NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

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 (CMS)

Brief Description of Measure: 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.

Developer Rationale: 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 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 to hospitals that is not otherwise possible.

Measure Type: Cost/Resource Use

Data Source: Claims, Other

Level of Analysis: Facility

New Measure Submission

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. High impact or high resource use:

The measure focus addresses:

– a demonstrated high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality).

AND

1b. Variation in cost or resource use**:**

Demonstration of resource use or cost problems and opportunity for improvement, i.e., data demonstrating

- Considerable variation cost or resource across providers; and/or
- Disparities in care across population groups

1a. High Impact or high resource use.

- The focus of this measure is *elective* hip and knee arthroplasty.

1b. Variation in cost or resource use.

- The developer examined the distribution of hospital payment scores to demonstrate the variation in payment among hospitals.
- The developer provided data demonstrating variation in terms of risk standardized payment across providers. For example, in the 2012-2013 period payment ranged from $16,421 to $35,123 with the median payment of $23,120. This range of performance was similar or greater across the other years in the data sample.
- The median hospital RSP in the combined three-year dataset was $22,408 (IQR $21,134 - $24,174)
- Of the 3,481 hospitals in the developer cohort, 21.06% of the hospitals had a payment "greater than the national payment"; 32.23% had a payment "no different from the national payment"; 27.89% had a payment "less than the national payment". 19.8% of hospitals had too few cases to reliably estimate the hospital risk standardized price.

1b. Disparities across populations.

- Hospitals with a low proportion of dual eligible patients (3.8%) had lower median risk standardized payments ($21,925) compared to hospitals with a high proportion (11.5%) of dual eligible patients ($23,974).
- Similiarily, hospitals with a low proportion of patients below the AHRQ SDS index score of 42.7 had lower median risk standardized payments ($22,110) compared to hospitals with a high proportion proportion of patients below the AHRQ SDS index score of 42.7 ($23,501).

Questions for the Committee:

- Has the developer demonstrated this is high impact, high-resource use area to measure?
- Is there a sufficient variation in performance across hospitals that warrants a national performance measure?

Staff preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b)

1a. High Impact or High Resource Use

Comments:

** yes

** yes, it the two procedures are commonly performed, affecting a large number of beneficiaries

**Yes
** yes, large component of Medicare hospitalizations and one of the most frequent surgical procedures.
** yes
** yes
** It addresses a sometimes elective procedure for a large number of people
** Yes
** Yes
** Yes
** To a point yes. By focusing on the cost of elective THA/TKA procedures, in addition to other measures that assess care coordination (readmissions) and complications (safety) it is unclear whether this particular measure's focus is necessary

1b. Performance Gap

Comments:

** Data was provided but interpretation was not provided. There is variability but the developer did not make a case for gaps - given the argument is to be able to look at the cost/quality tradeoff, it would have been helpful to see that data.
** Developer finds higher episode payments among hospitals with higher fraction of duals vs. non-duals. There is variation in total payments across facilities, though the interquartile range runs from $21K to $24K, so not huge
** Yes. Variability in RSP noted.
** The middle of the hospital-level distribution (quartiles) differ from the median by less than +/- 10%; however, the highest hospital RSP was more than twice the lowest.
** Variability - although it would be difficult to assess whether higher than average payment is better or worse without complementary quality data
** There is substantial variability in cost. Comparing cost with quality will allow for a value determination, important to people making a choice for this elective procedure.
** There is some variation in payment, but not as much as I expected. In fact, payments to hospitals with highest level of duals were highest- I would expect that as these are more complicated patients
** Yes. The data prosented shows significant variability in cost for the same 2 services across providers. This is consistent with commercial payor data for my regional market.
** Wide variability in standardized costing, driven by variations in post-hositalization use of care
** Uncertain
**The data presented demonstrates a gap in costs, but does not necessarily address a gap in care as some of those differences in cost are not in control of the hospital being measured

Disparities

Comments:

** Also, the disparity argument is not well formed - what does it mean that the costs are higher in hospitals with more dual eligibles? Not all differences are disparities. They needed to make the argument for the disparity. Data is not equal to information
** Information is provided on differences in payments based on the % of duals at the hospital, with hospitals with lower % duals showing lower episode payments
** Yes. Disparities noted by dual eligibilty and AHRQ SDS Index scores.
** hospitals with more dual eligibles or with more patients with lower SES scores had higher median RSPs
** Yes, relating to duals status and AHRQ SDS index
** Social risk factors are cited but dual eligibility is the status is the data used.
** It doesn't
** The developers used dual eligible patients and SDS index patients as proxies for disparity, but the data imply that lower SDS staus correlates with lower costs. Their data demonstrate difference in cost, but lower cost for the same service is not a disparity if the quality is comparable. Quality of the service is not addressed in this section.

** Yes. Low SES patients have distinct use patterns.
** To some extent
** Data demonstrated differences in cost in hospitals based upon proportion of duals served, but I'm not sure that it showed disparities in care. I would be more interested in data showing differences in costs based upon co-morbidities, since this is an elective procedure and the difference in costs usually about whether a patient needs inpatient rehabilitation services or outpatient rehabilitation services.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

2b. Validity: Alignment of Specifications with Intent (includes threats to validity [e.g., attribution, costing method, missing data**])** Testing**;** Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Multiple Data Sources**; and Disparities.**

Measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: Jen Perloff, Ron Walters, Susan White, Jack Needleman **(**Evaluation A: Methods Panel**)**

Methods Panel Individual Reliability Ratings: H-3, M-0, L-0, I-1
Methods Panel Individual Validity Ratings: H-2, M-2, L-0, I-0

Measure evaluated by Technical Expert Panel? ☒ **Yes** ☐ **No**

Evaluators: Timothy Henne, Bryan Little, Anthony Mascioli, Kimberly Templeton **(**Evaluation B: Technical Expert Panel)

Reliability

2a1. Specifications:

The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. All measures that use the ICD classification system must use ICD-10-CM.

2a2. Reliability testing:
Demonstration that 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 enough to distinguish differences in performance across providers.

2a2. Reliability Testing:

- Data Element:
  - The developer stated that the measure is calculated from claims which are adjucated by CMS. No formal testing or results were provided to demonstrate data element reliability. This does not meet NQF requirements for data element reliability testing.
- Score-Level:
  - Reliability of measure scores across hospitals was tested using the split sample test-retest method and interclass correlation coefficient (ICC) to determine the agreement of samples:
    - Index admissions were combined 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. As a metric of agreement, the intra-class correlation coefficient (ICC) was calculated.

        - The agreement between the two independent assessments of each hospital was 0.931. When 3 years of data is used the median reliability score is 0.938.
    - Hospitals with less than 25 cases are excluded from reliability assessment. Rates for these "small volume" hospitals are reported separately.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- Do you have any concerns with the reliability testing that was not identified by the Scientific Methods Panel?

Staff Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2a: Reliability

2a1. Reliability – Specifications

Comments:

** Data elements are clearly specified. One possible issue is the dual variable and whether the variable being used is the best variable. CMS has many (annual, month). It is clear that this measure can be consistently implemented, thought the exclusion criteria list is quite lengthy so requires programming skills to correctly implement.
** 1. ICD classification system should be consistent. 2. Exclusion of Part D costs seems to be pragmatic approach since not all Medicare beneficaries have Part D.
** as it is based on Medicare FFS claims data, it appears that the measure can be implemented consistently.
** No concern
** Reliability score is good
** No real concerns here
** No concerns in this area.
** Reliability moderate. All data clearly defined. Inclusion and exclusion codes should be reviewed by substantive experts.
** The measure is well defined. The codes are clear. The methods are well developed and presented. The measure will be straight-forward to implement.
** Concerned that there was no documented formal testing of reliability

2a2. Reliability – Testing

Comments:

** No
** Happy to see measure developer computed statistic on hospital-level reliability--mean score seems high. would be helpful to see the distribution of reliability scores
** 1. Unclear why 25 cases as the cutoff. 2. There is also an underlying assumption that the reliability of testing at the data element level is high due to the inherit process of claims data; albeit not tested.
** no, data element reliability relies on previous work; score-level intraclass correlation based on split samples was very high.
** No
** No
** none, but only useful for patent outliers
** Yes. I am concerned that reliability testing was not done.
** No.

** The measure has a modest pseudo R squared, meaning modest ability to adjust for confounders. The measure should work reasonably well, with reasonable confidence that is it measuring what it is suppored to measure.
**no

Validity

2b1. Specifications align with measure intent:

The measure specifications are consistent with the measure intent and captures the most inclusive target population.

2b2. Validity Testing:

Demonstration that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided.

2b3. Exclusions:

Exclusions are supported by the clinical evidence**,** AND/OR There is a rationale or analysis demonstrating that the measure results are sufficiently distorted due to the magnitude and/or frequency of then on-clinical exclusions; AND Measure specifications for scoring include computing exclusions so that the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion); 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).

2b4. Risk Adjustment:

For resource use measures and other measures when indicated: an evidence-based risk-adjustment strategy is specified and is based on patient factors (including clinical and sociodemographic risk factors) that influence the measured outcome and are present at start of care, and has demonstrated adequate discrimination and calibration, OR rationale/data support no risk-adjustment/-stratification.

2b5. Meaningful Differences:

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

2b6. Multiple Data Sources:

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

2c. Disparities: If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender), OR rationale rationale/data justifies why stratification is not necessary or not feasible.

2b1. Specifications Align with Measure Intent:

- The TEP expressed concerns about the clinical sites from which costs are captured in the measure. The TEP specifically questioned the inclusion of birthing centers in the list of included clinical sites as this does not seem relevant for this population (ages >65). There were also concerns with including costs of various other facilities as it may capture costs that are unrelated to the procedure as these patients will likely be high cost.
    - *Developer response*: Patients in the settings listed are not a priori included in the denominator unless they are subsequently admitted to an inpatient acute care facility for an elective TKA/THA procedure. The developer explained that the conditions that may be associated with the setting (e.g., psychiatric conditions for patients admitted from a psych facility) are included

in the risk model in order to adjust at the patient-level for co-occuring coditions that may impact outcomes. The measure design is Intended to enable capture of complications that may occur in those settings after discharge from the acute care facility. The developer pointed out that the range of settings and type of costs captured in the measure narrows during the episode. All costs are captured in all post acute settings in the first 30 days after discharge. Costs captured during days 31-90 are for a narrowed list of settings and events that defined as being related to TKA/THA procedures.

- The TEP sought clarity on the inclusion of readmissions in the measure and ED costs during the episode.
    - *Developer response*: All costs are captured in the first 30 days (this would include a pulmonary embolism (PE) or deep vein thrombosis (DVT)). After 30 days, only complications related to wound infection, surgical site infection, bleeding or mechanical issues are captured. This is in alignment with a harmonized THA/TKA complications readmissions measure and requires both diagnostic and procdure codes for the costs to be counted in the measure.
- The TEP stated specific concerns about including costs of homeless shelters and prisons in the first 30 days as these settings are often associated with costs to managing health issues related to sequelae of complex social issues that are often unrelated to the procedure itself. There are concerns that the social factors related to these types of patients are not accounted for in the risk model.
    - *Developer response:* The costs that are captured for patients to are in prison in the 30 days after discharge are limited to physician costs, and do not include the cost of prison itself. The developer stated that these patients would only account for a small percentage of the populatin included in the measure and would have little impact on overall hospital performance.
- Attribution:
    - This measure is attributed to the hospital. Analysis (table on page 58) indicates most variation in payments occur after hospitalization in post acute settings. This attribution approach was selected in order to drive hospitals to facilitate care coordination, assess referral practices, and understand their role in post-acute costs and resource use.
- Costing approach:
    - The costing approach is based on payments by Medicare to hopitals for services within the identified resource use service categories. Payments are based on agreed upon fee schedules for each setting.

2b2. Validity Testing:

- The developer cites multiple approaches to demonstrating validity; however, face validity that is systemtically assessed by a TEP is the only approach that meets NQF testing standards. Of the 13 member TEP, 5 “strongly agree” that the measure is a valid assessment, 6 “moderately agree”, and 2 “somewhat agree”. The highest possible rating for a measure with face validity testing is moderate.

2b3. Clinical Inclusions and Exclusions/Evidence to Support Clinical Logic

- The TEP expressed concerns that Medicare patients less than 65 not included in the measure. Diabled patients and those with ESRD are eligible for Medicare before the age of 65.
    - *Developer Response*: There are separate measurement programs for ESRD patients under the age of 65 so they are excluded. They are excluded due to the difficulty in capturing baseline functional status prior to elective procedure using administrative data. It is also very difficult to capture and accommodate these varying levels of functional status and levels of disability in the risk model. This approach is consistent with other CMS measures.
- The TEP sought clarity on how pathological fractures were handled in order to determine whether the patient should be excluded from the measure or if the fracture was related to the current episode. The

developer clarified that present on admission code modifiers are used to discern whether the fracture was acquired before or during admission. There was some discussion as to whether this modifier is consistently used and can be relied upon as an accurate method for identifying these patients, but it was ultimately deemed satisfactory by the TEP.

2b4/2c. Risk adjustment

- The developer's literature review of the impact of SES for hip and knee arthroplasty patients found that non-home discharge destinations are associated with higher costs. The literature also indicated those with social risk factors demonstrate longer lengths of stay and higher rates of readmissions.
- Two variables were used as proxy for SES in the analysis of the risk model: Dual Eligible status and AHRQ-validated SES index score
- Each of the SDS factors remained statistically significant in the multivariate models (1.12 for dual-eligibility and 1.04 for the AHRQ SDS index). The developers also tested whether there is a differences in the quasi-R square with the inclusion of these factors but found a negligible change. Given the variation in post-acute spending for the different subgroups and the results of their empirical analysis, the developer included the dual eligibility in the risk model.
- The TEP expressed concerns with using the cost measure as a proxy for complications and emphasized that complications are not the only sequalae of these procedures; functional status, and patient reported outcomes should also be examined. The TEP questioned whether using claims data for hip and knee arthroplasty patients to determine risk profile for the risk model has it been validtated in this popuation.
    - *Developer response*: CMS in the process of working on measures related to functional status and patient reported outcomes for this population. There are also hip/knee complications readmission measures that are endorsed, harmonized and in use. To date, they have not received any feedback on similar measures related to the validity of the risk adjustment approach. Face validity was sought from a clinical TEP consulted during the development of the measure. The developers acknowledged that administrative claims are not a proxy for clinical data, but when aggregated at the hospital level can be used to predict risk.

2b5: Meaningful Differences

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- What is the Committee's assessment of the inclusion of dual-eligibility in the risk model?
- Is there any concern with outliers, missing data (e.g., pharmacy [Part D] data)?
- Is there any concern with the risk adjustment model which includes patient-level, clinical factors relative to inpatient episode when the variation in cost is primarily in the post acute portion of the episode? Is this a valid approach to risk adjustment?

Staff preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2b: Validity

2b1. Validity –Testing

Comments:

** I am bothered by the notion that a driving factor in cost is the setting for rehabilitative care. The fact that patients with less favorable socioeconomic status have consistently higher cost related to location of post surgical care can't just be risk adjusted away. It's not clear that rating hospitals in a way that is influenced

heavily by socioeconomic status of the patient is reasonable. Also, waving off claims data as valid because it has been used before is not acceptable. Those of us who have worked extensively with claims data know that some data elements are less reliable and those should be identified.
** TEP found moderate to high face validity of measure. Measure developer conducted literature review to identify social risk factor evidence (which is limited) on effects on payments
** 1. Measure intents to look at complications, rather than a composite measurement of quality of care (complication is one factor). 2. Variability in the complex dynamics between social factors of patients (not included), sites included in the model (prisons, shelters, hospitals providing care to patients with more complex non medical issues) especially considering that most variation in payments happen in post acute setting.
** no major concerns with face validity testing; if joint replacements continue to move to the outpatient setting, more of the target population will be missed.
** No
**This measure scored relatively well when evaluated by the TEP
** none
** I agree with the TEP's concerns regarding which costs are included in the first 30 days vs the subsequent 60 days, how costs for homeless and incarcerated patients are counted, and attribution to hospitals but not to LTAC's or SNF's. While I understand their intent to pressue hospitals to change behavior, the significant impact of post-acute care settings is obscured by this approach.
** Inclusion and exclusion codes should be reviewed by substantive experts. Ideally, would like to see quantification of impact of odd services in first 30 days noted by TEP on relative scores and ranking.
** The validity testing was well done. The measure is consistent with the intent. The target population is well captured. I have no concerns with the validity testing. The measue uses appropriate elements and does reflect cost of care. Note that the pseudo R-squared is modest, althought higher than for other cost models.
** echo concerns with inclusion of costs unrelated to procedure in the 90 day episode (prison, homeless shelters, etc.); why are birthing centers included in measurement if the measure is limited to Medicare patients 65+

2b4-7. Threats to Validity: Meaningful Differences

Comments:

** it is not clear that measuring 90 day charges attributable to a hospital is meaningful for an accountability program. If would be helpful to have a discussion with the developer about what variables are driving the differences in cost to determine whether or not those are variables that we empower hospitals to control. The complications measrue is valuable but not necessarily more valuable in tandem with a charges measure.
** Mean reliability appears high. would be helpful to see full distribution
** As above.
** 21% of hospitals deemed to be "Greater than national payment"
** Very high ICC score - appears precise and able to capture meaningful differences
** There still appears to be some concern about attribution and exclusions
** I don't think that the payment differences among accountable units are necessarily meaningful. Useful to identify real outliers
** I share the TEP's concerns.
**Classification of above and below national average is standard for this type of measure.
** There is variability in cost. The measure will be able to identify this successfully.
** I think there is a real concern that this measure will not be able to help hospitals better determine which types of PAC services are appropriate for elective THA/TKA patients. There really isn't good data in this area for decision-making, and focusing on cost too early might hinder that work

2b4-7. Threats to Validity: Missing Data/Carve-outs

Comments:

** There is no discussion of the impact of missing pharmacy data though I suspect that pharmaceutical cost is not a major driver or costs for this measure. Pharmaceutical costs may help (to identify co-occuring conditions)

or may make the measure less useful (adding to variability in cost that is not attributable to the surgery or after care)
** Major issue flagged is part D expenses. Developer can't fix this problem.
** In addition to above, patients less than 65 years are excluded which raised some valid concerns, but would not likely be a significant threat to the validity.
**pharmacy data would be informative, but is likely not the greatest difference between hospitals.
** No concern
** This measure uses claims data. Clinical data would be preferred and patient reported outcomes would enhance the overall assessment
** none
** No. Pharmacy costs compose a relatively low proportion of the cost of TKA and THA compared to inpatient and rehab costs.
** None.
** The data are remarkable complete. There are no troublesome carve outs.
** no comments

2b2. Additional threats to validity: attribution, the costing approach, or truncation

Comments:

** it is not clear that hospitals have control over the location of after care; need to know how that continues to drive the variation even with risk assessment
** A 90 day episode encompasses a large time period for accountability. The hospital has reasonable control over the costs (after factoring in social risk factor proxy of dual status) for the inpatient portion, and the idea of an episode payment measure is to promote coordination. Not sure why the attribution doesn't expand beyond the hospital to other actors involved in the episode.
** None additional.
** most of the variability occurs after hospital discharge; surgeons and hospitals may influence where post-acute treatment occurs, but may have less influence on variability within those sites.
** I have no problem with the attribution model. Hospitals have the ability to direct patients to appropriate post-acute settings and providers. They have the ability to influence post-acute care providers on quality. Ideally, all stakeholders should be held accountable, I also believe that holding hospitals as the accountable entity is not a threat to validity.
** This is the best that can be done based on currently available data sets
** no concerns
** I am concerned that attribution focuses the measure's attention on acute care hospitals but not to LTAC's or SNF's. While I understand their intent to pressue hospitals to change behavior, the significant impact of post-acute care settings is obscured by this approach.
** Usual issues of attribution and standardized costing for this class of measures, but these have been accepted in the past.
** Attribution is at the hospital level and appropriate. The entity has some although not complete control over costs and resources. The measure is intended to drive change. The costing approach is appropriate. There is no truncation as patients with critical missing data are excluded.
** It seems quasi aspirational in that it is seeking to add a measurement of cost as a proxy for measuring complications (which is measured separately); similarly, the measure developer says it is measuring the hospitals as the care coordinators, though the intent of the separate readmissions measure is intended to measure a hospital's care coordination, so it's unclear what role this measure plays independent of the two current measures considering that the variation in cost is in PAC and not hospital setting

2b2. Additional Threats to Validity: Exclusions

Comments:

** yes, consistent though the list of included sites of care for the index admission seem a bit broad. This is not likely to be impacting the measure much due to low volumes
** Concern about excluding disabled patients

** As above.
** no
** No as I read it.
** Exclusions are appropriate
** no concerns
** Their exclusion of ESRD patients is reasonable.
** exclusions consistent and appropriate.
** There are no inappropraite exclusions.
** Would like more discussion about Medicare patients under age 65 not being included in the measure

2b2. Additional Threats to Validity: Risk Adjustment

Comments:

**yes, the math seems appropriate but the SES issue is not adequately addressed
**The developers tested dual status and AHRQ SES measure and elected to go only with dual status, which is significantly associated with performance on the measure. . while the developer states the variables in the risk adjustment model, the model results are not shown. Variables selected look reasonable, but there are a large number of variables and I'm wondering if there is a more parsiomoniuos model and what the model fit statistics are.
**As above.
**the only risk factor not necessarily present at start of care is the staged bilateral procedure. Patient preference may influence whether it occurs within 90 days of the index case. Risk adjustment approach is acceptable. Adjustment for dual eligibility appears appropriate. It is not clear whether both social risk factors were consider simultaneously or separately.
**I am satisfied with the risk adjustment using duals status.
**Yes, the risk adjustment approach appears appropriate
**no concerns
**It is not clear to me that the measure's results where risk adjusted. The decision to include dual eligible patients is reasonable.
**Agree with Scientific Methods Panel assessment.
**The social risk factors were well tested, and had little effect. The risk factors are available at the start. The risk adjustment was well developed. The results are acceptable. The strategy is appropriate.
**Concern that social risk factors are not included, considering impact on cost over 90 days and whether the clincial risk adjustment is able to capture the differences in a lifelong runner coming in for a knee replacement vs a patient with multiple co-morbidities

Evaluation A: Methods Panel

Measure Number: 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)

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☒ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☒ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☒ **New** ☐ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**
   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

   Panel Member #1: Yes the submitted specs are quite clear as to the involved population, inclusions and exclusions, the definition of the episode, the data sources, and the analytics involved. Code sets are provided for the claims data. The risk adjustment methodology is provided.

2. **Briefly summarize any concerns about the measure specifications.**
   Panel Member #1: I have no concerns.

   Panel Member #2: The specifications provide a conceptual description of the risk adjustment model, but not enough detail to reproduce the model. The authors indicate that they are willing to share their SAS code with those who want to apply the measure, but the MIF should include the equation being estimated. Not everyone has access to SAS. Also, the measure is really Medicare fee-for-service Hospital-level, risk standardized payment, not a generalized measure of THA/TKA resource use. Consider clarifying the name.
   Panel Member #3: This measure uses the general framework for the CMS episode of care measures that are triggered by a hospitalization. It uses standardized pricing. Key concerns:
   a. Measure has two sets of included services, everything in first 30 days post-hospitalization, and what are considered condition specific services in days 31-90. Day 31-90 services look reasonable. Days 0-30 include services that are unrelated to the surgery such as mass immunization or birthing center (unlikely to be a service for the 65+ population). The 0-30 services are clearly inflated beyond those related to the elective surgery but the extent is probably minimal once risk adjustment for other conditions that might require care is implemented.
   b. Substantive experts need to assess the procedures that are included in the measure and the exclusions that reflect injury or other trauma, etc. The ICD-9 and 10 codes relevant to these inclusions and exclusions are clearly specified.
   c. Exclusion of Part D drug costs. Inherent limitation of these measures, given not all Medicare beneficiaries have Part D. No major concerns that there might be wide variances in Part D costs associated with these elective procedures.
   d. The model for predicted and expected standardized payments are based on a hierarchical linear model. I believe that the hospital specific intercepts have Bayesian shrinkage, with more shrinkage for hospitals with lower volume. This has been the general model for previously approved CMS measures, but there has been objection by a minority of cost and resource use committee members to the use of shrinkage estimators.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?
   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**
   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: Reliability of testing at the data element level is known to be very high, given that it is claims data, stored electronically according to standard definitions.
   At the measure level, the data was tested by a test-retest method, comparing two distinct sets of half of the patients at each hospital. Intraclass correlation coefficient was 0.931. In addition, because the split half sets could overlap, an estimate for the whole cohort was derived from the Spearman-Brown prophecy formula. Furthermore, to give another assessment of facility level reliability, the formula presented by Adams, et al was used. The median reliability at the facility level was 0.938, or "almost perfect".
   Panel Member #2: The authors indicate data element testing, but there is no testing done. Instead, they feel that a measure of Medicare costs is reliable 'by definition'. Those who use claims cost data regularly know there are many errors, such as surgeries with no inpatient stay assigned, bills for service that were not clearly provided (e.g., a skilled nursing facility bill with no admission in the prior 7/30/60 days) and so on. As a result, I see data elements as untested.
   For reliability of the measure, they drew two random sample of a given hospital and compared results with an ICC – although this is not 'test-retest' as we have discussed in the Scientific Methods committee, it does provide some information on stability.
   Panel Member #3: Data level testing comparing claims and manually abstracted data was not done.

   a. Score level testing was assessed using
      - Split- sample testing and calculation of ICC

   b. Risk adjustment assessed using
      - Pseudo R-square
      - Overfitting metrics
      - Distribution of Standardized Pearson Residuals
      - Predictive ratios

   Panel Member #4: ICC
7. **Assess the results of reliability testing**
   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: Reliability testing at both the measure score level and the facility level are very strong.

   Panel Member #2: The ICC score is quite high, showing good reliability. Worth noting, this is a cross-sectional view episode cost. Coding changes over time could erode reliability, but probably only on the margin.

   Panel Member #3: Data level testing. Comparison of claims and manually abstracted data has been done for other similar measures and developers report high congruence. Do not consider this a problem

   Score level testing. Split sample ICC very high 0.93.

a. Risk adjustment discussed further below.

Panel Member #4: ICC = 0.931 for score

ICC = 0.938 for facility-level

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒**Yes**

   ☐**No**

   ☐**Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒**Yes**

   ☒**No**

   ☐**Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒**High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒**Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐**Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐**Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Panel Member #1: Very experienced data tester and very strong history of results obtained by multiple methodologies.

    Panel Member #2: Although I think there are some threats to reliability that were not considered, the measure appears to have moderate to high reliability based on the testing presented in the testing attachment.

    Panel Member #3: Measure uses standard methods used and NQF endorsed to construct measure, and standard test of stability of measure, using split sample methods. Hospitals with fewer than 25 cases (substantial number) excluded from reporting, although data from these patients used in construction of risk adjuster for constructing expected cost model.

    Panel Member #4: Developer stated that they tested data elements, but cannot find that material. They did test at the facility level.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**
    Submission document: Testing attachment, section 2b2.

    Panel Member #2: None.

    Panel Member #3: None

    Panel Member #4: I have no concerns as they are explicitly defined and satisfy the criteria of the measure intent of being inpatient and elective for two defined procedures in a defined population. Incomplete administrative data (0.99% of index admissions, those discharged AMA (0.01%), transfers to a federal hospital (very very small), more than two procedures during the index admission (very

very small), missing payment data (1.11%), and no payment due to no index DRG weight (0.65%) are all valid exclusions to the measure

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**
Submission document: Testing attachment, section 2b4.

Panel Member #1: Numerous citations are provided demonstrating the ability to report meaningful differences. in performance.

Panel Member #3: Given high ICC and inter-period ICC, measure appears to have precision to identify meaningful differences

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
Panel Member #1: None. All of the data sources are large, verified, well-known sources of data.

Panel Member #2: NA

Panel Member #3: NA

15. **Please describe any concerns you have regarding missing data.**
Submission document: Testing attachment, section 2b6.

Panel Member #1: The proposal outlines the major causes of missing data, mostly related to the claims process.
These are known, tabulated, and stated in the measure proposal at between 0.2 and 1.3%.
Panel Member #2: No concerns with missing data.
Panel Member #3: Minimal

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☒ No
16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No
16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No
16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☐ No
16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

16e**. Assess the risk-adjustment approach**

Panel Member #1: The large and inclusive databases are historically accurate at accounting for risk-adjustment with the variables included. Of course, not all desired variables are present on claims analyses, but the presence of conditions categories, and their groupings into clinically relevant ones, allowed for a stepwise generalized logistic regression model. Boostrap sampling included all candidate variables and showed a significant relationship. Then, those what were

significant above a 90% cutoff were included in the final analysis. The clinical rationale for the social risk factors is well-delineated and tested. Details are provided about those representing health status at admission, selection of patient into different quality hospitals, care within the hospital, and post discharge setting. The statistical results are provided in tabular for about 75 variables. The rational for social risk factors is clear and tested. The rationale for referral patterns is clear and tested.

Panel Member #2: The social risk factors is a little bit tricky here. The authors present a detailed literature review on potential social risk factors, but only include dual-eligibility status in the model. Seems like an area that would benefit from further research is the measure is used for payment. Underpayment for high risk population groups could lead to less access to care for high risk patients or prevent certain providers from joining alternative payment models.

Panel Member #3: Extensive risk adjustment model with basis for including specific measures described.

Model explains approximately 20% of variance, a reasonable level of adjustment.

Analysis of SES variables well constructed. Particularly like:

--Use of 9 digit zip code mapped to Census Block, rather than zip code. This is a real improvement.

--Analysis of differences in cost between high and low SES patients in hospitals with different SES mix. Shows clear differences in post hospitalization costs for low SES patients regardless of SES mix of hospital.

Other metrics used for assessing risk adjustment clearly reported and acceptable.

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**
    ☒ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**
18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

    Panel Member #2: It is not clear to me why 25 cases is the right minimum N – would be helpful to determine this empirically. The authors give us an observed range of $15,494 to $44,656 for the risk standardized cases, but it would be helpful to see more on the distributional properties of the measure and the relationship between variance and hospital size. I'm concerned that small providers could over-expose themselves to risk with this measure.

    Panel Member #3: Standard concerns that have been previously raised about attribution and standardized costing approach. Nonetheless, find measure reliable.

    I have none

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☒ **Both**
20. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☐ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**
21. **Assess the method(s) for establishing validity**
    Submission document: Testing attachment, section 2b2.2

    Panel Member #1: See above

    Panel Member #2: Data element validation was for diagnosis, not payment. Also, it would be helpful to know a bit more about who was on the TEP. It is very hard to know reasonable costs for an episode b/c so much care is provided in smaller units.

Panel Member #3: Basically face validity of included costs and exclusions, as assessed by Individual health economist

TEP

30 day public comment period

Measure also constructed using standard CMS approach to these measures

Panel Member #4: TEP:Validity of claims based measures established via CMS article

22. **Assess the results(s) for establishing validity**
Submission document: Testing attachment, section 2b2.3

Panel Member #1: A very thorough and extensive empirical model was developed and tested by numerous statistical techniques for each of the factors.

Panel Member #2: Surprising that the TEP was mixed in their overall rating. That said, the measure seems to have good face validity.

Panel Member #3: Validity supported by individual, TEP, public comment, prior work.

Panel Member #4: The validity results are compelling although no actual statistics are provided.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒**Yes**

☐**No**

☐**Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**
NOTE that data element validation from the literature is acceptable.

Submission document*: Testing attachment, section 2b1.*

☒**Yes**

☐**No**

☐**Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☒**High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒**Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☐**Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐**Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Panel Member #1: The testing rationale is quite clear, detailed, and the results provided.

Panel Member #2: The analysis of sub-groups included in the social risk factor section was helpful for understanding the validity of the measure – I would expect certain post-acute care patterns for high risk patients and those patterns are evident in those table. I have some lingering concerns for how well the measure performs at the upper end of the cost distribution mostly b/c that type of information is not covered in testing form.

Panel Member #3: Measure uses standard CMS approach to constructing cost and resource use measures. Biggest issue is whether all costs in 0-30 day post hospitalization period should be included.

ADDITIONAL RECOMMENDATIONS

27. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Panel Member #3: Substantive experts should review definition of included and excluded cases.

Evaluation B: Technical Expert Panel (Preliminary Evaluation Comments)

Measure Number: 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)

Type of Measure: Cost/Resource Use

_____

1. Clinical Logic Evaluation of Measure (questions S8.1.-S.8.6 in submission form)

1a. To what extent is the measure population clinically appropriate?

Panel Member #1: Population appears clinically appropriate. Appropriate exclusion criteria were utilized.

Panel Member #2: Mostly appropriate but 1) why not include patients on medicare <65? (S82)

Most <65 years Medicare patients are NOT severly disabled.

Panel Member #3: The measure is extremely clinically appropriate

Panel Member #4: The measure population appears to be appropriate, with consideration given to the variety of co-morbid conditions (other than what is meant by "major symptoms/abnormalities). As was noted, patients with low SES may have more co-morbid conditions at the time of surgery; what was not stated was that these conditions also tend be poorly controlled in patients with lower SES. It was also noted that African-American patients have worse function after joint arthroplasty than do white patients; however, this cannot be explained only by SES- African American patients tend to be referred late in the course for arthroplasty and thus reach surgery in worse physical condition and with worse function. The issue with late referral is also seen among women, who also tend to have worse function than men after arthroplasty. However, there was no mention of this in the measure, and there should be additional emphasis given to the differential outcomes of joint arthroplasty between women and men. In addition, it is stated that "we do not believe social risk factors should be adjusted for in the THA/TKA payment measure". Given the impact of SES on comorbidities, outcome, issues with family and social support after surgery (e.g., are family members of patients with low SES able to take off of work to care for their family member after surgery or are there increased admissions and associated costs for rehabilitation facility admissions due to this limitation in support? This would be a likely explanation of the increased number of admissions to rehab facilities after surgery and may be more likely than age or, potentially, co-morbidities), etc., it is hard to explain why social risk factors (other than dual eligibility) would not be a part of the calculation in a payment measure.

1b. To what extent are the definitions used to identify the measure population clinically consistent with the intent of the measure?

Panel Member #1: Definitions appear appropriate.

Panel Member #2: Strong

Panel Member #3: I think they are very consistent with the intent of the measure

Panel Member #4: The definitions to identify the measure population are appropriate, except for patients in some of the care settings noted, especially hospice facilities and inpatient psychiatric facilities. Patients admitted to hospice facilities either were initially extremely ill or had an untoward event during/after their arthroplasty; including these patients would skew the results. Patients in

psychiatric facilities likewise may have conditions that impact their surgical outcomes. In addition, including patients treated in homeless shelters, prisons, residential substance abuse centers, psychiatric residential facilities, and non-residential substance abuse faciltiies would result in inclusion of patients with additional (psycho)social and/or physical health conditions that can increase their risk of complications and result in increased costs but not reflecting the care provided by the hospital or physician(s). There is no rationale provided why patients in schools or birthing centers would be included and why they would receive services related to arthroplasty. In addition, the issues with quality are mentioned throughout the measure; however, again, the measure is looking only at complications as a measure of quality. This would seem to be a limited way to define quality after arthroplasty.

1c. To what extent does the submission adequately describe the evidence that supports the decisions/logic for grouping claims (i.e., identifying the measure population, exclusions) to measure the clinical condition for the episode?

Panel Member #1: Measure relies on claims data to provide risk stratification. Although claims data is noted to be verified in other patient populations, It isn't clear how accurately claims data creates risk stratification for arthroplasty patients.

Panel Member #2: Strong-Except why for the first 31 days are all claims deemed part of Episode (ie dialysis patient that has routine postop dialysis)

Panel Member #3: It is important to have a well defined measure population, the submission accomplishes this well.

Panel Member #4: The measure adequately describes the evidence behind the use of the Medicare vs Medicare/Medicaid populations for evaluation. There is also descrption of the differences in outcome based on SES. However, there is less description of the impact of the AHRQ SES Index Score. In addition, use of the measure of complications as a surrogate for outcomes seems limited. Use of functional and/or PROMs would seem to provide better information regarding quality of care provided.

1d. Given the condition being measured, and the intent of the measure, describe the alignment of the length of the episode (including what triggers the start and end) with the clinical course of this condition.

Panel Member #1: 90 days appropriately captures early complications. It does not set up a time frame that could be used for patient reported outcomes. These could be arguably at least as important a measurement of quality that early complication rate.

Panel Member #2: Day of surgery -> post-op 90 day Episode of care; Alignment Strong

Panel Member #3: Most patients undergoing these procedures reach a reasonable level of improvement within 90 days. Most if not all therapeutic interventions are also complete by this point. The 90 day length of episode aligns perfectly with the clinical condition that warrants an elective arthroplasty

Panel Member #4: The length of the measured episode seems appropriate. However, it is noted that readmissions and emergency department claims are not going to be assessed. What if these reflect complications from the arthroplasty (e.g., septic arthritis)? It is noted that additional co-morbidities increase the risk of readmission after joint arthroplasty, but it does not appear that these admissions would be included in this assessment. Costs from treatment during ED visits and readmissions should be included.

---

2. Adjustments for Comparability-Inclusion/Exclusion Criteria (question S.9.1. in submission form)

2a. Describe the clinical relevancy of the exclusions to narrowing the target population for the episode, condition/clinical course or co-occurring conditions, and measure intent.

Panel Member #1: If the intent is to capture complication rates, do the exclusion criteria exclude intra- operative fractures? Should not this diagnosis stay in the group being evaluated? This would be a different cohort than patients having arthroplasty for a fracture.

Panel Member #2: All exclusions reviewed are necessary to truly measure intent except age<65.

Panel Member #3: Each of the exclusions carries additional clinical baggage that will could alter the post operative clinical course. The complexity of the procedure and the recovery may also be increased with any of the excluded conditions. Therefore to get an accurate assessment of the episode, it is essential that control of co-occuring conditions exists.

Panel Member #4: Please see my comments, above. There should be additional exclusions to provide a more homogeneous group of patients (or at least to better weight their co-morbidities and risk) to assess hospital performance.

2b. Do the exclusions represent a large number or proportion of patients?

Panel Member #1: No.

Panel Member #2: Yes

Panel Member #3: NO

Panel Member #4: No, but the list of exclusions should be longer, although this would likely not significantly increase the number/proportion of exluded patients.

2c. Do the exclusions represent a large number or proportion of patients?

Panel Member #1: No.

Panel Member #2: I don't think it will be a large proportion.

Panel Member #3: NO

Panel Member #4: Duplicate question

2d. To what extent are the relevant conditions represented in the codes listed in the submission for clinical inclusions and exclusions?

Panel Member #1: Highly relevant.

Panel Member #2: Very well represented.

Panel Member #3: Very Much

Panel Member #4: It is not noted whether the primary diagnosis resulting in arthroplasty, in RA being listed as a risk factor (i.e., osteoarthritis vs inflammatory arthritis) will be assessed, as the risk factors for poor outcomes or complications and limitations in rehabilitation potential can differ between these two groups.

---

3. Adjustments for Comparability-Risk Adjustment (question S.9.3. in submission form)

3a. To what extent are the covariates (factors) included in the risk-adjustment model clinically relevant and consistent with the measure's intent? Are there other clinical factors or comorbidities that should be considered for inclusion in the model? Excluded from the model?

Panel Member #1: S.9.3 is answered as N/A. Accurate risk stratification is dependent on accuracy of claims data. This is reported to have been validated in other patient populations, it seems a reasonable jump to use this data in arthroplasty patients, but the accuracy of this data could be debated. Moreover, if the goal is to increase quality, risk stratification itself, although an important means to level the playing field among hospitals, disincentivizes current efforts in the arthroplasty community to reduce surgery on high risk patients with modifiable risk factors, for example with high hemoglobin A1cs, severe obesity, malnutrition, poor social support, chronic narcotic use. These patients have been demonstrated to have worse outcomes from arthroplasty. These variables deserve consideration. Perhaps thinking of risk stratification

in terms of modifiable and non modifiable risk factors would be worthwhile, but I suspect claims data that doesn't account for clinical severity might not be a sensitive enough tool to accomplish this, and no better tool is available.

Panel Member #2:

1) Very relevant and consistent
2) Dialysis and other co-morbidities that draw claims in first 31 days if all calims are included
3) Age< 65. If a patient is truly "severely" disabled, exclude from cohort.

Panel Member #3: I think they are relevant

Panel Member #4: See above- would assess OA vs inflammatory arthropathies.

Criterion 3. Feasibility

3. Feasibility

The extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims
- Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
- This measure uses variables from claims data submitted by hospitals for payment, data from Medicare fee schedules, data from Final Rules for Medicare prospective payment systems and payment policies, and CMS-published wage index data

Questions for the Committee:

- Are there any concerns regarding feasibility?

Staff preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**The availability of the SAS code makes the measure more accessible but still limited to the highly skilled. Preparing data sets for use is often more complex than running SAS. There is not much more the developer can do to lower barriers
**No barriers noted. Data are routinely available from claims data
**No major concerns.
**na
**No barriers to implementation that I can see. No concerns.
**No concerns regarding feasibility
**no concerns
**No concerns.
**No concerns about feasibility.
**All data elements are available from administrative data. There are not concerns over data collection strategy. Implementation will be staright forward. There are no associated fees or licensing required.
**no comments

Criterion 4: Usability and Use

Use

4a. Use. evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency.

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4a.2. Feedback on the measure by those being measured or others.

Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

---

4a1. Current uses of the measure

- **Publicly reported?** ☒ **Yes** ☐ **No**
- **Current use in an accountability program?** ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

  OR
- **Planned use in an accountability program?** ☒ **Yes** ☐ **No**

Accountability program details

- Hospital Inpatient Quality Reporting (IQR) program

4a2.Feedback on the measure by those being measured or others

- 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. The developer considers issues raised through the Q&A process about measure specifications or measure calculation in measure reevaluation.

Additional Feedback:

- The MAP reviewed this measure for the Hospital Value-Based Purchasing (VBP) program in the 2015-2016 cycle. MAP did not support this measure because it should first be placed in IQR and hospital compare for a year and then be in the VBP program. MAP has previously advocated keeping a parsimonious set of measures for the VBP program to avoid rewarding or penalizing a provider mulitple times for the same case.

***Questions for the Committee**:*

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Staff preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

Usability

4b. Usability.

The extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement.

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

4b2. Benefits vs. harms.

Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b1. Improvement results

- This is a new measure. The developer did not provide any data to demonstate any improvement.

4b2. Unintended consequences

- The developer did not identify any unintended consequences during measure development and testing.

4b2.Potential harms

- The developer did not identify any potential harms.

Questions for the Committee:

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- What benefits, potential harms or unintended consequences should be considered?
- Do the benefits of the measure outweigh any potential unintended consequences?

Staff preliminary rating for Usability and Use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency

Comments:

**No issues
**yes, it is being reported. My key concern is lack of information on ability of this measure to discriminate performance across providers. Measure developer has not computed this type of reliability metric. We only know something about the stablity of the measure based on split sample test.
**No major concerns.
**reported in Hospital Inpatient Quality program, potential to use it in payment.
**Appears to be used in public reporting.
**As I understand it, this is a new measure that will be used as both a transparency and accountability tool.
**unknown
**No concerns.
**Complementary measure to quality measure for elective hip/knee procedures.
**The measure may be reported by CMS. The results may be available to outside organizations. The implementation will be straight forward and it will be possible to evaluate performance.
**lack of clarity whether currently publicly reported or under consideration for public reporting

4a2. Use – Feedback

**No user feedback was presented beyond users requesting data. Given it has been 6 months, an update on user feedback seems appropriate now.
**yes, hospitals have been given results and asked about measure specifications, SAS code. Measure developer has a process for submitting comments from hospitals and considering the comments in possible revisions to the measure
**It would be preferably if here is a way where responses made to questions, be available to the public so that it can be used by other measured and non measured entities to improve their cost and efficiency. Data should be non indentifiable.
**public results available, no feedback has been incorporated?
**I can't answer this question.
**In process
**unknown
**No concerns.
**Needs discussion
**The performance on that measure can be made available to those being measured. There will be the opportunity for feedback.
**no comment

4b1. Usability – Improvement

Comments:

**If the hospitals are able to understand what costs in the episode are out of line for them, then this can be used for improvement. If the hospitals are simply told that they are overall more expensive, then it is not easy to use this measure for improvement.
**Measure is somewhat useful, for purposes of calling out variations in payments and trying to undestand driver (some of which are likely related to social risk factors and not fully captured by this measure).
**As above.
**theoretically, it could influence hospitals to coordinate more with post acute facilities
**This measure needs to be used together with #1550 and/or #1551, otherwise, I worry about the potential of unintended negative consequences. We cannot say, simply by looking at relative payments, whether above average spending is better or worse care. If post-acute care facilities are spending more to provide additional services that can improve care, that spending is "good" rather than "wasteful". We want to be careful about making judgement without complementary quality information.
**Because this is an elective procedure, patients will have the opportunity to select the highest value provider.
**can identify true cost outliers
**Attribution to acute care hospitals to the exclusion of LTACs, SNFs, and home health agencies limits the use of this measure to impact those loci of care.
**Assume CMS will use/present same grid of cost and quality to providers to see joint position in cost and quality space.
**The measure can be used to benchmark hosptials. The approach will be consistent with other CMS measures of hospital performance.
**unclear how this measure will differ from the existing measures in aiding QI efforts or appropriately furthering the goal of greater efficiency in healthcare

4b2. Usability – Benefits vs. harms

Comments:

**the risk is that hospitals with patients with lower SES are not routinely scoring worse would harm them in pay for performance programs.
**Not sure I see harms other than having hospitals devote resources to potentially reduce costs where the variation across hospitals is somewhat moderate.
**N.A
**none reported

**See comment above about potentially unintended consequences. Would not want providers to eliminate services because they are spending above average when those services might be appropriate for the population they are serving.
**none
**think higher acuity patients cost more- think hospitals with higher proportion of dual eligibles will have higher actual costs related to patient acuity. Think outliers in each category of hospital will become apparent, however, so benefits of the measure outweigh those potential unintended consequences. Think measure results should reflect acuity of patient
**I do not see significant harms.
**Measure encourages reduction in use of post-acute services, which may lead to skimping on valued rehabilitation services and less than full recovery of function. We should discuss.
**All measures of hospital performance can lead to hospital trying to score well on the measure at the expense of high quality care.
**Real concern that measure might muddy the waters on appropriate use of IRFs vs. outpatient rehab vs. home care when this is an area still needing greater attention for appropriate care transitions; putting cost front and center could incentivize lower cost care instead of appropriate levels of care

Criterion 5: Related and Competing Measures

- There are no competing measures for #3474 (i.e. same measure focus and target population)
- The developer identified the following NQF endorsed measures as related measures (different measure focus but same target population as #3474):
- 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)
- NQF identified the following NQF endorsed measures as related measures (same measure focus and approach but different target populations as #3474):
- 2431 Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Acute Myocardial Infarction (AMI)
- 2436 Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Heart Failure
- 2579 Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Pneumonia

Harmonization

- The developer states related measures 1550 and 1551 have been harmonized with the measure.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing

Comments:

**it is not clear how this measure is harmonized with the ETG measure
**No competing measures. Developer states related measures 1550 and 1551 have been harmonized with the measure
**#1550 & #1551 are related measued or THA/TKA in addition to #2431, #2436, #2579 for non THA/TKA.
**complication and readmission measures have been harmonized.
**Yes, there are related measures. No concerns about lack of harmonization.

**I do not think so
**not of which I am aware
**Has been harmonized with quality measures for elective knee and hip arthroscopy.
**There are no competing measures.
**I'd argue the two existing measures are competing measures and lack of clarity how this new measure provides distinguishing information for QI or patient decisionmaking

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January/30/2019
No NQF members who have submitted a support/non-support choice

Brief Measure Information

NQF #: 3474

De.2. 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)

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services (CMS)

De.3. Brief Description of Measure: 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.

IM.1.1. Developer Rationale: 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 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 to hospitals that is not otherwise possible.

De.1. Measure Type: Cost/Resource Use

S.5. Data Source: Claims

Other

S.3. Level of Analysis: Facility

1. 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.***

IM.1. Opportunity for Improvement

IM.1.1. Briefly explain the rationale for this measure (e.g., the benefits or improvements in performance envisioned by use of this measure)

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 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 to hospitals that is not otherwise possible.

IM.1.2. Provide performance scores on the measure as specified (current and over time) **at the specified level of analysis.** (This is required for endorsement maintenance. Include mean, stddev, min, max, interquartile range, 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 subcriterion on improvement (U.3.1.) under Usability and Use.**

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

IM.1.3. If no or limited performance data on the measure as specified is reported in IM.1.2., 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.

N/A

IM.1.4. 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.** (This is required for endorsement maintenance. 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 subcriterion on improvement (U.3.1.) under Usability and Use.**

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

Characteristic	Hospitals with a low proportion (=3.8%) Dual Eligible patients	Hospitals with a high proportion (=11.5%) Dual Eligible patients
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 low proportion of patients below AHRQ SES index score of 42.7	130,235 patients in hospitals with high proportion of patients below 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

IM.1.5. If no or limited data on disparities from the measure as specified is reported in IM.1.4., then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

N/A

IM.2. Measure Intent

IM.2.1. Describe intent of the measure and its components/ Rationale (including any citations) for analyzing variation in resource use in this way.

THA and TKA are common procedures among elderly patients with 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 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.

2. 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. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific *(check all the areas that apply):*

De.7. Care Setting *(Select all the settings for which the measure is specified and tested):*

Inpatient/Hospital

S.1. Measure-specific Web Page *(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.)*

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

S.2. Type of resource use measure *(Select the most relevant)*

Per episode

S.3. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED):*

Facility

S.4. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.5. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.5.1.

Claims

Other

S.5.1. Data Source or Collection Instrument *(Identify the specific data source or data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.)*

Data sources

Chronic Condition Data Warehouse (CCW)

We used the Chronic Condition Data Warehouse (CCW) to develop our measure. The CCW contains existing CMS beneficiary claims data from multiple care settings that can be linked by a unique patient identifier, allowing researchers to analyze individual patient data across the continuum of care. We used a 100% sample of all FFS Medicare beneficiaries from July 2010 - June 2012 who underwent elective hip or knee replacement and met all cohort inclusion criteria.

The measure was developed using claims data from seven standard analytic files contained in the CCW data (inpatient, outpatient, skilled nursing facility, home health agency, hospice, carrier [physician/supplier Part B items], and durable medical equipment).

Medicare Administrative Claims

The data sources for these analyses include Medicare administrative claims and enrollment information for patients with hospitalizations between April 1, 2012 and March 31, 2015 (2016 reporting period). The period for public reporting of the THA/TKA measure aligns with the 90-day THA/TKA complication measure. Medicare administrative claims for the 12 months prior to and during the index admission are used for risk adjustment.

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 2016 was applied to the claims to calculate the measures.

Medicare Enrollment Database (EDB)

This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This dataset was used to obtain information on enrollment, date of birth, post-discharge mortality status, and dual-eligibility. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992).

Medicare Fee Schedules

Fee schedules are lists of pre-determined reimbursement amounts for certain services and supplies (e.g. physician services, independent clinical labs, ambulance services, durable medical equipment) and are used by Medicare in the calculation of payment to providers. We used the applicable fee schedules when calculating payments for claims that occurred in each care setting.

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 through the CMS website.

The American Community Survey (2008-2012)

The American Community Survey data is collected annually and an aggregated 5-years data was used to calculate the AHRQ socioeconomic status (SES) composite index score.

Reference

Fleming, C., Fisher, E., Chang, C., Bubolz, T., & Malenka, D. (1992). Studying Outcomes and Hospital Utilization in the Elderly: The Advantages of a Merged Data Base for Medicare and Veterans Affairs Hospitals. Medical Care, 30(5), 377-391.

S.5.2. Data Source or Collection Instrument Reference *(available at measure-specific Web page URL identified in S.1 OR in the file attached here) (Save file as: S_5_2_DataSourceReference)*

<SamplingMethodologySpecificDataSourceAttachment nodeType="0" />

S.6. Data Dictionary or Code Table *(Please provide a web page URL or attachment if exceeds 2 pages. NQF strongly prefers URLs. Attach documents only if they are not available on a web page.)*

Data Dictionary:

URL:

Please supply the username and password:

Attachment: Del18aNQFHipKneeDataDictionary.xls

Code Table:

URL:

Please supply the username and password:

Attachment:

Construction Logic

S.7.1. Brief Description of 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.

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 constructed a cohort of patients who underwent elective primary THA/TKA based on primary discharge diagnosis in administrative claims data. Specifically, we included Medicare FFS patients age 65 or older with a primary discharge diagnosis of elective primary THA/TKA procedure. We then applied six exclusion criteria as detailed in section S.9.1. Once our cohort was finalized, we examined 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 included payments for all care settings, except Part D. We standardized 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 adjusted 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 as well as social risk assessed by dual eligibility status. We then used 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

S.7.2. Construction Logic *(Detail logic steps used to cluster, group or assign claims beyond those associated with the measure's clinical logic.)*

To construct the measure, we use data from CMS's Chronic Condition Data Warehouse (CCW). The CCW data contain claims for all care settings, supplies, and services as outlined in Section S.7.8. (except 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.

S.7.2a. CONSTRUCTION LOGIC ATTACHMENT or URL: If needed, attach supplemental documentation (Save file as: S_7_2_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.

URL:

Please supply the username and password:

Attachment:

S.7.3. Concurrency of clinical events, measure redundancy or overlap, disease interactions *(Detail the method used for identifying concurrent clinical events, how to manage them, and provide the rationale for this methodology.)*

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 primary 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.

S.7.4. Complementary services *(Detail how complementary services have been linked to the measure and provide rationale for this methodology.)*

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.

S.7.5. Clinical hierarchies *(Detail the hierarchy of codes or condition groups used and provide rationale for this methodology.)*

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. For example, it would allow for the inclusion of both HCC 34 (peptic ulcer, hemorrhage, other specified gastrointestinal disorders) and HCC 33 (inflammatory bowel disease) rather than only HCC 33, which is considered the more “severe” condition.

S.7.6. Missing Data *(Detail steps associated with missing data and provide rationale for this methodology (e.g., any statistical techniques to impute missing data)*

: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.

S.7.7. Resource Use Service Categories (Units) (Select all categories that apply)

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.)

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.)

Durable Medical Equipment (DME)

S.7.8. Identification of Resource Use Service Categories (Units)

(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.)

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 large 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 day 31 through day 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

S.7.8a. 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):

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

Please supply the username and password:

Attachment:

Clinical Logic

S.8.1. Brief Description of Clinical Logic (Briefly describe your clinical logic approach including clinical topic area, whether or not your account for comorbid and interactions, clinical hierarchies, clinical severity levels and concurrency of clinical events.)

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.

S.8.2. Clinical Logic *(Detail any clustering and the assignment of codes, including the grouping methodology, the assignment algorithm, and relevant codes for these methodologies.)*

We focused on a 90-day episode of care triggered by admission for an elective primary THA/TKA as identified using ICD-9 and 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 arthroplasty procedure;
- A concurrent revision procedure;
- A concurrent resurfacing 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;
- Removal of implanted devices/prostheses.
- 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.

S.8.3. Evidence to Support Clinical Logic Described in S.8.2 *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 IM3)*

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

S.8.3a. CLINICAL LOGIC ATTACHMENT or URL: If needed, attach supplemental documentation (Save file as: S_8_3a_Clinical_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.

URL:

Please supply the username and password:

Attachment:

S.8.4. Measure Trigger and End mechanisms *(Detail the measure's trigger and end mechanisms and provide rationale for this methodology)*

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.

S.8.5. Clinical severity levels *(Detail the method used for assigning severity level and provide rationale for this methodology)*

The measure uses administrative claims data to risk adjust for patient comorbidities but does not include adjustments for clinical severity. 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.

S.8.6. Comorbid and interactions *(Detail the treatment of co-morbidities and disease interactions and provide rationale for this methodology.)*

The goal of risk adjustment for this measure is to account for patient age 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 15,000 ICD-9 codes and about 68,000 ICD-10 codes into clinically coherent variables for risk adjustment, the measure employs the publicly available CMS condition categories (CCs) to group ICD-9 and ICD-10 codes into CCs, and selects comorbidities on the basis of both clinical relevance and statistical significance [1].

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.

Adjustments for Comparability

S.9.1. Inclusion and Exclusion Criteria *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)*

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 procedure
- Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication, and readmission rates.
- A concurrent resurfacing procedure
- Rationale: Resurfacing procedures are a different type of procedure involving only the joint's articular surface. Resurfacing procedures are typically performed on younger, healthier patients.
- 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.
- Removal of implanted devices/prostheses
- Rationale: Elective procedures performed in these patients may be more complicated.
- Transfer from another acute care facility for the THA/TKA
- Rationale: The THA/TKA complication measure does not include admissions for patients transferred in to the index hospital, as they likely do not represent elective THA/TKA procedures.

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).

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-9-CM and ICD-10-CM procedure codes are listed in the attached data dictionary

S.9.2. Risk Adjustment Type (Select type)

Statistical risk model

If other:

S.9.3. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets)*

N/A

S.9.4 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.

Standardized pricing

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/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228774267858

S.10. Type of score*(Select the most relevant):*

Continuous variable

If other:

Attachment:

S.11. Interpretation of Score *(Classifies interpretation of a ratio score(s) according to whether higher or lower resource use amounts is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score, etc.)*

Results of the measure alone do not necessarily reflect the quality of care provided by hospitals but simply whether the total episode payments are greater than or less than would be expected for an average hospital with a similar case mix. Hospitals are classified as having a less than average, no different than average, or greater than average payment as compared to national average payment for an episode. Accordingly, a classification of lower than average payment should not be interpreted as better care. The THA/TKA risk-standardized payment (RSP) is most meaningful when presented in the context of a THA/TKA outcome measure, such as the publicly reported THA/TKA complication measure. This is because a measure of payments to hospitals that is aligned with a quality measure facilitates profiling hospital value (payments and quality).

S.12. Detail Score Estimation *(Detail steps to estimate measure score.)*

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.

Reporting Guidelines

This section is optional and will be available for users of the measure as guidance for implementation and reporting.

S.13.1. Describe discriminating results approach

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

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.

S.13.2. 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.

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.

S.13.3. Identify and define peer group

Identify the peer group and detail how peer group is identified and provide rationale for this methodology.

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.

S.13.4. Sample size

Detail the sample size requirements for reporting measure results.

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

S.13.5. Define benchmarking and comparative estimates

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

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.

Measure Testing (subcriteria 2a2, 2b1-2b6)

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☒ other: The American Community Survey

1.2. 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*).

The datasets/data sources we used in testing include: Chronic Condition Data Warehouse (CCW), 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. Census data were used to assess socioeconomic factors and 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], 2009-2013.

The dataset used varies by testing type; see Section 1.7 for details.

Reference

Bonito AJ, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries: final report. Rockville (MD): Agency for Healthcare Research and Quality; 2008 Jan. (AHRQ Publication No. 08-0029-EF). Available from: https://archive.ahrq.gov/research/findings/final-reports/medicareindicators/medicareindicators.pdf

1.3. What are the dates of the data used in testing? July 2010-March 2015

Dates of the data used ranged from July 2010-March 2015. The dates used vary by testing type; see Section 1.7 for details.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were 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*)

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 1.7 for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of admissions/patients varies by testing type; see Section 1.7 for details.

1.7. 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 reported below.

The datasets, dates, number of measured entities and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2)

For reliability testing, we randomly split **Dataset 2** into two samples. The reliability of the model was tested by randomly selecting 50% of the dataset and calculating the risk-standardized payments for this group. We then calculated risk-standardized payments for the remaining 50% of patients and compared the results from each sample to assess reliability of the measure score (**Dataset 2**).

Dataset 2 (2016 public reporting cohort): Medicare Administrative Claims Data, Medicare Enrollment Data (for determination of dual-eligibility for Medicare and Medicaid)

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

For validity testing (Section 2b1)

No empirical testing was done. We used established measure development guidelines, and performed a systematic assessment of measure face validity by a Technical Expert Panel (TEP) of national experts and stakeholder organizations (based on analysis of **Dataset 1**).

Dataset 1 (development dataset): 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

For testing of measure exclusions (Section 2b2)

We examined overall frequencies and proportions of the admissions excluded for each exclusion criteria for all elective primary THA/TKA admissions using **Dataset 2** (2016 public reporting cohort)

For testing of measure risk adjustment (Section 2b3)

During model development, we computed four summary statistics for assessing model performance using the development (A1 - random 50% sample of July 2011-June 2012) and validation (A2 - remaining 50% of July 2011-June 2012; B – full July 2010-June 2011 sample) cohorts (**Dataset 1**):

(1) Quasi-R-square

(2) Over-fitting indices (Calibration γ0, γ1) (3) Distribution of Standardized Pearson Residuals

(4) Predictive ratios

We also assessed model performance using the quasi-R-squared and over-fitting indices for the 2016 public reporting cohort (**Dataset 2**).

Dataset 1 (development dataset)

Dataset 2 (2016 public reporting cohort)

For testing to identify meaningful differences in performance (Section 2b4)

Consistent with other publicly reported payment 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 using **Dataset 2**.

Dataset 2 (2016 public reporting cohort)

For testing of sociodemographic factors in risk models (Section 2b3.4b)

We examined differences in payments according to the proportion of patients in each hospital who were dual eligible for both Medicare and Medicaid. We also used the AHRQ SES index score to study the association between payments and socioeconomic status. These analyses were performed using **Dataset 2** and **Dataset 3**.

Dataset 2 (2016 public reporting cohort)

Dataset 3 (The American Community Survey [ACS]), The American Community Survey, 2009-2013

Data Elements

- Dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from CMS enrollment data (**Dataset 2**)
- Validated AHRQ SES index score is a composite of 7 different variables found in the American Community Survey data (**Dataset 3**). We attributed this score to each index admission (in **Dataset 3**) using patients' 9-digit zip code at the census block group level.

1.8 What were 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.

We selected social risk factors such as SES variables to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various social risk factors to worse health status, higher mortality over a lifetime, hospital outcomes such as readmission or complication of care, discharge destination and cost of care more broadly (Adler and Newman, 2002; Blum et al., 2014; Bozic et al. 2014; Courtney 2017; Freburger et al. 2011; Gilman et al., 2014; Hu et al., 2014; Joynt and Jha, 2013). Income is the most commonly examined variable. Although the literature directly examining how different social risk factors might influence the cost of care of older, insured, Medicare patients following hip/knee surgery is limited, studies have indicated that patients with social risk factors are more likely to be discharged to an institution such as a rehabilitation facility or a skilled nursing facility (SNF) rather than to home (with or without home health services) (see e.g., Courtney et al. 2017; Inneh et al. 2016; Schwarzkopf et al 2016, Vina et al., 2017). And non-home discharge destinations are associated with higher costs. In addition, 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). Specifically, individuals with social risk factors demonstrate increased length of stay (Lan & Kamath, 2017; Lin & Kaplan, 2004) and higher rates of readmission (Vina et al., 2017)

The SES variables used for analysis were:

- Dual eligible status (**Dataset 2**)
- AHRQ-validated SES index score (summarizing the information from the following variables: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12^{th} grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room) (**Dataset 3**).

In selecting variables, our intent was to be responsive to the NQF guidelines for measure developers. Our approach has been to examine all patient-level indicators of SES that are reliably available for all Medicare beneficiaries, are linkable to claims data, and have established validity.

We similarly 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 takes into account both income and assets and is consistently applied across states. For the dual-eligible variable, there is a body of literature demonstrating differential health care and health outcomes among beneficiaries indicating that this variable, while not ideal, also allows us to examine some of the pathways of interest.

Finally, we selected the AHRQ-validated SES index score because it is a well-validated variable that describes the average SES 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. In this submission, we present analyses using the census block level, the most granular level possible using 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. We used 2009-2013 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the AHRQ SES Index at the census block group level. 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 SES 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. In the THA/TKA measure cohort, we were able to assign an AHRQ SES Index score to 99.6% of patient admissions. 89.4% of patient admissions had calculated AHRQ SES Index scores linked to their 9-digit ZIP codes. 10.2% of patient admissions had only valid 5-digit ZIP codes; we utilized the data for the median 9-digit ZIP code within that 5-digit ZIP code.

References

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Affairs (Project Hope). 2002; 21(2):60-76.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation. Cardiovascular quality and outcomes. May 2014; 7(3):391-397.

Bozic KJ, Grosso LM, Lin Z, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. The Journal of Bone & Joint Surgery .2014;96:640-647

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 & Joint Surgery surgery.. 2014;96(21):e180.

Courtney PM, Huddleston JI, Lorio 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

Eapen ZJ, McCoy LA, Fonarow GC, Yancy CW, Miranda ML, Peterson ED, Califf RM, HernandezAF. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circulation Heart Failure. May 2015; 8(3):473-80.

Gilman M, Adams EK, Hockenberry JM, Wilson IB, Milstein AS, Becker ER. California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Affairs (Millwood). Aug 2014; 33(8):1314-22.

Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health Affairs (Project Hope). 2014; 33(5):778-785.

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

Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. Jan 23 2013; 309(4):342-3.

Lan RH, Kamath AF. Race and Rehabilitation Destination After Elective Total Hip Arthroplasty: Analysis of a Large Regional Data Set. Arthroplast Today. Mar 2017 6;3(3):187-191.

Lin JJ, Kaplan RJ. Multivariate analysis of the factors affecting duration of acute inpatient rehabilitation after hip and knee arthroplasty. Am J Phys Med Rehabil. May 2004;83(5):344-52.

Mahomed NN, Barrett JA, Katz JN, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. The Journal of Bone and Joint Surgery American. 2003;85-a:27-32

Ong KL, Kurtz SM, Lau E, Bozic KJ, Berry DJ, Parvizi J. Prosthetic joint infection risk after total hip arthroplasty in the Medicare population. The Journal of Arthroplasty. 2009;24(6 Suppl):105-109.

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. Circulation. Jun 14 2005; 111(23):3063-3070.

van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. European Journal of Epidemiology. Aug 2012; 27(8):605-613.

Vina ER, Kallan MJ, Collier A, Nelson CL, Ibrahim SA. Race and Rehabilitation Destination After Elective Total Hip Arthroplasty: Analysis of a Large Regional Data Set. Geriatr Orthop Surg Rehabil. Dec 2017;8(4):192-201.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level 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*)

Data Element Reliability

Because this measure is calculated from claims submitted by hospitals and other providers, adjudicated by CMS, and stored electronically, the reliability of the data is extremely high. When the measure is computed on the same set of admissions, for the same providers, using the same time period, precisely the same results are obtained. That is, hip/knee payment is a deterministic measure, reproducible by any third party. Therefore, NQF standards are met.

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), and assessed the values according to conventional standards (Landis and Koch, 1977). 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.

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.

Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

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.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Measure Score Reliability

As a metric of agreement, we calculated the ICC (Landis 1977; 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, which, according to the conventional interpretation, is "almost perfect" (Shrout et al. 1979).

Facility-level Reliability

The median reliability score of 0.938 calculated with 3 years of data, is considered "almost perfect" (Landis, Koch, 1977).

Table – Interpretation of metric of agreement

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Source: (Landis, Koch, 1977).

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

References

Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159-174.

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

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)

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

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **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*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. 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)*

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).

Data Element Validity: Validity of Claims-Based Measures

Our team has demonstrated for a number of prior measures the validity of claims-based measures for profiling hospitals by comparing either the measure results or individual data elements against medical records. CMS validated the six NQF-endorsed, claim-based measures currently in public reporting (AMI, heart failure, and pneumonia mortality and readmission) 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 rates estimated using the claims-based risk-adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based models for public reporting. Our group has reported these findings in the peer-reviewed literature (Krumholz et al. 2006; Krumholz et al. 2011; Krumholz et al. 2006a; Keenan et al. 2008; Bratzler 2011; Lindenauer 2011).

References

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, 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 heart failure. Circulation. 2006b Apr 4;113(13):1693-70.

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.

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.

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.

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.

ICD-9 to ICD-10 Conversion

Statement of Intent

☒ Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

☐ Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

☐ The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using General Equivalence Mapping (GEM) software. We then enlisted the help of clinicians with expertise in orthopedic surgery to select and evaluate which of the ICD-10 codes that mapped to the ICD-9 codes were appropriate for use in this measure. Once the ICD-10 system was implemented in October 2015, we performed a series of analyses of the frequency of use of codes, the size of measure cohorts and number of index admissions per hospital in ICD-10-coded claims as compared with historical samples of claims coded with ICD-9.

We have continued to reevaluate ICD-10-based measure specifications. We have studied the FY 2017 version of the ICD-10-CM/PCS system, with particular attention to newly added codes and codes that were removed. We solicited input from clinical and measure experts to determine which, if any, of the newly implemented ICD-10 codes in the 2017 code set should be added to the cohort definitions.

We then solicited input from clinical and measure experts to confirm the clinical appropriateness of the changes to the specifications given the newly implemented ICD-10 codes

To update the measure's risk model, we studied the FY 2017 version of the V22 CMS-Hierarchical Condition Categories (HCC) crosswalk maintained by RTI International, to determine how the newly implemented ICD-10

codes in the 2017 code set were classified, and to examine any code shifts that may have occurred from the previous version of the HCC to the most current version. We then solicited input from clinical and measure experts to confirm the clinical appropriateness of the HCC classifications of the newly implemented ICD-10 codes and any changes warranted due to the code shifts that occurred. The experts also reviewed the newly implemented ICD-10 codes in the FY 2017 version of the ICD-10-CM/PCS to determine which, if any, should be added to the singular ICD-10 code lists that are also used in risk adjustment (conditions that are not captured by CC codes). These processes led to the following changes:

- We expanded the cohort definitions to include ICD-10 codes for use with discharges on or after October 1, 2015 (Previously-specified ICD-9 codes continue to be used for discharges before October 1, 2015).

- We expanded the definitions of the wound/joint infections and mechanical complications used in assessing THA/TKA payments to include ICD-10 codes for use with discharges on or after October 1, 2015 (Previously-specified ICD-9 codes continue to be used for discharges before October 1, 2015).

- We re-specified the risk models:

a) The CC-based risk variables were updated to the ICD-10-compatible Hierarchical Condition Categories (HCC) system version 22, maintained by RTI International; and,

b) The CC-based risk variables were updated to the ICD-10-compatible Hierarchical Condition Categories (HCC) system version 22, maintained by RTI International; and,

c) Certain risk variables (for example, history of PTCA) previously defined using ICD-9 codes were re-defined using ICD-10 codes, for use with inpatient, outpatient, and/or physician Medicare administrative claims on or after October 1, 2015.

The goal of these updates was to maintain the intent of the original measure. Changes are effective in claims for discharges on or after October 1, 2015.

ICD-9 and ICD-10 codes are attached in the Data Dictionary.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

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.

2b1.4. What is 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?*)

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.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b3

2b2.1. Describe the method of testing exclusions and what it tests (*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*)

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.

2b2.2. What were 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*)

In Dataset 2 (prior to exclusions being applied):

Exclusion	N	%	Median % (IQR)	Range	Number of Hospitals
1. Incomplete administrative data	9,412	0.99	1.21 (0.72 - 1.96)	0.12 - 100.00	2284
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	NA	NA	NA
5. THA/TKA stays with missing payment data	10,549	1.11	1.27 (0.72 - 2.26)	0.09 - 50.00	2076
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	1164

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., 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*)

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.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with** 57 **risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

See risk model specification in Section 2b3.3a and the attached data dictionary.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

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.

Social Risk Factors

We selected relevant social risk factors based on a literature review and the availability of the variables in current datasets. In Section 1.7, we describe the variables that we considered and analyzed based on this review. 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 Review of Social Risk Factors and THA/TKA Payment

We performed a literature review to examine the relationship between social risk factors, such as socioeconomic status (SES) or race, and hospital 90-day RSPs following elective primary THA/TKA. The literature review excludes international studies, articles published more than 10 years ago, articles without primary data, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on social risk factors or cost following hip/knee surgery.

Our literature review showed that most studies examine the following social risk factors: 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. In some cases, researchers use patient-level information such as dual eligibility to measure SES. Other research relies on neighborhood/community-level variables such as median household income or composite measures like the AHRQ SES index as proxies for individual patient-level data (see, e.g., Courtney et al. 2017).

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). The higher costs could 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 SES nor race is related to increased costs or that costs for these patients 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 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.

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, SES and race) and costs following THA/TKA surgery, few studies directly address the complex causal pathways. Our literature review 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

Patients with social risk factors such as low SES 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 African-American 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). Furthermore, there is evidence that being non-white (i.e., a Hispanic or an African-American) is associated with increased costs related to hip or knee replacement (Pugely et al. 2014).

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). Our measure risk adjusts for comorbidities to account for health at admission.

Selection of patients into different quality hospitals

Some studies examining the link between social risk factors and costs suggest that the relation is mediated by hospital quality. Disadvantaged patients are more likely to select and be admitted to lower quality hospitals. 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 specialized hospitals might charge a premium.

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 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). Different factors might explain this pattern in post-discharge settings. Low-SES 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 why they need more intensive post-discharge

care. Care providers could 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-SES 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.

Other studies have shown evidence that low-SES 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 blacks are more likely to be discharged home compared to non-blacks (Pugely et al. 2014). In this case, the effect of SES or race may be associated with lower episode of care payment.

References

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

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

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

R. 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

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.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

Variable	PR (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)	
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)
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)

Variable	PR (95% CI)
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)

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Variation in prevalence of social risk factors across measured entities

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-SES patients using the AHRQ SES Index score is 12.7% (IQR: 86.2% – 23.6%).

Empirical association with the outcome (univariate)

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-SES patients using the AHRQ SES Index score was $24,206 (SD: 10,666), compared with $22,700 (SD: 9,646) for non-low SES patients.

Incremental effect of social risk factors in a multivariable model

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).

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. The relative change 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.

Referral patterns by hospital characteristics

To further understand the effect of social risk factors on payments following hip/knee surgery, we examined differences in observed payments between subgroups of patients (measure stratification). Specifically, we stratified payments by 1) dual and non-dual eligible patients, and, 2) low SES and non-low SES patients using the AHRQ SES Index score (see tables below).

We were also interested in understanding whether the association between social risk and payments is mainly driven by a patient- or hospital-level effect. To untangle patient- from hospital-level effects, 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 overall 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 SES 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 SES Index scores in the post-acute care setting. Regardless of hospitals' characteristics, the payment for dual eligible patients and low SES patients was consistently higher than for non-dual eligible patients and non-low SES 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.

Table - Overall Referral Patterns for Hospitals with High ("Top Quartile") and Low Proportion ("Bottom Quartile") of Dual Medicare and Medicaid Beneficiaries, Stratified by Patient-Level Dual Eligibility

			Hospital Proportion of Dual Medicare and Medicaid Beneficiaries			
Care Setting	**All hospitals**		**Bottom Quartile**		**Top Quartile**	
	Dual Beneficiaries	**Non-dual Beneficiaries**	**Dual Beneficiaries**	**Non-dual Beneficiaries**	**Dual Beneficiaries**	**Non-dual Beneficiaries**
Index Hospitalization						
# of patients	58,751	820,151	8,262	315,958	21,683	81,841
$ per patient	$14,884	$14,629	$14,813	$14,610	$14,974	$14,749
Post-Acute Care (total)						
# of patients	58,509	809,696	8,226	311,844	21,594	80,876
% of patients	99.6	98.7	99.6	98.7	99.6	98.8
$ per patient	$12,689	$8,041	$12,631	$7,655	$13,320	$9,613
Skilled Nursing Facilities						
# of patients	31,811	291,579	4,636	108,556	11,818	33,270
% of patients	54.4	36	56.4	34.8	54.7	41.1
$ per patient	$13,100	$10,027	$13,109	$9,794	$13,304	$10,566
Inpatient Rehabilitation						
# of patients	6,696	67,730	853	24,267	2,969	9,685
% of patients	11.4	8.4	10.4	7.8	13.7	12
$ per patient	$13,951	$12,972	$13,304	$12,572	$14,139	$13,336
Non-Acute Inpatient Settings						
# of patients	173	913	22	232	71	153
% of patients	0.3	0.1	0.3	0.1	0.3	0.2
$ per patient	$11,606	$9,735	$7,918	$8,392	$12,501	$11,692

Table - Overall Referral Patterns for Hospitals with High and Low Proportion of Low-SES Patients, Stratified by Patient-Level Low-SES Patients (AHRQ SES Index)

			Hospital % of Low SES			
Care Setting	**All hospitals**		**Q1**		**Q4**	
	Low SES	**Non-low SES**	**Low SES**	**Non-low SES**	**Low SES**	**Non-low SES**
Index Hospitalization						
# of patients	116,749	758,919	9,131	252,356	45,308	84,006
$ per patient	$14,715	$14,635	$14,709	$14,662	$14,807	$14,785
Post-Acute Care (total)						
# of patients	115,431	749,580	9,002	248,935	44,822	83,142
% of patients	98.9	98.8	98.6	98.6	98.9	99
$ per patient	$9,600	$8,166	$9,636	$8,035	$9,864	$8,916
Skilled Nursing Facilities						
# of patients	46,560	275,766	4,094	98,544	17,856	30,880
% of patients	40.3	36.8	45.5	39.6	39.8	37.1
$ per patient	$11,398	$10,148	$11,096	$9,824	$11,421	$10,480
Inpatient Rehabilitation						
# of patients	11,652	62,532	854	19,922	5,062	8,961
% of patients	10.1	8.3	9.5	8	11.3	10.8
$ per patient	$13,602	$12,959	$12,833	$12,300	$13,931	$13,750
Non-Acute Inpatient Settings						
# of patients	226	855	12	229	115	177
% of patients	0.2	0.1	0.1	0.1	0.3	0.2
$ per patient	$10,192	$9,968	$5,630	$9,388	$11,301	$11,652

2b3.5. 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 controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

Risk-Adjustment Model Development and Validation in Medicare FFS (Dataset 1)

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

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.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

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

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

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%

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

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

2b3.9. Results of Risk Stratification Analysis:

N/A

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Quasi-R^2

For a traditional linear model (i.e. ordinary least squares regression) R^2 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 R^2 that is akin to the R^2 of a traditional linear model. In order to provide the NQF Committee with a statistic that is conceptually similar, we produced a "quasi-R^2" 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-R^2 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-R^2 is in-line with R^2s from other patient-level risk-adjustment models for health care payment (Pope et al. 2011).

References

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.

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.

Reference

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.

Overall Interpretation

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

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

_____

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. 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 related to performance gap in 1b)*

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: http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g.,

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)

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.

2b4.3. What is 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? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

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.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

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 eMeasures). 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.***

2b5.1. 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; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to 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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2b6.3. What is 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 nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., 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, provide rationale for the selected approach for missing data*)

N/A

3. Feasibility

F.1. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

F.1.1. Data Elements Generated as Byproduct of Care Processes.

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

F.2. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

F.2.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*)

ALL data elements are in defined fields in electronic claims

F.2.1a. 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 other than electronic sources.

F.2.2. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.

Attachment:

F.3. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

F.3.1. Describe what you have learned/modified 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.

Using administrative claims variables for risk adjustment

This measure uses variables from claims data submitted by hospitals for payment, data from Medicare fee schedules, data from Final Rules for Medicare prospective payment systems and payment policies, and CMS-published wage index data. Prior research has demonstrated that administrative claims data can be used to develop risk-adjusted outcomes measures for both mortality and readmission following hospitalization for acute myocardial infarction1,2 heart failure3,4, and pneumonia5,6, and that the models produce estimates of risk-standardized rates that are very similar to rates estimated by models based on medical record data. This high level of agreement supports the use of the claims-based risk-adjusted models for public reporting. The models have also demonstrated consistent performance across years of claims data.

The approach to gathering risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is coded at every visit, for Medicare FFS patients we use inpatient, outpatient, and physician claims data from the year prior to admission, and diagnosis codes during the index admission, for risk adjustment. The one year time frame provides a more comprehensive view of a patient's medical history than is provided by the secondary diagnosis codes from the index hospitalization alone. If a diagnosis appears in some visits and not others, it is included, minimizing the effect of incomplete coding. We were careful, however, to include information about each patient's status at admission and not to adjust for possible complications of the admission. Although some codes, by definition, represent conditions that are present before admission (e.g. cancer), other codes and conditions cannot be differentiated from complications during the hospitalization (e.g. infection or shock). If these secondary diagnoses are coded only in the index admission, they are not adjusted for in the analysis because they may represent complications of care.

Using Medicare Enrollment data variable for dual eligibility risk adjustment

There is a large body of literature linking socioeconomic status, higher mortality over a lifetime, hospital outcomes such as readmission or complication of care, discharge destination and cost of care more broadly7-12

The hip/knee payment measure includes patient-level dual eligibility status (i.e., eligibility in both Medicare and Medicaid) in the risk adjustment model as a proxy for wealth (income and assets). Information on dual eligibility comes from Medicare Enrollment Data (April 1, 2012 – March 31, 2015).

In selecting variables, our intent was to be responsive to the NQF guidelines for measure developers. Our approach has been to examine all patient-level indicators of SES that are reliably available for all Medicare beneficiaries, are linkable to claims data, and have established validity.

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 takes into account both income and assets and is consistently applied across states. For the dual-eligible variable, there is a body of literature demonstrating differential health care and health outcomes among beneficiaries indicating that this variable, while not ideal, also allows us to examine some of the pathways of interest (ASPE Report 2016)13.

References

1. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation. 2006;113(13):1683-1692.

2. Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circ Cardiovasc Qual Outcomes. 2011;4(2):243-252.

3. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation. 2006;113(13):1693-1701.

4. Keenan PS NS, 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.

5. Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS One. 2011;6(4):e17401.

6. Lindenauer PK, Normand SL, Drye EE, et al. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. J Hosp Med. 2011;6(3):142-150.

7. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood). 2002;21(2):60-76.

8. Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circ Cardiovasc Qual Outcomes. 2014;7(3):391-397.

9. Bozic KJ, Grosso LM, Lin Z, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2014;96(8):640-647.

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

11. Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health Aff (Millwood). 2014;33(5):778-785.

12. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. 2013;309(4):342-343.

13. Office of The Assistant Secretary for Planning and Evaluation. Assistant Secretary for Planning and Evaluation (ASPE). Report to Congress: Social Risk Factors and Performance Under Medicare´s Value-Based Purchasing Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs.

F.3.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, and algorithm)?

There are no fees associated with the use of claims-based measures

F.3.3. If there are any fees associated with the use of this measure as specified, attach the fee schedule here. (Save file as: F3_3_FeeSchedule)

4. 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.

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 performance improvement.

U.1.1. Current and Planned Use

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Hospital Inpatient Quality Reporting (Hospital IQR) program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html

U.1.2. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

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, as well as the number of patients included in the measure, varies 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.

U.1.3. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

U.1.4. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*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.*)

N/A

U.2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation. 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.

The exact number of measured entities (acute care hospitals) varies with each new measurement period. In 2016, 3,481 hospitals were included in 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 resent publically reported measure results included all hospital discharges with a primary elective THA/TKA procedure between April 2014 and March 2017.

Each hospital receives their 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

U.2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

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 is updated each year are released upon request beginning in July of each year.

8. Measure Fact Sheets: provides 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.

U.2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1. Describe how feedback was obtained.

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.

U.2.2.2. Summarize the feedback obtained from those being measured.

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 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.

U.2.2.3. Summarize the feedback obtained from other users.

Summary of Question 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.

U.2.3. Describe how the feedback described in 4a2.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not

N/A

U.3.1. Progress on Improvement. (Not required for initial endorsement unless available.) Performance results on this measure (current and over time) should be provided in IM.1.2 and IM.1.4.

Discuss:

- **Purpose Progress (trends in performance results)**
- **Geographic area and number and percentage of accountable entities and patients included**

N/A

U.3.2. If no improvement was demonstrated, what are the reasons? 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.

N/A

U.4.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

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.

U.4.2. Please explain any unexpected benefits from implementation of this measure.

N/A

5. 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.

H.1. Relation to Other NQF-endorsed Measures

If there are related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

H.1.1. List of related or competing measures (selected from NQF-endorsed measures)

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)

1609 : ETG Based HIP/KNEE REPLACEMENT cost of care measure

H.1.2. If related or competing measures are not NQF endorsed please indicate measure title and steward.

N/A

H.2. Harmonization

H.2.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications completely harmonized?

Yes

H.2.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

H.3. Competing Measure(s)

H.3.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services (CMS)

Co.2 Point of Contact: Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-

Co.3 Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)

Co.4 Point of Contact: Karen, Dorsey, karen.dorsey@yale.edu, 203-688-2475-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? Yearly

Ad.5 When is the next scheduled review/update for this measure? 11, 2019

Ad.6 Copyright statement: N/A

Ad.7 Disclaimers: N/A

Ad.8 Additional Information/Comments: N/A