We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous outcome. However, the threshold for over 65-year-old Medicare patients is valuable, as it considers both income and assets and is consistently applied across states for the older population. We acknowledge that it is important to test a wider variety of social risk factors including key variables such as education and poverty level; therefore, we also tested a validated composite (AHRQ SES Index – see below) based on census data linked to as small a geographic unit as possible.

AHRQ-validated SES Index score: neighborhood socioeconomic factors as proxies for patient-level socioeconomic status

Data source: Enrollment database and Census data (American Community Survey)

The American Community Survey (ACS) provides several social risk indicators that are available at the ZIP code level and can be linked directly to Medicare claims at the 9-digit ZIP code level. We used the Agency for Healthcare Research and Quality (AHRQ)-validated composite index of socioeconomic status which has been used and tested among Medicare beneficiaries (Bonito, 2008). This index is a composite of seven different variables found in the Census data which may capture social risk better than any single variable. The variables are: (1) median household income, (2) percentage of persons living below the federal poverty level, (3) percentage of persons who are aged

>16 years and in the labor force but not employed, (4) median value of owner-occupied homes, (5) percentage of persons aged >25 years who completed at least a 12th grade education, (6) percentage of persons aged >25 years who completed at least four years of college, and (7) percentage of households that average one or more persons per room.

We selected the AHRQ SES Index because it is a well-validated variable that describes the average socioeconomic status of people living in defined geographic areas (Bonito et al, 2008). Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of social risk factors. We ultimately did not include the ADI at the time, partly due to the fact that the coefficients used to derive ADI had not been updated for many years. Recently, the coefficients for ADI have been updated and therefore we compared the ADI with the AHRQ SES Index and found them to be highly correlated.

In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low socioeconomic status neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes.

Black race (Black, other)

Data source: Medicare Enrollment Database

We used the Medicare enrollment database to identify the patient-level race variable (Black) that we used in these analyses. The Black variable has been shown to be reliable for use in this dataset (Waldo, 2004).

Rurality

Data source: Rural Commuting Area Codes

To evaluate the impact of rurality on the measure, we used the following dataset: The Rural-Urban Commuting Area Codes dataset, used to assign a patients' admission as rural vs. not rural, using pre-established coding categories.

Please see section 2b.13 for the conceptual model and literature summary related to social risk factors.

References

Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;2.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation. Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2017.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/pdf-report/second-impact-report-to-congressAccessed January 4, 2021.

Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk within Healthcare Performance Measurement. National Quality Forum; 2021. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=96087

National Academies of Sciences, Engineering, and Medicine (NASEM); Accounting for Social Risk Factors in Medicare Payment: Data. Washington DC: National Academies Press; 2016.

Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. Health Care Financing Review. 2004;26(2). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194866/

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

Previous submission

Measure Score Reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. In line with this thinking, our approach to assessing reliability was to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we took a "test-retest" approach in which hospital performance was measured once using a random subset of patients, then measured again using a second random subset exclusive of the first. Finally, we compared the agreement between the two resulting performance measures across hospitals (Rousson et al., 2002).

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half of patients. Thus, each hospital was measured twice, but each measurement was made using an entirely distinct set of patients. To the extent that the calculated measures of these two samples agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979). Specifically, we used the **Dataset 2** split sample and calculated the RSPs for each hospital for each sample. The agreement of the two RSPs was quantified for hospitals using the ICC (2,1) as defined by Shrout and Fleiss (1979).

Using two independent samples provides a stringent estimate of the measure's reliability, compared with using two random but potentially overlapping samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical generalized linear regression, and a known property of hierarchical generalize linear regression models is that smaller volume hospitals contribute less ´signal´, a split sample using a single measurement period would introduce extra noise. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman 1910, Brown 1910). We used this formula to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

Second, we estimate the facility-level reliability. While test re-test reliability is the most relevant metric from the perspective of overall measure reliability, it is also meaningful to consider the separate notion of "unit" reliability, that is, the reliability with which individual units (here, hospitals) are measured. This is because the reliability of any one facility's measure score will vary depending on the number of procedures performed. Facilities with more procedural volume will tend to have more reliable scores, while facilities with less procedural volume will tend to have less reliable scores. Therefore, we also use the formula presented by Adams and colleagues (2010) to calculate facility-level reliability as an additional, complementary metric.

Current submission

For this Fall 2022 endorsement maintenance submission, we have provided updating testing results using both the split-sample and signal-to-noise reliability approaches and the Fall 2022 Endorsement Maintenance Dataset.

References

Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021. Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Previous submission

Measure Score Reliability

As a metric of agreement, we calculated the ICC (Shrout and Fleiss 1979). To calculate the ICC, we used Dataset 2. The agreement between the two independent assessments of each hospital was 0.931.

Facility-level Reliability

The median reliability score of 0.938 is considered high.

Taken together, these results indicate that there is sufficient reliability in the measure score.

Current submission

Measure score reliability (*split-sample*): In total, 948,457 admissions were included in the analysis, using 3 years of data. After randomly splitting the sample into two halves, there were 475,086 admissions from 3,417 hospitals in one half and 473,371 admissions from 3,362 hospitals in the other half. As a metric of agreement, we calculated the ICC for all hospitals.

Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the RSP for each hospital was 0.889.

Facility-level reliability (signal-to-noise): Table 2 below shows the distribution of signal-to-noise reliability (Adams et al., 2010) for the 2,751 hospitals with at least 25 cases, using three years of performance data. Median signal-to-noise reliability was 0.948.

Mean	Std	Min	5th Percentile	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	95th Percentile	Max
0.915	0.084	0.37	0.742	0.79	0.876	0.948	0.977	0.988	0.991	0.998

Table 2: Distribution of signal-to-noise reliability (Fall 2022 Endorsement Maintenance Dataset)

Reference

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86:3420-3428.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Previous submission

The ICC score demonstrates very strong agreement across samples, indicating that the measure score is reliable.

Current submission

This measure's reliability is sufficiently high, as shown by the results of both split sample (0.889) and facility-level signa-to noise-reliability results (median, 0.948).

[Response Ends]

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

Previous submission

Measure score validity is demonstrated by systematic assessment of measure face validity by a Technical Expert Panel (TEP) of national experts and stakeholder organizations. Additionally, we have performed prior validity testing on our other claims-based measures, and applied established measure development guidelines.

Measure Score Validity: Face Validity as Determined by TEP

To systematically assess face validity, we surveyed the Technical Expert Panel (TEP) and asked each member to rate the following statement using a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5= Moderately Agree, and 6=Strongly Agree): "The Hip/Knee Payment measure as specified will provide a valid assessment of the relative costs of a 90-day hip/knee arthroplasty episode of care for Medicare patients admitted to a given hospital?"

Measure Score Validity: Validity as Assessed by External Groups

To increase transparency and to gain broader input into the measure, we obtained expert and stakeholder input via three mechanisms: regular consultations with an expert health economist, convening a national TEP, and a 30-day public comment period.

The health economist with whom we consulted had years of experience in economic analysis and working with claims data. We worked with the consultant to address key issues surrounding measure development, including detailed discussions regarding the appropriate cohort for inclusion in the measure. Having regular meetings with a consultant provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In alignment with the CMS Measure Management System (MMS), we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals who represent a range of perspectives including clinicians, consumers, and purchasers, as well as individuals with experience in quality improvement, performance measurement, and healthcare disparities. We convened two structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We made modifications to the measure based on TEP feedback.

Following completion of the measure, we solicited public comment on the measure through CMS, and the public comments were posted publicly. The resulting input was taken into consideration during the final stages of measure development.

Measure Score Validity: Validity Indicated by Established Measure Development Guidelines

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcomes measures, CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz, Brindis, et al. 2006; NQF 2010).

Current submission

Empiric Validity

Stewards of NQF-endorsed measures going through the re-endorsement process are required to demonstrate external validity testing at the time of maintenance review, or if this is not possible, justify the use of face validity only. To meet this requirement for the THA/TKA payment measure, we identified and assessed the measure's correlation with other measures that target the same domain (payment or utilization) for the same or similar populations. After literature review and consultations with measure experts in the field, there were very few measures identified. Given that challenge, we selected the hospital Medicare Spending per Beneficiary (MSPB) measure for comparison. We report an unweighted Pearson's correlation coefficient for this analysis.

The hospital Medicare Spending per Beneficiary (MSPB) measure is a risk-adjusted, price-standardized measure that assesses Medicare Part A and Part B payments for services provided to Medicare beneficiaries for episodes that spanning from three days prior to an inpatient hospital admission through 30 days after discharge. More information about the hospital MSPB measure can be found here:
https://qualitynet.cms.gov/inpatient/measures/mspb

Because the MSPB measure assesses payments for all Medicare FFS patients for all conditions during the measurement period, and the THA/TKA payment measure is focused on a narrow set of procedures, we predicted that THA/TKA payment measure scores would be weakly-to-moderately, positively correlated with MSPB measure scores.

As additional evidence of measure score validity, we also present a measure of internal validity of the outcome by examining the distribution of payment types across the quartiles of risk-standardized payments. Our expectation is that detailed level observed payments would be greater in hospitals with higher risk-standardized payments.

References

Bratzler DW, Normand SL, Wang Y, O'Donnell WJ, Metersky M, Han LF, Rapp MT, Krumholz HM. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. Public Library of Science One. 2011 Apr 12;6(4):e17401.

Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y-F, Herrin J, Chen J, Federer JJ, Mattera JA, Wang Y, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation: Cardiovascular Quality and Outcomes. 2008 Sep;1(1):29-37.

Krumholz HM, Brindis RG, Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. 2006;113(3):456-462.

Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand SL. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation: Cardiovascular Quality and Outcomes. 2011 Mar 1;4(2):243-52.

Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation. 2006a Apr 4;113(13):1683-92.

Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation. 2006b Apr 4;113(13):1693-70.

Lindenauer PK, Normand SL, Drye EE, Lin Z, Goodrich K, Desai MM, Bratzler DW, O'Donnell WJ, Metersky ML, Krumholz HM. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. Journal of Hospital Medicine. 2011 Mar;6(3):142-50.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/Publications/2011/07/National_Voluntary_Consensus_Standards_for_Patient_Outcomes_2009.aspx Accessed August 19, 2010.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

Previous submission

Validity was assessed by the TEP. The TEP provided input on the model to strengthen the measure and supported the final measure. Among the thirteen of fifteen TEP members who provided a response, two responded "Somewhat Agree," six responded "Moderately Agree," and five reported "Strongly Agree" that this measure

provides a valid assessment of payments for Medicare patients for a 90-day THA/TKA episode of care, removing policy adjustments unrelated to care decisions, risk adjusting based upon case mix, and providing CMS with a tool that it can use to compare payments across hospitals and identify hospitals with notably higher and lower payments.

Current submission

External Empiric Validity

Correlations with Medicare Spending per Beneficiary

The results of the correlation analyses described in section 2b.02 are shown below in Figure 2. As expected, the THA/TKA payment measure score was positively correlated with the Medicare Spending Per Beneficiary (MSPB) measure, with a correlation coefficient of 0.480 meaning that higher spending across all Medicare FFS beneficiaries correlated with higher spending on patients undergoing THA/TKA procedures.

Figure 2 shows the box-whisker plots of the THA/TKA risk-standardized payment (RSP) within each quartile of the MSPB measure score. The blue circles represent the mean RSPs of THA/TKA payment score quartiles.

Figure 2. Box-whisker plots of HF payment RSPs within each quartile of the Medicare Savings Per Beneficiary (MSPB) measure score

Disposition of Payments

Below we show the disposition of payments for the observed outcomes (inpatient and post-acute care), within quartiles of the provider RSP (Table 3).

Description	1st Quartile of RSPs	2nd Quartile of RSPs	3rd Quartile of RSPs	4th Quartile of RSPs
Total Number of Patients in Each RSP Quartile	376,903	266,416	202,942	102,196
Total Observed Episode Payment per Patient ($)	18,835	20,877	22,640	25,877
Index Inpatient Payment/Patient ($)	14,271	14,362	14,568	14,930
Index Inpatient Facility Payment/Patient ($)	12,288	12,321	12,461	12,841
Index Inpatient Physician Payment/Patient ($)	1,983	2,041	2,108	2,089
Patient with PAC (%)	97.6	98.9	99.3	99.3
PAC Payment/Patient ($)	4,580	6,540	8,097	11,017

Table 3: Subcategories of observed payments within the THA/TKA Payment Quartiles of Provider RSPs

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Previous submission

These results demonstrate TEP agreement with overall face validity of the measure score as specified. Measure validity is also ensured through the processes employed during development, including regular expert and clinical input, and modeling methodologies with demonstrated validity in claims-based measures.

Current submission

The validity of the THA/TKA Payment measure is supported by three types of evidence: face validity results derived from a systematic survey of a Technical Expert Panel (TEP), empiric validity demonstrated by correlations, and internal consistency.

The validity of the THA/TKA Payment measure is supported by face validity as indicated by the Technical Expert Panel (TEP) vote. There was unanimous TEP support for the face validity of the measure: 8 of 8 TEP members strongly, mostly, or somewhat agreed with the validity statement.

The validity of the measure is further supported by the empiric evidence that shows a correlation in the expected strength and direction with a related and valid payment measure.

Finally, the observed payment breakdowns appropriately align with the distribution of the provider-level risk-standardized payments.

[Response Ends]

2b.05. Indicate whether the measure uses exclusions.

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.06. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

Previous submission

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure as well as alignment with the Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) (NQF #1550) with which this payment measure is paired in public reporting.

To ascertain the impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 2**). These exclusions are consistent with similar NQF-endorsed outcome measures.

Current submission

We examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (Fall 2022 Endorsement Maintenance Dataset). These exclusions are consistent with similar NQF-endorsed outcome measures.

[Response Ends]

2b.07. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

Previous submission

Exclusion	N	%	Median % (IQR)	Range	Number of Hospitals
1. Incomplete administrative data	9,412	0.99	1.21 (0.72 - 1.96)	0.12 - 100.0	2,284
2. Discharged against medical advice (AMA)	125	0.01	0.27 (0.14 - 0.56)	0.01 - 25.00	120
3. Transferred to federal hospitals	43	0.005	0.29 (0.14 - 0.56)	0.04 - 11.11	39
4. More than two THA/TKA procedure codes during the index admission	0	0.00	N/A	N/A	N/A
5. THA/TKA stays with missing payment data	10,549	1.11	1.27 (0.72 - 2.26)	0.09 - 50.00	2,076
6. Patients without an index admission DRG weight and provider received no payment	6,142	0.65	1.11 (0.41 - 2.56)	0.04 - 50.00	1,164

Table 4: Measure exclusions

Current submission

Exclusion	N	%	Distribution across hospitals (N=2,789: Min, 25th, 50th, 75th percentile, Max)
1. Incomplete administrative data	10,437	1.03	(0.00, 0.38, 0.93, 1.64, 11.4)
2. Discharged against medical advice (AMA)	156	0.02	(0.00, 0.00, 0.00, 0.00, 2.94)
3. Transferred to a federal hospital	50	0.00	(0.00, 0.00, 0.00, 0.00, 3.03)
4. More than two THA/TKA procedure codes during the index admission	0	0	n/a
5. Not matched to admission in the THA/TKA complication measure	10,732	1.06	(0.00, 0.00, 0.68, 1.49, 99.4)
6. Patients without an index admission DRG weight and provider received no payment	0	0	n/a

Table 5: Measure exclusions (Fall 2022 Endorsement Maintenance Dataset)

After exclusions #1-6 are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent. This step excludes an additional 40,953 admissions, or about 4% of the cohort. For the three-year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. July admissions within the 30-day outcome window of the June admission are excluded to avoid assigning payments for the same claims to two admissions. There were 1,433 admissions in July, representing 0.14% of the cohort.

[Response Ends]

2b.08. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

Previous submission:

Exclusion 1 (Incomplete administrative data) accounts for 0.99% of all index admissions excluded from the initial index cohort. This exclusion is necessary for valid calculation of the measure. It affects very few patients.

Exclusion 2 (patients who are discharged AMA) accounts for 0.01% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 3 (patients who are transferred to a federal hospital) accounts for less than <0.005% of all index procedures excluded from the initial index cohort. This exclusion is intended to remove admissions from the cohort for patients transferred to federal hospitals. It is necessary for valid calculation of the measure. Very few patients are affected by this exclusion.

Exclusion 4 (patients with more than two THA/TKA procedure codes during the index hospitalization) accounts for <0.00% of all index procedures excluded from the initial index cohort. Although clinically possible, it is highly unlikely that patients would receive more than two elective THA/TKA procedures in one hospitalization, which may reflect a coding error.

Exclusion 5 (THA/TKA stays with missing payment data) affects 1.11% of all index admissions excluded from the initial index cohort. It is necessary for valid calculation of the measure. This exclusion affects very few patients.

Exclusion 6 (Patients without an index admission DRG weight and provider received no payment) affects 0.65% of all index admissions excluded from the initial index cohort. It is necessary for valid calculation of the measure. This exclusion affects very few patients.

Current submission:

Updated exclusion analyses show results similar to the Previous submission, for each exclusion. For each exclusion only small proportion of patients (0 to about 1.0%) are excluded.

[Response Ends]

2b.09. Check all methods used to address risk factors.

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

57

[Response Ends]

2b.10. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

See the attached data dictionary that includes risk factor codes with descriptors.

The goal of risk adjustment for this measure is to account for patient and procedure characteristics and comorbid conditions that are clinically relevant and have strong relationships with the outcome, while illuminating important quality differences between hospitals.

Comorbidities for inclusion in risk adjustment are identified in administrative claims during the 12 months prior to and including the index admission. To assemble the more than 15,000 ICD-9 diagnosis codes into clinically coherent variables for risk adjustment, the measure employs the publicly available CMS CCs to group ICD-9 diagnosis codes into CCs and selects comorbidities on the basis of clinical relevance and statistical significance (Pope et. Al, 2000).

The measure does not adjust for the patient's admission source or discharge disposition (for example, a skilled nursing facility) because these factors are associated with the structure of the health care system and the different care patterns the measure seeks to illuminate. Because hospitals should not be held to different standards of care based on the demographics of their patients, the measure does not adjust for socioeconomic status (SES), race, or ethnicity. Variation in payments associated with these characteristics may indicate differences in the care provided to vulnerable populations and adjusting for these factors would obscure these disparities. The measure does not adjust for hospital characteristics either (for example, teaching status), since this would hold different types of hospitals to different standards, and because such characteristics may exist on a causal pathway to the outcome rather than act as confounders. This approach was consistent with NQF guidance at the time of measure development (Measure Evaluation Criteria, 2011).

Complications of Index Hospitalization

Complications occurring during index hospitalization are not comorbid illnesses and may reflect hospital care; therefore, they should not be used for risk adjustment. Although adverse events during hospitalization may increase the payments for a THA/TKA episode of care, including them as covariates in a risk-adjusted model could

obscure payment differentials related to the quality of care delivered by hospitals. CORE previously reviewed every CMS-CC and identified those which, if they only occur during the index hospitalization and not in the 12 months prior, would be considered potential complications rather than comorbidities. For example, fluid, electrolyte or base disorders; sepsis; and acute liver failure are CMS-CCs that could potentially be complications of care (see Appendix A of the original methodology report).

Case Mix Adjustment: Choice of Functional Form

As is typical with data for healthcare payments, our dependent variable – total payment for a THA/TKA 90-day episode of care – is both right-skewed and leptokurtotic (skewness = 2.5; kurtosis = 13.1). This remains the case after Winsorization of patient-level 90-day payments at the 99.9% of the full distribution. To address estimation problems that can arise with non-normally distributed data, we employed the algorithm suggested by Manning & Mullahy. Using this algorithm and Sample A1, we compared several alternative models in order to determine the best estimation approach. Based on these assessments, we chose to estimate a generalized linear model with a log link and an inverse Gaussian distribution.

References

Manning W, Mullahy J. Estimating Log Models: To Transform or not to Transform? J Health Econ. Jul 2001;20(4):461-494.

Measure Evaluation Criteria. 2011; http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx Accessed September 26, 2012.

Pope G, Ellis R, Ash A, et al. Principal Inpatient Diagnostic Cost Group Model for Medicare Risk Adjustment. Health care financing review. Spring 2000;21(3):93-118.

[Response Ends]

2b.11. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

[Response Ends]

2b.12. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature

Internal data analysis

[Response Ends]

2b.13. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

Previous submission

The goal of risk adjustment for this measure is to make fair comparisons among hospitals by accounting for differences in their patient case mix, including age and comorbid conditions that are clinically relevant and have strong relationships with the outcome, while illuminating important payment differences between hospitals. The measure adjusts for case mix differences based on the comorbidities of the patient at the time of index admission. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in risk adjustment. Although they may increase the risk of mortality and complications, including them as covariates in risk adjustment could attenuate the measure's ability to characterize payments influenced by care delivered by hospitals.

The candidate variables for the model are derived from secondary diagnoses of the index hospital stay (excluding potential complications), as well as inpatient claims data, outpatient hospital claims data, and Part B claims for physician, radiology, and laboratory services during the 12 months prior to the index hospital stay.

For candidate variable selection using the development sample A1 (random 50% of Dataset 1), we started with the 189 Condition Categories (CCs). We used the ICD-9-to-CC assignment map, which is maintained by RTI and posted at www.qualitynet.org. To select candidate variables, a team of clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the THA/TKA payment outcome (e.g., attention deficit disorder, female infertility). Clinically relevant CCs were selected as candidate variables; some of these CCs were combined into clinically coherent groups. Other adjustment variables included age, gender, location of procedure (THA or TKA), and procedure type (single, bilateral, or staged).

To inform variable selection, we performed a modified approach to stepwise generalized linear model regression. We used sample A1 to create 1,000 bootstrap samples. For each sample, we ran a generalized linear model that included all candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with THA/TKA payment (at the $p<0.05$ level) in the 1,000 bootstrap samples (e.g., 70% would mean that the candidate variable was significant at $p<0.05$ in 70% of the bootstrap samples).

The working group reviewed these results and decided to retain all risk-adjustment variables above a 90% cutoff (i.e., to retain variables that were significant at the $p<0.05$ level in at least 90% of the bootstrap samples). We chose the 90% cutoff because variables above this threshold demonstrated a relatively strong association with THA/TKA payment and were clinically relevant.

The final set of risk-adjustment variables are listed in the data dictionary (see tabs 3 and 4).

For the current submission we have updated the literature summary that was presented in the previous submission:

Social Risk Factors

We selected relevant social risk factors based on the literature and the availability of the variables in current datasets. See section 2a.08 for a description of the variables that we considered and analyzed. Below we describe mechanisms by which social risk factors may influence episode-of-care payment. Our conceptualization of the mechanisms by which social risk factors affect 90-day payment is informed by the literature.

Literature Summary of Social Risk Factors and THA/TKA Payment

We reviewed the literature to examine the relationship between social risk factors and hospital 90-day RSPs following elective primary THA/TKA.

Our literature summary found that most studies examine income, insurance status, race, and gender. Only a limited number of studies examined the relationship between social risk factors and costs following hip/knee surgery for the older, insured Medicare population. Some of the evidence comes from studies that look drivers of costs, such as readmission, and use of post-acute care (Peel et. al, 2015).

The relationship between social risk factors and episode-of-care payment is complex and not well understood. Most studies indicate that low-income, lower insurance status, and non-white race are associated with higher costs (Browne et al. 2014; Pugely et al. 2014). Higher costs, however, can be due to lower quality of care (e.g., increased complication of care, hospital length of stay or readmission rates) or non-home post-discharge

destinations (e.g., discharged to a rehabilitation facility or a SNF). However, other studies suggest that neither socioeconomic status nor race is related to increased costs or that costs for patients with social risk factors may be lower due to less utilization of post-discharge resources (Freburger et al. 2011).

Additionally, it is important to consider whether higher costs associated with socially disadvantaged patients who undergo hip and knee replacement influences clinical or functional outcomes. For example, if there is a pattern showing that increased spending results in better outcomes for socially disadvantaged patients, it might be appropriate to risk adjust. However, given the evidence of complex relationships among, social risk, costs, and outcomes, if there were no consistent association between payment and quality, it may not be appropriate to risk adjust. In fact, studies show that while drivers of higher costs (such as non-home discharge and length of stay) for patients with social risk factors can be higher, outcomes for those patients can also be worse (Alverez et al., 2022; Chun et al., 2021).

Potential Mechanisms by which Social Risk Factors Affect Costs

Potential causal mechanisms by which social risk factors influence costs following THA/TKA surgery are varied and complex. Although some recent literature assesses the relationship between patient social risk factors (e.g., gender, socioeconomic status and race) and costs following THA/TKA surgery, few studies directly address the complex causal pathways. Our literature summary has identified four potential mechanisms at the patient- and hospital-level: (1) Health at admission and other patient characteristics, (2) selection of patients into different quality hospitals, (3) care within hospital, and (4) post-discharge setting.

Health at admission and other patient characteristics

Comorbidities

Patients with social risk factors such as low socioeconomic status may have more comorbid conditions at the time of surgery related to historical or lifelong social disadvantage. For example, research shows that Medicaid insured and Black patients present more clinical comorbidities and worse function than patients with higher income or than white patients when they undergo total hip/knee replacement (Schwarzkopf et al. 2015). A recent study that examined the impact of CMS's Comprehensive Care for Joint Replacement (CJR) program found persistent differences in pre-operative comorbidities when comparing Black and white patients (Okewunmi et al., 2022).

Worse health at admission makes patients with social risk factors potentially more expensive to care for. For instance, increased comorbidities are associated with increased resource use (e.g., cost-to-charge ratios) and hospital length of stay (Pugely et al. 2014). Comorbidities increase the likelihood of complications of care and readmission within 90-days of discharge (Courtney et al. 2017); readmissions are a substantial driver of costs (Swenson et al., 2017). The THA/TKA Payment measure adjusts for comorbidities to account for health at admission.

Other patient-level factors

There is evidence that being female, or non-Hispanic Black race is associated with increased length of stay (Weiner et al, 2020) and costs related to hip or knee replacement (Pugely et al. 2014). In addition, Black and Hispanic patients were found to have a higher risk of 90-day post-operative ED visits and 30- and 90-day readmission (White et al., 2018). More recent studies have shown a relationship between marital status and length of stay (which may impact payments), as well as between tobacco store density and length of stay; tobacco store density and food deserts were also associated with increased total cost of care for THA/TKA procedures (Delanois et al., 2021).

Other patient-level risk factors such as frailty, age, and gender, have been examined for their relationship to THA/TKA outcomes such as complications, which may increase utilization, including length of stay. However, social risk factors differ in their association with specific complications. For example, age is associated with a greater risk of cardiac complications (Elsiwy et al, 2019), and women are more likely to experience venous thromboembolism (Zhang et al, 2015), whereas men are at higher risk of death, AMI, pneumonia, and surgical site infections (Basques et al, 2019). Frailty has been shown to be associated with a higher risk of reoperation and readmission (Runner et al, 2017).

Selection of patients into different quality hospitals

Some studies examining the link between social risk factors and costs suggest that the relationship is mediated by hospital quality. Disadvantaged patients are more likely to select and be admitted to lower quality hospitals. For

example non-white patients are more likely to have a THA/TKA procedure at lower-volume facilities (Dy et al., 2015). Furthermore, studies have also suggested other disparities related to hip/knee surgery, including significant socioeconomic and racial differences in access to care and quality of care following total hip/knee replacement (Keswani et al. 2016; Mahomed, 2003, Usiskin and Misra, 2022; Alvarez et al., 2022).

Low- and high-quality hospitals can both increase costs. On the one hand, low-quality hospitals increase costs because the lower quality of care may require more frequent and intense follow-up care. But high-quality hospitals can also lead to increased costs by offering higher quality care, and specialty hospitals might charge a premium. For example, in their 2007 study, Cram et al. showed that specialty hospitals were more likely to admit patients with fewer comorbidities and from more affluent neighborhoods. These hospitals tended to have better outcomes than non-specialty hospitals. Although they did not assess costs explicitly, the suggestion of the article is that these specialty hospitals take advantage of Medicare's diagnoses-related-group-based reimbursement system by selecting only low-risk patients. Cram et al. conclude that specialty hospitals may contribute to differential healthcare costs among socioeconomic groups through patient selection.

Care within hospital

Social risk factors can contribute to costs if patients do not receive equivalent or patient-centered care within a facility. For example, a study using linked hospital and census data found that low income or minority patients may experience differential, lower quality, or discriminatory care within a given facility (Trivedi 2014). Alternatively, patients with social risk factors may require and not necessarily receive differentiated care, such as provision of lower literacy information. For example, hospitals may provide the same care for all patients (e.g. the same discharge instructions) but this care might be insufficient for patients with social risk factors (e.g. due to low literacy). Failure to meet the needs of socially disadvantaged patients can lead to costly complications requiring readmission following hip or knee replacement or provision of inpatient rehabilitation in costly settings. Failure to meet the needs of socially disadvantaged patients can lead to costly complications requiring readmission following hip or knee replacement or provision of inpatient rehabilitation in costly settings. However, to the extent that social risk factors exert effect on care within a hospital through the above mechanisms, we do not believe social risk factors should be adjusted for in the THA/TKA payment measure.

Post-discharge setting

Numerous studies have shown that low-income patients and patients of races other than white are more likely to be discharged to a skilled nursing facility (SNF) or a rehabilitation facility rather than to a home setting with or without health services (Courtney 2017; Inneh et al.; Keswani et al. 2015, Alvarez et al., 2022; Usiskin & Misra, 2022). Different factors might explain this pattern in post-discharge settings. Low-socioeconomic status and non-white patients tend to undergo THA/TKA surgery at an older age, present more comorbidities and poorer function preoperatively (Schwarzkopf et al. 2015). They are more likely to experience post-operative complications, which would explain the need for more intensive post-discharge care. Care providers may also prefer to send patients living in some rural or urban areas to inpatient institutions to ensure that geographically isolated patients or those living in inadequate housing receive adequate care. Lack of social support (i.e., not being married/having a partner) and access to social services (childcare, transportation, housing stability) can also help explain why low-socioeconomic status patients are more often discharged to inpatient facilities. Finally, providers' beliefs about patients' health behaviors (e.g., low compliance in filling and taking medications, following discharge instructions, attending appointments, etc.) may incentivize them to discharge patients with social risk factors to inpatient facilities despite the fact that early discharge and discharge to home (with and without health services) are associated with better health outcomes and lower costs. In fact, analyses of data from the Comprehensive Care for Joint Replacement (CJR) Model show that when providers are at risk for payments related to post-acute care, post-acute care utilization goes down (Agarwall, 2020).

Other studies have shown evidence that low-socioeconomic status patients have more limited access to healthcare providers and intensive care in post-discharge settings, which may lower costs for these patients (Freburger et al. 2011). Furthermore, some studies indicate that race is not a predictor of discharge to institutions in a Medicare-only cohort and suggest that Black patients are more likely to be discharged home compared to others (Pugely et al. 2014).

Other social risk factor considerations

More recent studies have shown a relationship between marital status and length of stay (which may impact payments), as well as between tobacco store density and length of stay (Delanois et al., 2021); tobacco store density and food deserts were also associated with increased total cost of care for THA/TKA procedures.

In addition to race and socioeconomic risk factors, other factors such as frailty, age, gender, and provider density have been examined for their relationship to THA/TKA outcomes such as complications, which may increase length of stay and utilization. However, social risk factors differ in their association with specific complications. For example, age is associated with a greater risk of cardiac complications (Elsiwy et al, 2019), and women are more likely to experience venous thromboembolism (Zhang et al, 2015), whereas men are at higher risk of death, AMI, pneumonia, and surgical site infections (Basques et al, 2019). Frailty has been shown to be associated with a higher risk of reoperation and readmission (Runner et al, 2017). Finally, one study found no association between THA/TKA outcomes and primary care provider density (Mehta et al, 2021).

Conceptual Model

Based on the literature described above, and on additional conceptual relationships, we identified and characterized the individual social risk factors that may impact the outcome (Figure 3, Table 6). For some of these factors, providers may be able to implement mitigating interventions. We note that we did not include insurance status as a risk factor because the target audience for this measure is Medicare Fee-For-Service patients. Figure 3 diagrams the pathways that influence the outcome of payment for patients undergoing inpatient THA/TKA replacement. Table 6 identifies the specific risk factors, the variables available for testing, if they are currently accounted for in the measures' risk model, and how hospitals might mitigate the impact of the risk factor.

Figure 3: Conceptual model for the THA/TKA payment outcome

Factors that influence the outcome:

1. **Hospital-provided care directly related to the outcome**, which can reduce a patient's risk or increase a patient's risk. Some of this care can directly impact the outcome (such as following standards of care, hiring and maintaining high quality staff, and following standardized processes). This includes providing access to high-quality timely care and the appropriate needed services.

2. Patient level risk factors present at the start of care, which can be clinical, demographic, or social. These can be broken into two categories – factors for which hospitals can provide mitigating care, and factors which hospitals cannot control.

2a. Factors for which **hospital-provided care can likely mitigate** the impact (such as providing translators or easy-to-understand home care instructions), at least in part.

2b. Factors that are **beyond the hospital's control** and cannot be mitigated by hospital-based interventions**.**

Based on our literature review and empiric data, including feedback from subject matter experts in social risk factor adjustment, we have identified patient-level risk factors present at the start of care that can influence the outcome (Table 6). For each patient-level factor, we indicate how the factor may be related to the outcome and if there is a variable that is available for testing. We note that casual pathways can be very complex and that this is a high-level overview of each factor and its potential role in the outcome. In Table 7 we indicate strategies for how hospitals may be able to mitigate the impact of each risk factor.

In our conceptual model and outlined in Table 6, we have identified four patient-level variables which are **already accounted for in the risk adjustment model:**

- Age
- Sex
- Comorbidities
- Frailty

In our conceptual model and outlined in Table 6, we have identified 7 patient-level variables which are not accounted for in the model and that have the potential for mitigation by hospitals, outlined in Table 7.

Risk Factor including Social Risk Factors	How it conceptually impacts the outcomes	Variable(s)	Is a variable available for use with claims?	Is the variable currently in the risk model?
Age	Higher age higher risk	Age	Yes	Yes
Sex	Women have higher risk	Sex	Yes	Yes
Comorbidities	Certain comorbidities increase risk	Various ICD-10 groupings	Yes	Yes
Frailty	More frail higher risk	ICD-10 group	Yes	Yes
Exposure to racism	Non-white at higher risk	Race	Yes	No
Income	Low income at higher risk	AHRQ SES/Dual Eligibility	Yes	No
Health literacy	Lower health literacy higher risk	Self-reported health literacy	No	No
Education	Lower education at higher risk	AHRQ SES	Yes	No
Access to timely post-operative care	More rural higher risk	Geographic location (urban/rural)	Yes	No
Access to high quality hospital/surgeon	Proximity to low-quality hospitals increases risk	No variable available	No	No
Social support (e.g. marital status)	Less support at higher risk	Marital status	No	No

Table 6: Patient-level risk factors (social/demographic/functional) at the start of care, the impact on the outcome (payments), available variables, and potential for mitigation by hospitals

The conceptual model (Figure 3) identifies that hospitals can implement mitigating strategies to counteract the impact of patient-level social risk factors outlined in Table 6. We note that some the evidence for mitigation comes from quality improvement efforts related to THA/TKA procedures. For example, there is some evidence from the implementation of the Comprehensive Care for Joint Replacement (CJR) model, that, while disparities continue to exist for some outcomes, disparities in care have narrowed for outcomes that impact payment, such as 90-day readmission (Okewunmi et al, 2022). This may be due to the types of processes that facilities implemented in response to this payment model, including redesigning and standardizing care pathways that include pre-procedure education, discharge planning, preemptive multimodal pain control, and early rehabilitation, which have been shown to improve outcomes, including for patients with social risk factors (Riepen et al, 2021). Other facility-specific care redesign features associated with improved intermediate outcomes to payment include use of risk prediction tools, having care managers engage with patients before admission, nurse navigators, and 7-day access to an after-hours clinic (Gray et al, 2018). Furthermore, from work related to the CJR model, assigning the financial accountability for post-acute care to the hospital resulted in decreased post-acute care costs (Dummit et al., 2016; Navathe et al., 2017). Some of the evidence to support mitigating solutions is generalized from broader research much of which focuses on preventing readmission (CMS Guide to Reducing Disparities in Readmissions, 2018), which is a driver of costs.

Because several of these social risk factors have similar underlying drivers of poor outcomes (such as education and income, and race and income), we address the mitigating strategies below as a series of topics and the recommended strategies. We also acknowledge that mitigating exposure to racism (Commonwealth, 2021), in particular, is a complex issue and that hospitals may initially struggle to develop and implement effective approaches. In addition, additional research is needed to identify the most effective approaches (Ricks et al, 2021).

Strategy	Description of interventions	Related social risk factors
Improving post-procedural follow up care	- Advance care transition planning and follow up for patients at high risk - Communicate with patients about importance of follow up care; assist with scheduling appointments - Offer telehealth options - Expand clinic hours to avoid ED use - Engage family/caregivers - Use nurse navigators	Income Education Exposure to racism Access to post-operative care Marital status
Improve access to a usual source of care	- Ensure patient is connected with a usual source of care	Income Education Exposure to racism Access to post-operative care

Strategy	Description of interventions	Related social risk factors
Reduce language/literacy barriers	- Identify patients at risk (language and literacy barriers) - Ensure access to translation services - Communicate at home or follow up care instructions in patients' native language and in a culturally competent manner - Simplify instructions - Communicate instructions at the appropriate literacy level - Engage family/caregivers	Low health literacy Limited English proficiency
Reduce socioeconomic barriers	- Connect patients with community-based resources that address the need (e.g., housing and food insecurity, transportation, employment) - Connect underinsured patients with supplemental insurance - Connect with social support services	Income Education Exposure to racism
Reduce biased care	- Track metrics stratified by race and ethnicity - Quality improvement - Staff training - Diversity of staff, trainees, and Board of Directors	Exposure to racism Income
Improve access to high-quality care	- Recruit, train, and retain high-quality staff - Follow standards of care and use a learning healthcare system - Address workforce shortage and burnout	Exposure to racism Income Education

Table 7: Strategies and interventions to reduce the impact of social risk factors

Based on the literature showing a relationship between social risk factors and complications, and in assessing the available risk variables that can be linked to claims data (outlined in section 1.8), we tested the impact of adjusting for the following social risk variables on the IP THA/TKA measure:

- Dual eligible status
- Low AHRQ SES
- Race
- Rurality

Please refer to section 2a.08 for a detailed description of each social risk factor.

References

Alvarez PM, McKeon JF, Spitzer AI, Krueger CA, Pigott M, Li M, Vajapey SP. Race, Utilization, and Outcomes in Total Hip and Knee Arthroplasty: A Systematic Review on Health-Care Disparities. JBJS Rev. 2022 Mar 1;10(3).

Agarwal R, Liao JM, Gupta A, Navathe AS. The Impact Of Bundled Payment On Health Care Spending, Utilization, And Quality: A Systematic Review. Health Aff (Millwood). 2020 Jan;39(1):50-57.

Browne JA, Novicoff WM, D'Apuzzo MR. Medicaid payer status is associated with in-hospital morbidity and resource utilization following primary total joint arthroplasty. The Journal of Bone and Joint Surgery. 2014; 96(21) e180-1-6

Cram P, Vaughan-Sarrazin MS, Wolf B, Katz JN, Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. Journal of Bone & Joint Surgery. 2007; 89(8): 1675-1684

Courtney PM, Huddleston JI, Iorio R, Markel DC. Socioeconomic Risk Adjustment Models for Reimbursement Are Necessary in Primary Total Joint Arthroplasty. The Journal of Arthroplasty. 2017; 32-1: 1-5

Delanois RE, Tarazi JM, Wilkie WA, Remily E, Salem HS, Mohamed NS, Pollack AN, Mont MA. Social determinants of health in total knee arthroplasty : are social factors associated with increased 30-day post-discharge cost of care and length of stay? Bone Joint J. 2021 Jun;103-B(6 Supple A):113-118.

Dummit LA, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, Press MJ, Flood S, Muldoon LD, Gu Q, Hassol A, Bott DM, Bassano A, Conway PH. Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. JAMA. 2016 Sep 27;316(12):1267-78.

Dy CJ, Marx RG, Ghomrawi HM, Pan TJ, Westrich GH, Lyman S. The potential influence of regionalization strategies on delivery of care for elective total joint arthroplasty. J Arthroplasty 2015;30:1–6.

Freburger J.K., Holmes G.M., Ku Cutchin L.J., Heatwole-Shank K., Edwards L.J. E. Disparities in post-acute rehabilitation care for joint replacement. Arthritis Care & Research 2011; 63(7): 1020–1030

Inneh IA, Clair AJ, Slover JD, Iorio R. Disparities in Discharge Destination After Lower Extremity Joint Arthroplasty: Analysis of 7924 Patients in an Urban Setting. The Journal of arthroplasty. 2016; 31(12): 2700-2704

Keswani A, Tasi MC, Fields A, Lovy AJ, Moucha CS, Bozic KJ. Discharge Destination After Total Joint Arthroplasty: An Analysis of Postdischarge Outcomes, Placement Risk Factors, and Recent Trends. The Journal of Arthroplasty. 2016; 31(6): 1155-1162

Navathe AS, Troxel AB, Liao JM, Nan N, Zhu J, Zhong W, Emanuel EJ. Cost of Joint Replacement Using Bundled Payment Models. JAMA Intern Med. 2017 Feb 1;177(2):214-222.

Okewunmi J, Mihalopoulos M, Huang HH, Mazumdar M, Galatz LM, Poeran J, Moucha CS. Racial Differences in Care and Outcomes After Total Hip and Knee Arthroplasties: Did the Comprehensive Care for Joint Replacement Program Make a Difference? J Bone Joint Surg Am. 2022 Jun 1;104(11):949-958. doi: 10.2106/JBJS.21.00465. Epub 2022 Apr 15. PMID: 35648063.

Peel TN, Cheng AC, Liew D, Buising KL, Lisik J, Carroll KA, Choong PF, Dowsey MM. Direct hospital cost determinants following hip and knee arthroplasty. Arthritis Care Res (Hoboken). 2015 May;67(6):782-90.

Pugely AJ, Martin CT, Gao Y, Belatti DA, Callaghan JJ. Comorbidities in patients undergoing total knee arthroplasty: do they influence hospital costs and length of stay?. Clinical Orthopaedics and Related Research 2014; 472(12): 3943–3950.

Ricks TN, Abbyad C, Polinard E. Undoing Racism and Mitigating Bias Among Healthcare Professionals: Lessons Learned During a Systematic Review. J Racial Ethn Health Disparities. 2021 Sep 3:1–11. doi: 10.1007/s40615-021-01137-x. Epub ahead of print. PMID: 34480317; PMCID: PMC8415190.

Riepen DW, Gelvez D, Collett GA, Nakonezny P, Estrera KA, Huo MH. Standardized total knee arthroplasty pathway improves outcomes in minority patients. Am J Manag Care. 2021 May 1;27(5):e152-e156.

Schwarzkopf J. Ho, N. Snir. Factors influencing discharge destination after total hip arthroplasty: a California state database analysis. A California State Database Analysis. geriatric orthopedics Surgery & rehabilitation. 2015; 6(3): 215-219

Swenson ER, Bastian ND, Nembhard HB, Davis Iii CM. Reducing cost drivers in total joint arthroplasty: understanding patient readmission risk and supply cost. Health Syst (Basingstoke). 2017 Nov 14;7(2):135-147. doi: 10.1080/20476965.2017.1397237. PMID: 31214344; PMCID: PMC6452843.

Trivedi AN, Nsa W, Hausmann LRM, et al. Quality and Equity of Care in U.S. Hospitals. New England Journal of Medicine. 2014;371(24):2298-2308.

Usiskin I, Misra D. Racial Disparities in Elective Total Joint Arthroplasty for Osteoarthritis. ACR Open Rheumatol. 2022 Apr;4(4):306-311.

Weiner JA, Adhia AH, Feinglass JM, Suleiman LI. Disparities in Hip Arthroplasty Outcomes: Results of a Statewide Hospital Registry From 2016 to 2018. J Arthroplasty. 2020 Jul;35(7):1776-1783.e1. doi: 10.1016/j.arth.2020.02.051. Epub 2020 Feb 28. PMID: 32241650.

[Response Ends]

2b.14. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

Previous submission

Variable	Payment Ratio (95% CI)
Age minus 65 (years above 65, continuous)	1.01 (1.01 - 1.02)
Male	0.94 (0.94 - 0.94)
Index admissions with an elective THA procedure	1.01 (1.00 - 1.01)
Procedure type (bilateral joint replacement)	1.74 (1.73 - 1.75)
Procedure type (staged joint replacements)	1.75 (1.73 - 1.77)
Procedure type (single joint replacement; reference group)	REFERENCE
Morbid obesity (ICD-9 diagnosis code 278.01)	1.12 (1.12 - 1.12)
Congestive heart failure (CC 80)	1.06 (1.05 - 1.06)
Acute coronary syndrome (CC 81-82)	1.02 (1.02 - 1.02)
Valvular or rheumatic heart disease (CC 86)	1.01 (1.01 - 1.01)
Hypertension and hypertension complications (CC 89-91)	1.03 (1.02 - 1.03)
History of infection (CC 1, 3-6)	1.04 (1.04 - 1.04)
Metastatic cancer or acute leukemia (CC 7)	1.03 (1.02 - 1.04)
Cancer (CC 8-12)	0.99 (0.99 - 1.00)
Benign neoplasms of skin, breast, eye (CC 14)	0.98 (0.98 - 0.99)
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)	1.05 (1.05 - 1.05)
Protein-calorie malnutrition (CC 21)	1.17 (1.16 - 1.19)
Other significant endocrine and metabolic disorders (CC 22)	1.03 (1.02 - 1.03)
Obesity/disorders of thyroid, cholesterol, lipids (CC 24, excluding ICD-9 diagnosis code 278.01)	0.99 (0.99 - 0.99)
Appendicitis (CC 35)	0.96 (0.94 - 0.99)
Bone/joint/muscle infections/necrosis (CC 37)	1.04 (1.04 - 1.05)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 38)	1.02 (1.02 - 1.03)
Disorders of the vertebrae and spinal discs (CC 39)	1.01 (1.01 - 1.01)
Osteoarthritis of hip or knee (CC 40)	1.08 (1.07 - 1.08)
Other musculoskeletal and connective tissue disorders (CC 43)	1.03 (1.03 - 1.03)
Severe hematological disorders (CC 44)	1.09 (1.08 - 1.11)
Coagulation defects and other specified hematological disorders (CC 46)	1.02 (1.01 - 1.02)
Delirium and encephalopathy (CC 48)	1.04 (1.03 - 1.05)
Dementia or other specified brain disorders (CC 49-50)	1.11 (1.10 - 1.11)
Major psychiatric disorders; personality disorders (CC 54-57)	1.08 (1.08 - 1.09)

Variable	Payment Ratio (95% CI)
Depression/anxiety (CC 58-59)	1.04 (1.04 - 1.04)
Other psychiatric disorders (CC 60)	1.02 (1.02 - 1.02)
Mental retardation or developmental disability (CC 61-65)	1.22 (1.19 - 1.25)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.07 (1.06 - 1.07)
Polyneuropathy (CC 71)	1.04 (1.04 - 1.04)
Multiple sclerosis (CC 72)	1.12 (1.10 - 1.14)
Parkinson's and Huntington's disease (CC 73)	1.19 (1.18 - 1.20)
Seizure disorders and convulsions (CC 74)	1.06 (1.06 - 1.07)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	1.01 (1.01 - 1.02)
Stroke (CC 95-96)	1.04 (1.03 - 1.04)
Vascular or circulatory disease (CC 104-106)	1.03 (1.02 - 1.03)
Chronic Obstructive Pulmonary Disease (COPD) (CC 108)	1.04 (1.04 - 1.05)
Pleural effusion/pneumothorax (CC 114)	0.98 (0.98 - 0.99)
Other lung disorders (CC 115)	1.01 (1.01 - 1.02)
Legally blind (CC 116)	1.12 (1.10 - 1.14)
Dialysis status (CC 130)	1.39 (1.35 - 1.42)
Renal failure (CC 131)	1.04 (1.04 - 1.04)
Incontinence (CC 134)	1.05 (1.05 - 1.06)
Urinary tract infection (CC 135)	1.01 (1.01 - 1.01)
Other urinary tract disorders (CC 136)	1.01 (1.01 - 1.01)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	1.09 (1.08 - 1.09)
Cellulitis, local skin infection (CC 152)	1.02 (1.02 - 1.03)
Other dermatological disorders (CC 153)	0.99 (0.99 - 0.99)
Trauma (CC 154-156, 158-161)	1.05 (1.05 - 1.06)
Vertebral fractures (CC 157)	1.04 (1.03 - 1.05)
Other injuries (CC 162)	1.01 (1.01 - 1.01)
Major symptoms, abnormalities (CC 166)	1.03 (1.03 - 1.03)
Minor symptoms, signs, findings (CC 167)	1.02 (1.02 - 1.02)
Patient level dual eligibility	1.12 (1.12 - 1.13)

Table 8: Risk variables and payment ratios for the THA/TKA Payment measure (Dataset 2)

Risk Variable	Payment Ratio (95% CI)
Age minus 65 (years above 65, continuous)	1.01 (1.01 , 1.01)
Male	0.96 (0.95 , 0.96)
Index admissions with an elective THA procedure	0.99 (0.99 , 0.99)
Procedure type (bilateral joint replacement)	1.77 (1.75 , 1.78)
Procedure type (staged joint replacements)	1.75 (1.73 , 1.77)

Risk Variable	Payment Ratio (95% CI)
Procedure type (single joint replacement; reference group)	REFERENCE
Morbid obesity (ICD-9 diagnosis code 278.01)	1.09 (1.09 , 1.09)
Congestive heart failure (CC 80)	1.07 (1.06 , 1.07)
Acute coronary syndrome (CC 81-82)	1.01 (1.01 , 1.02)
Valvular or rheumatic heart disease (CC 86)	1.01 (1.01 , 1.01)
Hypertension and hypertension complications (CC 89-91)	1.02 (1.02 , 1.02)
History of infection (CC 1, 3-6)	1.04 (1.04 , 1.04)
Metastatic cancer or acute leukemia (CC 7)	1.05 (1.04 , 1.06)
Cancer (CC 8-12)	0.99 (0.99 , 1.00)
Benign neoplasms of skin, breast, eye (CC 14)	0.98 (0.98 , 0.98)
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)	1.04 (1.04 , 1.04)
Protein-calorie malnutrition (CC 21)	1.15 (1.14 , 1.16)
Other significant endocrine and metabolic disorders (CC 22)	1.03 (1.02 , 1.03)
Obesity/disorders of thyroid, cholesterol, lipids (CC 24, excluding ICD-9 diagnosis code 278.01)	0.99 (0.98 , 0.99)
Appendicitis (CC 35)	0.96 (0.94 , 0.98)
Bone/joint/muscle infections/necrosis (CC 37)	1.06 (1.05 , 1.07)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 38)	1.02 (1.02 , 1.03)
Disorders of the vertebrae and spinal discs (CC 39)	1.01 (1.01 , 1.01)
Osteoarthritis of hip or knee (CC 40)	1.06 (1.05 , 1.06)
Other musculoskeletal and connective tissue disorders (CC 43)	1.02 (1.02 , 1.02)
Severe hematological disorders (CC 44)	1.11 (1.10 , 1.13)
Coagulation defects and other specified hematological disorders (CC 46)	1.01 (1.01 , 1.02)
Delirium and encephalopathy (CC 48)	1.07 (1.06 , 1.08)
Dementia or other specified brain disorders (CC 49-50)	1.11 (1.10 , 1.11)
Major psychiatric disorders; personality disorders (CC 54-57)	1.07 (1.07 , 1.07)
Depression/anxiety (CC 58-59)	1.03 (1.03 , 1.03)
Other psychiatric disorders (CC 60)	1.02 (1.02 , 1.02)
Mental retardation or developmental disability (CC 61-65)	1.13 (1.11 , 1.15)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.11 (1.10 , 1.12)

Risk Variable	Payment Ratio (95% CI)
Polyneuropathy (CC 71)	1.01 (1.01 , 1.02)
Multiple sclerosis (CC 72)	1.13 (1.11 , 1.15)
Parkinson's and Huntington's disease (CC 73)	1.19 (1.18 , 1.20)
Seizure disorders and convulsions (CC 74)	1.07 (1.06 , 1.08)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	1.02 (1.01 , 1.02)
Stroke (CC 95-96)	1.04 (1.03 , 1.05)
Vascular or circulatory disease (CC 104-106)	1.03 (1.03 , 1.03)
Chronic Obstructive Pulmonary Disease (COPD) (CC 108)	1.06 (1.05 , 1.06)
Pleural effusion/pneumothorax (CC 114)	0.99 (0.98 , 1.00)
Other lung disorders (CC 115)	1.01 (1.01 , 1.01)
Legally blind (CC 116)	1.12 (1.10 , 1.15)
Dialysis status (CC 130)	1.46 (1.43 , 1.50)
Renal failure (CC 131)	1.05 (1.05 , 1.05)
Incontinence (CC 134)	1.04 (1.04 , 1.04)
Urinary tract infection (CC 135)	1.02 (1.02 , 1.02)
Other urinary tract disorders (CC 136)	1.00 (1.00 , 1.01)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	1.09 (1.09 , 1.10)
Cellulitis, local skin infection (CC 152)	1.03 (1.03 , 1.03)
Other dermatological disorders (CC 153)	0.98 (0.98 , 0.99)
Trauma (CC 154-156, 158-161)	1.05 (1.04 , 1.05)
Vertebral fractures (CC 157)	1.04 (1.03 , 1.05)
Other injuries (CC 162)	1.02 (1.02 , 1.02)
Major symptoms, abnormalities (CC 166)	1.03 (1.03 , 1.03)
Minor symptoms, signs, findings (CC 167)	1.01 (1.01 , 1.01)

Table 9: Risk variables and payment ratios for the THA/TKA Payment measure (2022 EM Testing Dataset)

[Response Ends]

2b.15. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

Variation in prevalence of social risk factors across measured entities

Previous submission

The prevalence of social risk factors in the hip/knee payment cohort varies across measured entities. The median percentage of dual eligible patients is 6.6% (interquartile range [IQR]: 3.8% – 11.5%). The median percentage of low-socioeconomic status patients using the AHRQ SES Index score is 12.7% (IQR: 86.2% – 23.6%).

Current submission

For these analyses we used the percentage of patients with an AHRQ socioeconomic status index score equal to or below 46.0 to define the lowest quartile of the AHRQ socioeconomic status Index and refer to this variable as "low AHRQ SES." We used rural area commuting code 10, Rural areas: primary flow to a tract outside a urbanized area or urban cluster, to define rurality variable.

Table 10 shows the prevalence of social risk factors in the THA/TKA payment cohort using the 2022 EM Dataset. The mean prevalence of patients with low AHRQ SES using updated data is similar to the previous submission (11.7% vs. 12.7%) but the mean prevalence of the dual eligibility variable is lower (3.4% vs. 6.6%).

Social Risk Factor	Median Hospital Prevalence (IQR)
Dual Eligibility	3.4% (1.4-7.9%)
Low AHRQ SES	11.7% (5.0-24.0%)
Race (Black)	1.6% (0.0-6.5%)
Rurality	2.1% (0.0-9.7%)

Table 10: Prevalence of social risk factors in the THA/TKA payment cohort

Empirical association with the outcome (univariate)

Previous submission

Mean observed payments were about $5,000 higher for dual eligible patients ($27,521; SD: 12,080), compared to non-dual eligible patients ($22,568; SD: 9,532). Similarly, mean observed payments for low-socioeconomic status patients using the AHRQ SES Index score was $24,206 (SD: 10,666), compared with $22,700 (SD: 9,646) for non-low socioeconomic status patients.

Current submission

Mean observed payments were higher for patients with dual eligibility, low AHRQ SES, and race (Black) compared with patients without the social risk factors (Table 11). Mean observed payments for patients living in rural areas was similar to patients living in non-rural areas.

Social Risk Factors	Number of patients with the social risk factor vs. without	Mean observed payments within population with the social risk factor vs. without ($)	Difference in mean payments between population with and without the social risk factor ($)
Dual Eligibility	37,798 (vs 910,617)	26,509 (vs. 20,752)	5,757
Low AHRQ SES	102,894 (vs 796,458)	22,505 (vs. 20,779)	1,726
Race (Black)	44,863 (vs 903,594)	23,323 (vs. 20,865)	2,458
Rurality	50,015 (vs 897,827)	21,051 (vs. 20,978)	73

Table 11: Mean observed payments for patients with social risk factors

Incremental effect of social risk factors in a multivariable model

Previous submission

We also examined the strength and significance of the social risk factors in the context of a multivariable model. Consistent with the above findings, when we include any of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of each of these variables remains significant, but somewhat lower, than the coefficient for the bivariate association (the parameter estimate decreased from 1.19 to 1.12 for dual eligibility and from 1.05 to 1.04 for the AHRQ SES Index).

Current submission

Updated results show that the payment ratios for the social risk factors remain significant for each of the social risk factors (Table 12), however the effect size for dual eligibility is lower (1.05 vs. 1.12) compared with the prior submission.

MODEL	Social risk factor	Payment ratio	p-value
Base Model plus Dual Eligibility	Dual Eligibility	1.05	<.0001
Base Model plus Low AHRQ SES	Low SES	1.04	<.0001
Base Model plus Race (Black)	Race (Black)	1.07	<.0001
Base Model plus Rurality	Rurality	1.01	<.0001

Table 12: Strength and significance of each social risk factor within a multivariate model

Model performance

Previous submission

To further understand the relative importance of these risk-factors in the measure, we compared hospital performance with and without the addition of each social risk variable. Results show that the quasi-R square is almost unchanged with the addition of any of these variables into the model: The quasi-R square of the original model (i.e., that does not include any social risk factor) is 0.21; the quasi-R square of the original model with the dual eligible variable added is 0.23; and the original model with the AHRQ SES index variable added is 0.22.

Current submission

Updated results show similar model performance when adding each social risk factor, separately, to the base model (Table 13).

MODEL	quasi-R-squared
Base model	0.18
Base Model plus Low AHRQ SES	0.18
Base Model plus Dual Eligibility	0.18
Base Model plus Race (Black)	0.18
Base Model plus Rurality	0.18

Table 13: Model performance for the THA/TKA payment measure with the addition of social risk factors

Impact on measure scores

Previous submission

Differences in RSPs after adding social risk factors compared with the current model is small and the mean change is -$3.91 when just adding AHRQ SES index factor (median: $31.60; IQR: $-57.45 – $84.45) and -$10.89 when just adding dual eligibility (median: $86.46; IQR: $-41.52– $160.09)

Overall, we find that social risk factors that could feasibly be incorporated into our model do have a significant relationship with the outcome in multivariable modeling.

Current submission

Median absolute (not relative) changes in measure scores (RSPs) after adding social risk factors (compared with measure scores calculated without social risk factors) show that adding the dual eligibility, low AHRQ SES, or race (Black) variables to the model has a greater impact on mean payments compared with adding the rurality variables

to the model (Table 14). However, changes are small (less than 0.3% of total mean payments), and correlations in measure scores comparing models with and without social risk factors, show that measure scores are highly correlated (Table 14).

Social risk factor	Median change in RSP (Percent of Average Payments)	Interquartile Range (IQR)	Correlation in measure scores (Pearson Correlation Coefficient)
Dual Eligibility	$76.84 (0.29%)	$-18.50 to $125.20	0.994
Low AHRQ SES	$47.73 (0.21%)	$ -61.50 to $121.40	0.992
Race (Black)	$44.26 (0.19%)	$-13.00 to $70.80	0.998
Rurality	$7.77 (0.04%)	$-4.70 to $12.60	1.00

Table 14: Social risk factor impact on measure scores: differences in measure scores and correlation in measure scores calculated with and without social risk factors

To further understand the impact on measure scores, we examined the impact on the classification of hospitals into categories that are reported to the public on CMS's public reporting website, *Care Compare*.

To categorize hospital payments, CMS estimates each hospital's RSP and the corresponding 95% interval estimate. CMS assigns hospitals to a payment category by comparing each hospital's RSP interval estimate to the national mean payment. Comparative payments for hospitals with 25 or more eligible cases are classified as follows:

- "Less than the National Average Payment" if the entire 95% interval estimate surrounding the hospital's RSP is lower than the national mean payment.
- "No Different than the National Average Payment" if the 95% interval estimate surrounding the hospital's RSP includes the national mean payment.
- "Greater than the National Average Payment" if the entire 95% interval estimate surrounding the hospital's RSP is higher than the national mean payment.
- If a hospital has fewer than 25 eligible cases for a measure, CMS assigns the hospital to a separate category: "Number of Cases Too Small." This category is used when the number of cases is too small (fewer than 25) to reliably estimate the hospital's RSP. If a hospital has fewer than 25 eligible cases, the hospital's RSP and interval estimate will not be publicly reported for the measure.

Because this measure was initially NQF endorsed with dual eligibility in the risk model but was implemented without dual eligibility, we wanted to characterize the degree of reclassification when adding the social risk factor to the model. Our results showed that with the addition of dual eligibility to the model in total, 105 of 1,074 hospitals (3%) shifted one performance category; no hospital shifted more than one performance category. Specifically, with the addition of dual eligibility to the model, more hospitals (60) hospitals shifted one category higher (from less than the national average, to no different; or from no different, to greater than the national average); 45 hospitals shifted one category lower (from no different to less than the national average; or from greater than to no different than the national average).

Referral patterns by hospital characteristics

To better understand whether the association between social risk and payments was a patient- or hospital-level effect we examined referral patterns and observed payments for dual eligible and non-dual eligible patients among hospitals with a high overall proportion of dual eligible patients and hospitals with a low proportion of dual eligible patients. The purpose of this analysis was to understand whether patterns of use of post-acute care settings and payments associated with that care were driven mostly by the patient's dual eligible status or the fact that they received care at a hospital that cares for a large proportion of dual eligible patients (see table below). We performed the same analyses on hospitals with high and low proportions of low socioeconomic status patients using the AHRQ SES index (see table below).

Results showed that both types of hospitals are, on average, spending more for dual eligible patients and patients with higher AHRQ Index scores in the post-acute care setting. Regardless of hospitals' characteristics, the payment

for dual eligible patients and low socioeconomic status patients was consistently higher than for non-dual eligible patients and non-low socioeconomic status patients.

This finding suggests that hospital-level effects were not entirely driving higher payments for patients with social risk factors, but that patients with social risk factors may need more support to recover after knee/hip replacement; thus, higher payment would be appropriate.

Based on these results, we recommend risk-adjusting hip/knee payment for dual eligibility.

*	**Hospital Proportion of Dual Medicare and Medicaid Beneficiaries: Bottom Quartile**	**Hospital Proportion of Dual Medicare and Medicaid Beneficiaries: Bottom Quartile**	**Hospital Proportion of Dual Medicare and Medicaid Beneficiaries: Top Quartile**	**Hospital Proportion of Dual Medicare and Medicaid Beneficiaries: Top Quartile**
Care Setting	**Dual Beneficiaries**	**Non-dual Beneficiaries**	**Dual Beneficiaries**	**Non-dual beneficiaries**
Index Hospitalization	*	*	*	*
# of patients	8,262	315,958	21,683	81,841
$ per patient	$14,813	$14,610	$14,974	$14,749
Post-Acute Care (total)	*	*	*	*
# of patients	8,226	311,844	21,594	80,876
% of patients	99.6	98.7	99.6	98.8
$ per patient	$12,631	$7,655	$13,320	$9,613
Skilled Nursing Facilities	*	*	*	*
# of patients	4,636	108,556	11,818	33,270
% of patients	56.4	34.8	54.7	41.1
$ per patient	$13,109	$9.794	$13,304	$10,566
Inaptient Rehabilitation	*	*	*	*
# of patients	853	24,267	2,969	9,685
% of patients	10.4	7.8	13.7	12
$ per patient	$13,304	$12,572	$14,139	$13,336
Non-Acute Inpatient Settings	*	*	*	*
# of patients	22	232	71	153
% of patients	0.3	0.1	0.3	0.2
$ per patient	$7,918	$8,392	$12,501	$11,692

Table 15: Overall referral patterns for hospitals with high and low proportion of low-socioeconomic status patients, stratified by patient-level dual eligibility

*Cells intentionally left blank

*	Hospital Proportion of Patients with Low AHRQ SES: Bottom Quartile	Hospital Proportion of Patients with Low AHRQ SES: Bottom Quartile	Hospital Proportion of Patients with Low AHRQ SES: Top Quartile	Hospital Proportion of Patients with Low AHRQ SES: Top Quartile
Care Setting	**Low AHRQ SES**	**Non-Low-AHRQ SES**	**Low AHRQ SES**	**Non-Low-AHRQ SES**
Index Hospitalization	*	*	*	*
# of patients	9,131	252,356	45,308	84,006
$ per patient	$14,709	$14,662	$14,807	$14,785
Post-Acute Care (total)	*	*	*	*
# of patients	9,002	248,935	44,822	83,142
% of patients	98.6	98.6	98.9	99
$ per patient	$9,636	$8,035	$9,864	$8,916
Skilled Nursing Facilities	*	*	*	*
# of patients	4,094	98,544	17,856	30,880
% of patients	45.5	39.6	39.8	37.1
$ per patient	$11,096	$9,824	$11,421	$10,480
Inpatient Rehabilitation	*	*	*	*
# of patients	854	19,922	5,062	8,961
% of patients	9.5	8	11.3	10.8
$ per patient	$12,833	$12,300	$13,931	$13,750
Non-Acute Inpatient Settings	*	*	*	*
# of patients	12	229	115	177
% of patients	0.1	0.1	0.3	0.2
$ per patient	$5,630	$9,388	$11,301	$11,652

Table 16: Overall referral patterns for hospitals with high and low proportion of low-socioeconomic status patients, stratified by patient-level low-socioeconomic status patients (AHRQ SES Index)

*Cells intentionally left blank

Current submission

In the current submission we have updated the analyses of referral patterns and observed payments for patients with and without social risk factors among hospitals with a high overall proportion of patients with that social risk factor and hospitals with a low proportion of patients with that social risk factor. Similar to the previous submission and consistent with the literature, for dual eligibility (Table 17) we found that while both hospitals with a high and low proportion of dual eligible patients have similar total index payments, they are, on average, spending more in the Skilled Nursing Facility portion of post-acute care for dual eligible patients compared with non-dual eligible patients. We see similar but less marked results for the low AHRQ SES variable (Table 18). (The distribution of Black patients and rural patients is skewed and therefore not amenable to quartile analysis.) This finding suggests that patients with dual eligibility or low AHRQ SES may need more support to recover after knee/hip replacement; thus, higher payment may be appropriate. However, studies have also found that when

hospitals are held accountable for post-acute care costs, those costs are lower (Dummit et al., 2016; Navathe et al., 2017).

Payment Type	Statistic	Facilities in the Bottom Quartile for % Dual Eligibility: Patients Without Dual Eligibility	Facilities in the Bottom Quartile for % Dual Eligibility: Patients With Dual Eligibility	Facilities in the Top Quartile for % Dual Eligibility: Patients Without Dual Eligibility	Facilities in the Top Quartile for % Dual Eligibility Quartile: Patients With Dual Eligibility
Total Index Admission Payments	# of Patients	247,671	1,980	80,964	15,755
Total Index Admission Payments	**$ per Patient**	**$14,346**	**$14,712**	**$14,488**	**$14,790**
Total Post-Acute Care Payments	# of Patients	247,585	1,978	80,920	15,747
Total Post-Acute Care Payments	% of Patients	100.0	99.9	99.9	99.9
Total Post-Acute Care Payments	**$ per Patient**	**$5,862**	**$12,347**	**$7,614**	**$11,807**
Skilled Nursing Facility	# of Patients	51,522	933	22,382	7,465
Skilled Nursing Facility	% of Patients	20.8	47.2	27.7	47.4
Skilled Nursing Facility	**$ per Patient**	**$9,443**	**$15,064**	**$11,036**	**$13,769**
Inpatient Rehabilitation	# of Patients	9,144	131	4,219	1,067
Inpatient Rehabilitation	% of Patients	3.7	6.6	5.2	6.8
Inpatient Rehabilitation	**$ per Patient**	**$16,692**	**$17,381**	**$16,524**	**$16,069**
Non-Acute Inpatient Settings	# of Patients	94	2	80	29
Non-Acute Inpatient Settings	% of Patients	0.0	0.1	0.1	0.2
Non-Acute Inpatient Settings	**$ per Patient**	**$10,200**	**$17,447**	**$14,827**	**$11,156**

Table 17: Distribution of payments within the top and bottom quartiles of the facility proportion of patients with dual eligibility, for patients with and without dual eligibility.

Payment Type	Statistic	Facilities in the Bottom Quartile for % low AHRQ SES: Patients Without low AHRQ SES	Facilities in the Bottom Quartile for % low AHRQ SES: Patients WITH low AHRQ SES	Facilities in the TOP Quartile for % low AHRQ SES: Patients Without low AHRQ SES	Facilities in the TOP Quartile for % low AHRQ SES: Patients WITH low AHRQ SES
Total Index Admission Payments	# of Patients	248,111	7,466	61,861	32,890
Total Index Admission Payments	**$ per Patient**	**$14,442**	**$14,530**	**$14,573**	**$14,686**
Total Post-Acute Care Payments	# of Patients	248,029	7,466	61,826	32,858

Payment Type	Statistic	Facilities in the Bottom Quartile for % low AHRQ SES: Patients Without low AHRQ SES	Facilities in the Bottom Quartile for % low AHRQ SES: Patients WITH low AHRQ SES	Facilities in the TOP Quartile for % low AHRQ SES: Patients Without low AHRQ SES	Facilities in the TOP Quartile for % low AHRQ SES: Patients WITH low AHRQ SES
Total Post-Acute Care Payments	% of Patients	100.0	100.0	99.9	99.9
Total Post-Acute Care Payments	**$ per Patient**	**$6,199**	**$7,535**	**$8,053**	**$9,126**
Skilled Nursing Facility	# of Patients	64,193	2,458	16,193	10,013
Skilled Nursing Facility	% of Patients	25.9	32.9	26.2	30.5
Skilled Nursing Facility	**$ per Patient**	**$9,352**	**$10,795**	**$11,797**	**$12,979**
Inpatient Rehabilitation	# of Patients	6,512	225	4,663	2,612
Inpatient Rehabilitation	% of Patients	2.6	3.0	7.5	7.9
Inpatient Rehabilitation	**$ per Patient**	**$16,442**	**$16,805**	**$17,291**	**$17,429**
Non-Acute Inpatient Settings	# of Patients	115	8	126	61
Non-Acute Inpatient Settings	% of Patients	0.0	0.1	0.2	0.2
Non-Acute Inpatient Settings	**$ per Patient**	**$9,839**	**$6,055**	**$12,878**	**$15,321**

Table 18: Distribution of payments within the top and bottom quartiles of the facility proportion of patients with low AHRQ SES, for patients with and without low AHRQ SES.

Conclusion

We found that observed payments were higher for patients with the social risk factors that we analyzed: dual eligibility, low AHRQ SES, race (Black), and rural residence. We also found that odds ratios for each social risk factors (including rurality) remained significant, even in a multivariable model that controls for all of the clinical and demographic risk factors in the base model. When comparing impacts on measure scores, while differences in payment were small (less than 0.3%) we found that the dual eligibility variable had the most impact, followed by the low AHRQ SES and race (Black) variables; the rural variable had the least impact on measures scores. In addition, measure scores calculated with and without each social risk factor were highly correlated. Finally, we found that the overall impact on hospitals' classification was small: for example, for the dual eligibility variable, about 3% of hospitals' assignments changed one performance category. We also show (see section 2b.19) that the risk model has good risk prediction for patients with each of the social risk factors.

As described earlier, the version of this measure that went through NQF endorsement previously was adjusted for dual eligibility. However, because the measure is reported together with the THA/TKA Complications measure, which is not adjusted for dual eligibility, CMS implemented public reporting of this measure without adjustment. We also note that this THA/TKA Payment measure is not part of a pay for performance program.

CMS has chosen to account for social risk factors, such as dual eligibility, at the program level within their payment programs and in this case adjusts for dual eligibility within the payment program for lower extremity joint replacement, the Comprehensive Care for Joint Replacement Model or CJR. CJR, run through the Center for Medicare and Medicaid Innovation (CMMI), is a retrospective bundled payment model where CMS provides participant hospitals with a target price prior to the start of each performance year. Following the end of a model performance year, actual total spending for the episode is compared to the target price for the participant hospital where the beneficiary had the initial surgery. Target prices now include beneficiary-level risk adjustment, including adjustment for dual eligibility. Then, depending on the participant hospital's quality and episode spending performance, the hospital may receive an additional payment from Medicare or be required to repay Medicare for a portion of the episode spending.

CMS has further chosen to not adjust this THA/TKA payment measure for race. CMS believes that only in rare cases should quality measures be adjusted for race, due to the unintended consequences of masking disparities.

References

Dummit LA, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, Press MJ, Flood S, Muldoon LD, Gu Q, Hassol A, Bott DM, Bassano A, Conway PH. Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. JAMA. 2016 Sep 27;316(12):1267-78.

Navathe AS, Troxel AB, Liao JM, Nan N, Zhu J, Zhong W, Emanuel EJ. Cost of Joint Replacement Using Bundled Payment Models. JAMA Intern Med. 2017 Feb 1;177(2):214-222.

[Response Ends]

2b.16. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

Previous submission

As is typical with data for healthcare payments, our dependent variable – total payment for a THA/TKA 90-day episode of care – was both right-skewed and leptokurtotic (skewness= 2.5; kurtosis = 13.1). To address estimation problems that can arise with non-normally distributed data, we employed the algorithm suggested by Manning & Mullahy. Using this algorithm and Sample A1 (%50 July 2011-June 2012 sample), we compared several alternative models in order to determine the best estimation approach. Based on these assessments, we chose to estimate a generalized linear model with a log link and an inverse Gaussian distribution.

Approach to Assessing Model Performance (Dataset 1 and Dataset 2)

During model development, we computed four summary statistics for assessing model performance using Dataset 1 randomly split into two samples (A1; 50% sample of July 2011-June 2012) and validation (A2; remaining 50% of July 2011-June 2012) cohorts:

1. Quasi-R-squared
2. Over-fitting indices (Calibration γ0, γ1)

   Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)
3. Distribution of Standardized Pearson Residuals
4. Predictive ratios

Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; good discrimination indicated by a wide range between the lowest decile and highest decile)

As a part of measure reevaluation, each year we assess temporal trends in model performance in the combined 3-year public reporting data (Dataset 2) using the following summary statistics:

1. Quasi-R-Squared
2. Predictive Ratios

Current submission:

As a part of measure reevaluation, each year we assess temporal trends in model performance in the combined 3-year public reporting data (Fall 2022 EM Dataset) using the following summary statistics:

1. Quasi-R-Squared
2. Predictive Ratios

CORE notes that after initial measure development we do not re-test our risk models for overfitting using a dataset that is external to the testing sample. In our risk models, coefficients are updated each time the measure is calculated. Therefore, random statistical fluctuations in model coefficients across repeated reporting cycles are part of the overall random error in the facility performance estimates. CORE believes that this approach is not a validity issue for this type of model, unlike the case of a static risk model.

References

Harrell F.E. and Shih Y.C.T., Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

Manning WG, Mullahy J. Estimating log models: to transform or not to transform? Journal of health economics. Jul 2001;20(4):461-494.

[Response Ends]

2b.17. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

Previous submission

The quasi-R2 results were:

Dataset 1 (Development dataset)

Sample A1 (random 50% sample of July 2011-June 2012) – 0.22

Sample A2 (remaining 50% of July 2011-June 2012) – 0.22

Sample B (full July 2010-June 2011 sample) – 0.23

Dataset 2 (2016 reporting period)

0.23

The predictive ability results were:

Dataset 1 (Development dataset)

Predictive ability (lowest decile %, highest decile %):

Sample A1: 0.99, 1.01

Sample A2: 0.99, 1.00

Sample B: 0.99,1.01

Dataset 2 (2016 reporting period)

Predictive ability (lowest decile %, highest decile %): 0.98, 0.99

Current submission (2020 EM Dataset)

Quasi-R2: 0.18

Predictive ability (lowest decile %, highest decile %): 0.98, 0.98

[Response Ends]

2b.18. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

Previous submission

Over-fitting indices results were:

Dataset 1 (development dataset)

Sample A1 (random 50% sample of July 2011-June 2012) – (0, 1)

Sample A2 (remaining 50% of July 2011-June 2012) – (0.03, 1.00)

Sample B (full July 2010-June 2011 sample) – (-0.11, 1.02)

Standardized Pearson Residuals lack of fit:

Dataset 1 (development dataset)

<-2 = A1 0.01%; A2 0.02%; B 0.02%

[-2, 0) = A1 62.75; A2 62.72%; B 62.17%

[0, 2) = A1 32.02%; A2 32.06%; B 32.65%

[2+ = A1 5.21%; A2 5.20%; B 5.16%

[Response Ends]

2b.19. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

Previous submission

Below is the table of predictive ratios by decile and top 1% of predicted payment for the development and validation samples (Dataset 1):

First Decile: A1 0.99; A2 0.99; B 0.99

Second Decile: A1 1.00; A2 1.00; B 1.00

Third Decile: A1 1.01; A2 1.01; B 1.01

Fourth Decile: A1 1.01; A2 1.01; B 1.01

Fifth Decile: A1 1.01; A2 1.01; B 1.01

Sixth Decile: A1 1.01; A2 1.01; B 1.01

Seventh Decile: A1 1.00; A2 1.01; B 1.00

Eighth Decile: A1 0.99; A2 0.99; B 0.99

Ninth Decile: A1 0.98; A2 0.98; B 0.99

Tenth Decile: A1 1.01; A2 1.00; B 1.01

Top 1%: A1 1.10; A2 1.09; B 1.08

Current submission

Updated analyses for the predicted ratios by decile are shown below.

Decile Predictive ratio

1 .98

2 1.00

3 1.01

4 1.01

5 1.02

6 1.01

7 1.01

8 1.00

9 .99

10 .98

Risk-decile plots for the entire cohort (Figure 4), and for each social risk factor (Figures 4a-4d), are shown below.

Figure 4: Risk-decile plot for the entire THA/TKA Payment cohort

Figure 4a. Risk-decile plot for the patients with dual eligibility

Figure 4b. Risk-decile plot for patients with low AHRQ SES

Figure 4c. Risk-decile plot for Black patents

Figure 4d. Risk-decile plot for rural patients

[Response Ends]

2b.20. Provide the results of the risk stratification analysis.

[Response Begins]

Not applicable.

[Response Ends]

2b.21. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

Previous submission

Quasi-R2

For a traditional linear model (i.e. ordinary least squares regression) R2 is interpreted as the amount of variation in the observed outcome that is explained by the predictor variables (patient-level risk factors). Generalized linear models (GLMs), however, do not output an R2 that is akin to the R2 of a traditional linear model. In order to provide the NQF Committee with a statistic that is conceptually similar, we produced a "quasi-R2" by regressing the total payment outcome on the predicted outcome (Jones 2010). Specifically, we regressed the total payment on the payment predicted by the patient-level risk factors. This regression produced a quasi-R2 of 0.23 (Dataset 2), suggesting that about 23 percent of the variation in payment can be explained by patient-level risk factors. This quasi-R2 is in-line with R2s from other patient-level risk-adjustment models for health care payment (Pope et al. 2011).

Over-fitting (Calibration γ0, γ1)

Over-fitting can result in the phenomenon in which a model describes the relationship between predictor variables and the outcome well in the development sample, but fails to provide valid predictions in new patients. If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting.

Standardized Pearson Residuals

Standardized Pearson residuals also assess model fit. If a substantial number of standardized Pearson residuals exceed 2 in absolute value, lack of fit may be indicated.

Predictive Ratios

A predictive ratio is an estimator's ratio of predicted outcome to observed outcome. A predictive ratio close to 1.0 indicates an accurate prediction. A ratio substantially greater than 1.0 indicates overprediction, and a ratio substantially less than 1.0 indicates underprediction.

Current submission

We continue to see adequate model performance with updated data. In addition, the model performs well (in terms of calibration) for each of the sub-cohorts of patients with social risk factors. This quasi-R2 is in-line with R2s from other patient-level risk adjustment models for health care payment (Pope et al., 2011) and indicates that about 23 percent of the variation in payment can be explained by patient-level risk factors.

Overall Interpretation

Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

References

Ash AS, Byrne-Logan S. How Well Do Models Work? Predicting Health Care Costs. Proceedings of the Section on Statistics in Epidemiology. American Statistical Association. 1998

Jones AM. Models for Health Care. Health, Econometrics and Data Group (HEDG) Working Papers. 2010.

Pope, G. C., Kautter, J., Ingber, M. J., Freeman, S., Sekar, R., & Newhart, C. RTI International, (2011). Evaluation of the CMS-HCC risk adjustment model (Final Report). pp.6.

[Response Ends]

2b.22. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

Not applicable.

[Response Ends]

2b.23. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

Consistent with the other publicly reported measures, we calculate interval estimates for the risk-standardized payment to characterize the amount of uncertainty associated with the payment, compare the interval estimate to the average national payment, and categorizes hospitals as "higher than," "less than," or "no different than" the average national payment (Kim et al. 2014).

Reference

Kim N, Ott L, Lin Z, Zhou S, Keshawarz A, Spivack S, Xu X, George E, Parisi M, Reilly E, Zribi R, Suter L, Krumholz HM. Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 1.0) 2014 Measure Methodology Report. December 2014; Centers for Medicare & Medicaid Services (CMS). Available at: https://qualitynet.cms.gov/inpatient/measures/payment/methodology

[Response Ends]

2b.24. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Previous submission

Dataset 2 (2016 public reporting data)

Between April 2012-March 2013 and April 2014-March 2015, the national mean payment decreased from $23,706 to $22,338 ($2014).

After adjusting for patient case mix, the RSP at the hospital level has a median (interquartile range) of $22,877 ($21,413, $24,576). The mean ± SD risk-standardized hospital payment is $23,135 ± $2,536, ranging from $15,494 to $44,656 across 3,481 hospitals.

Of 3,481 hospitals in the study cohort, 733 (21.06%) had a payment "Greater than the National Payment," 1,087 (32.23%) had a payment "No Different than the National Payment," and 971 (27.89) had a payment "Less than the National Payment." 690 (19.82%) were classified as "Number of Cases Too Small" (fewer than 25) to reliably estimate the hospital's RSP.

Current submission

Table 19 shows the distribution of RSPs across all hospitals and Figure 5 shows a histogram of hospital RSPs.

# Hospitals	3,412
# Admissions	948,457
Mean (SD)	21,121 (2,506)
Min - Max	15,890 - 40,898

# Hospitals	3,412
25th Percentile	19,405
75th Percentile	22,445
1st Percentile	16,799
99th Percentile	28,375
10th Percentile	18,217
20th Percentile	19,066
Obs	1
30th Percentile	19,709
40th Percentile	20,274
Median	20,849
60th Percentile	21,397
70th Percentile	22,023
80th Percentile	22,906
90th Percentile	24,289

Table 19: Distribution of RSPs ($) across all hospitals

Figure 5. Distribution of THA/TKA RSPs among all facilities (Fall 2022 EM Dataset)

Classification of hospital performance

Out of 3,417 hospitals in the U.S., 1,074 performed less than the U.S. national average payment, 1,055 performed no different from the U.S. national average payment, and 622 performed greater than the U.S. national average payment. 666 were classified as number of cases too small (fewer than 25) to reliably tell how well the hospital is performing.

[Response Ends]

2b.25. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

Previous submission

The variation in rates suggests that there are meaningful differences across hospitals in risk standardized payments associated with a 90-day episode of care for patients undergoing elective primary THA/TKA.

Current submission

With updated data there continues to be variation in rates showing meaningful differences across hospitals in risk-standardized THA/TKA payments.

[Response Ends]

2b.26. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

Not applicable.

[Response Ends]

2b.27. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

Not applicable.

[Response Ends]

2b.28. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

Not applicable.

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.29. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.30. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.31. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.32. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

Feasibility

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in a combination of electronic sources

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

This is a claims-based measure, and there has been no difficulty in obtaining the needed data to calculate this measure.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]

There are no fees associated with the use of this measure.

[Response Ends]

Usability and Use

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

Public Reporting

[Public Reporting Please Explain]

Previous Submission

Public Reporting

Program Name, Sponsor: Hospital Inpatient Quality Reporting (Hospital IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital IQR program was originally mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. Initially, the MMA provided for a 0.4 percentage point reduction in the annual market basket (the measure of inflation in costs of goods and services used by hospitals in treating Medicare patients) update for hospitals that did not successfully report. The Deficit Reduction Act of 2005 increased that reduction to 2.0 percentage points.

In addition to giving hospitals a financial incentive to report the quality of their services, the Hospital IQR program provides CMS with data to help consumers make more informed decisions about their health care. Some of the hospital quality of care information gathered through the program is available to consumers on the Hospital Compare website at: www.hospitalcompare.hhs.gov.

Geographic area and number and percentage of accountable entities and patients included:

The Hospital IQR program includes all Inpatient Prospective Payment System (IPPS) non-federal acute care hospitals and VA hospitals in the United States. The number and percentage of accountable hospitals included in the program and the number of patients included in the measure vary by reporting year. For 2016 reporting data, the RSP was reported for 2,791 hospitals across the U.S. The final index cohort includes 887,061 admissions.

Current Submission

The THA/TKA payment measure continues to be reported through the Hospital IQR program and on *Care Compare*, the successor website to *Hospital Compare*.

The IQR Program is a pay-for-reporting program. Hospitals eligible for the Hospital Inpatient Quality Reporting (IQR) Program are included annually in one of three lists:

Hospitals receiving full Annual Payment Update (APU) - hospitals that satisfactorily met the requirements for the Hospital IQR Program. These hospitals will receive the full annual market basket update.

Hospitals not receiving full APU - hospitals that did not satisfactorily meet the criteria for the Hospital IQR Program. These hospitals will receive their annual market basket update with a reduction by one-fourth of the applicable market basket update.

Hospitals choosing not to participate - hospitals that actively chose not to participate. These hospitals will receive their annual market basket update with a reduction by one-fourth of the applicable market basket update.

Updated URLs for information about the measure include:

Hospital IQR Program page (CMS): https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU

Hospital IQR Program page (for hospitals): https://qualitynet.cms.gov/inpatient/iqr

Care Compare: https://www.medicare.gov/care-compare/

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Measure Currently in Use

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

N/A. This measure is currently in use.

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

N/A. This measure is currently in use.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

Previous Submission

The exact number of measured entities (acute care hospitals) varies with each new measurement period. In 2016, 3,481 hospitals were included in the measure calculation, and results were provided confidentially in hospital-specific reports (HSRs) (Simoes et al., 2016). In 2017 and 2018, the results of this measure were publicly reported and included in the HIQR payment determination for all the nation's non-federal short-term acute care hospitals (including Indian Health Services hospitals) and critical access hospitals. The most recent publicly reported measure results included all hospital discharges with a primary elective THA/TKA procedure between April 2014 and March 2017.

Each hospital receives its measure results in April of each calendar year through CMS's QualityNet website. The results are then publicly reported on CMS's Hospital Compare website in July of each calendar year. Since the measure is risk-standardized using data from all hospitals. Hospitals cannot independently calculate their score.

However, CMS provides each hospital with several resources that aid in the interpretation of their results (described in detail below). These include Hospital-Specific Reports (HSRs) with details about every patient from their facility that was included in the measure calculation (for example, dates of admission and discharge, discharge diagnoses, outcome [died or not], transfer status, and facility transferred from). HSRs provide detailed payment information for index hospitalization stays and post-acute care. These reports facilitate quality improvement (QI) activities such as review of individual deaths and patterns of deaths; make visible to hospitals post-discharge outcomes that they may otherwise be unaware of, and allow hospitals to look for patterns that may inform QI work (e.g., among patient transferred in from particular facilities).

The Hospital-Specific Reports (HSRs) also provide hospitals with more detailed benchmarks with which to gauge their performance relative to peer hospitals and interpret their results, including comorbidity frequencies for their patients relative to other hospitals in their state and the country.

Additionally, the code used to process the claims data and calculate measure results is written in SAS (Cary, NC) and is provided each year to hospitals upon request to make the measure methodology completely transparent.

Reference

Ott L, Kim N, Hsieh A, et al. 2016: 2016 Measure Updates and Specifications Report Hospital-Level Risk-Standardized Payment Measures. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228774789978

Current Submission

While this measure has already been implemented, CMS continues to follow the process described above, providing measured entities with Hospital-Specific Reports (HSRs) that provide detailed measure results, discharge-level data, and state and national results. The most recent preview period, during which hospitals can review their HSRs prior to public reporting, was from May 16 to June 15, 2022, for July 2022 public reporting. CMS provides information on how to interpret the results through an HSR User Guide and Video and access to the Q&A website. The URL for this page is: https://qualitynet.cms.gov/inpatient/measures/payment/reports

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

Previous Submission

In April of each year, hospitals have access to the following list of updated resources related to the measure, which is provided directly or posted publicly for hospitals to use:

1. Hospital-Specific Reports (HSR): available for hospitals to download from QualityNet in April of each calendar year; includes information on the index admissions included in the measure calculation for each facility, detailed measure results, and state and national results.
2. HSR User Guide: available with the HSR and posted on QualityNet; provides instructions for interpreting the results and descriptions of each data field in the HSR.
3. Mock HSR: posted on QualityNet; provides real national results and simulated state and hospital results for stakeholders who do not receive an HSR.
4. Inpatient Quality Reporting (IQR) Preview Reports and Preview Report Help Guide: available for hospitals to download from QualityNet in April of each calendar year; includes measure results that will be publicly reported on Hospital Compare.
5. Measure Updates and Specification Reports: posted in April of each calendar year on QualityNet with detailed measure specifications, descriptions of changes made to the measure specifications with rationale and impact analysis when appropriate, updated risk variable frequencies and coefficients for the national cohort, and updated national results for the new measurement period.
6. Frequently Asked Questions (FAQs): includes general and measure-specific questions and responses, as well as infographics that explain complex components of the measure's methodology, and are posted in April of each calendar year on QualityNet.
7. The SAS code used to calculate the measure with documentation describing what data files are used and how the SAS code works. This code and documentation are updated each year and are released upon request beginning in July of each year.
8. Measure Fact Sheets: provide a brief overview of measures, measure updates, and are posted in April of each calendar year on QualityNet.

In July of each year, the publicly-reported measure results are posted on Hospital Compare, a tool to find hospitals and compare their quality of care that CMS created in collaboration with organizations representing consumers, hospitals, doctors, employers, accrediting organizations, and other federal agencies. Measure results are updated in July of each calendar year.

Current Submission

CMS uses the same process to provide measure results. The time of the preview period may differ with each reporting cycle but occurs prior to July 2022 reporting. CMS has updated its public reporting website, which is now called Care Compare: https://www.medicare.gov/care-compare/

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

Previous Submission

Questions and Answers (Q&A)

The measured entities (acute care hospitals) and other stakeholders or interested parties submit questions or comments about the measure through an email inbox (cmsepisodepaymentmeasures@yale.edu). Experts on measure specifications, calculation, or implementation, prepare responses to those inquiries and reply directly to the sender. We consider issues raised through the Q&A process about measure specifications or measure calculation in measure reevaluation.

Literature Reviews

In addition, we continually scan the literature for scholarly articles describing research related to this measure. We summarize new information obtained through these reviews every three years as a part of comprehensive reevaluation as mandated by the Measure Management System (MMS) Blueprint.

Current Submission

CMS continues to collect feedback from users using the Q&A tool, which can be found here: https://cmsqualitysupport.servicenowservices.com/qnet_qa

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

Previous Submission

Summary of Questions or Comments from Hospitals submitted through the Q & A process

For the Hip/Knee payment measure, we have received the following inquiries from hospitals since the implementation of the measure maintenance in July 2018:

1. Requests for detailed measure specifications, including ICD-9 and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
2. Requests for the SAS code used to calculate measure results;
3. Queries about how to access hospital preview results for hip/knee payment measure;
4. Requests for updates on detailed measure specifications, including measure inclusion criteria.

Current submission

Summary of Questions or Comments from Hospitals

Since January 1, 2016, feedback from hospitals included the following topics:

1. Requests for SAS code used to calculate measure results
2. Requests related to Hospital-Specific Reports (HSRs)
3. Data sources
4. COVID-19
5. Definition of the outcome
6. Statistical methods around performance categories
7. Statistical methods around risk adjustment

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

Previous Submission

Summary of Questions and Comments from Other Stakeholders

For the Hip/Knee payment measure, we have received the following inquiries from other stakeholders since the implementation of the measure maintenance in July 2018:

1. Requests for detailed measure specifications, including ICD-9 and ICD-10 codes, used to define the measure cohort or in the risk-adjustment model;

2. Requests for the SAS code used to calculate measure results;
3. Queries about how to calculate the measure and interpret the results;
4. Requests for hospital-specific measure information, such as data included in the HSRs;
5. Requests for data on particular hospitals.

Current Submission

Summary of Questions and Comments from Other Stakeholders

From January 1, 2016, through September 2022, CMS received 33 questions from other stakeholders, including the following topics:

1. Requests for publicly available materials, including SAS packs
2. Cohort definition
3. Outcome definition
4. National rates/benchmarking
5. Data sources
6. Present-on-admission (POA) codes

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

Feedback from stakeholders (in addition to other input, including published literature, and updates to code sets) is considered within each cycle of measure re-evaluation. In this past cycle, as described in section spma.02, updates to the measure include coding updates, and removal of the dual eligibility variables.

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

As shown in Table A in section 1b.02, year-over-year performance on THA/TKA measure scores (risk-standardized payments or RSPs) improved across the entire distribution each year. For example, the 10th percentile was $19,007 in 2016/2017, compared with $18,346 in 2017/2018 and $17,957 in 2018/2019; the 50th percentile was $21,695 in 2016/2017, compared with $20,725 in 2017/2018 and $20,071 in 2018/2019; the 90th percentile was $24,911 in 2016/2017, compared with $23,708 in 2017/2018 and $22,827 in 2018/2019.

Performance on the THA/TKA payment measure has also improved when comparing the 2016/2019 reporting period with the 2012/2015 reporting period. For example, mean RSPs were $23,146 in 2012/2015 compared with $20,314 in 2018/2019; median RSPs were 22,844 in 2012/2015 compared with $20,017 in 2016/2019.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

Previous Submission

We did not identify any unintended consequences during measure development and testing. We are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care or coding/billing practices, increased patient morbidity and mortality, and other negative unintended consequences for patients.

Current Submission

CMS has not identified any unintended consequences during measure implementation.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

N/A

[Response Ends]

Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)

1551: Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

Related measure: Elective Primary Hip Arthroplasty episode-based cost measure within the Merit-Based Incentive Program (MIPS).

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

N/A

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

Previous Submission

N/A

Current Submission

There are no competing measures.

[Response Ends]

Contact Information

Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Measure Steward Point of Contact: Dollar-Maples, Helen, helen.dollar-maples@cms.hhs.gov

Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

Measure Developer Point(s) of Contact: Peter, Doris, doris.peter@yale.edu

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

Available in attached file

[Response Ends]

Attachment: 3474_3474_2020 Measure Updates and Specifications Report_HKPayment-508.pdf

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

Technical Expert Panel Members:
Blair Biase, MMSc, PA-C, MBA
Global Knee Reconstruction, OrthoSensor, Inc.

John Birkmeyer, MD
University of Michigan, Department of Surgery

Kate Chenok, MBA
Pacific Business Group on Health

Cheryl Crumpton, MS, RN, CEN
Cheyenne Regional Medical Center

Vinod Dasa, MD
Louisiana State University Health Sciences Center: Adult
Reconstruction and Sports Medicine; Ochsner Kenner Medical Center

Cheryl Fahlman, PhD, MBA, BSP
Premier Healthcare Solutions, Inc.

Vivian Ho, PhD
Rice University, Department of Economics

David Hopkins, PhD
Pacific Business Group on Health

Cynthia Jacelon, PhD, RN, CRRN, FAAN
University of Massachusetts School of Nursing

Brian McCardel, MD
Sparrow Health System, Orthopedic Surgery Section

Derek Nordman, MPT, ATC
Gentiva Health Services

Amita Rastogi, MD, MHA, CHE, MS
Health Care Incentives Improvement Institute (HCI3)

Jonathan Schaffer, MD, MBA
The Cleveland Clinic Foundation: Department of Orthopaedic Surgery, Information Technology Division

Kathleen Willhite, MS
BayCare Health Systems

AJ Yates, MD
University of Pittsburgh School of Medicine, Dept. of Orthopaedic Surgery

Anonymous Patient

Working Group Member:
Kevin Bozic, MD, MBA Professor and Vice Chair
University of California, San Francisco
Department of Orthopaedic Surgery
Core Faculty, Philip R. Lee Institute for Health Policy Studies

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

2019

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

Fall 2022

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

Annually

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

Fall 2023

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]