



## Measure Information

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

### Brief Measure Information

**NQF #:** 3559

**Corresponding Measures:**

**Measure Title:** Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)

**Measure Steward:** Centers for Medicare & Medicaid Services

**sp.02. Brief Description of Measure:** This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.

**1b.01. Developer Rationale:** The goal of this measure is to improve patient outcomes by providing information to patients, physicians, and hospitals about hospital-level, risk-standardized patient-reported outcomes, such as pain and functional status, following elective primary THA/TKA. Measurement of patient-reported outcomes allows for a broad view of quality of care. Complex and critical aspects of care — such as communication among providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment — all contribute to patient outcomes but are difficult to measure by individual process-of-care measures. As patient outcomes are not only influenced by care given during the time of hospitalization but also by patient status on presentation, outcome measures ideally are risk adjusted for patients' comorbid conditions.

THA/TKA procedures provide a particularly rich test bed for developing quality measures based upon patient-reported experiences. These procedures are commonly performed in older patients who have marked pain and functional limitation preoperatively, and who often experience significant improvements postoperatively. Patients who have undergone THA/TKA procedures have already indicated their support of such outcomes in the published literature (Liebs et al., 2013) and voiced their support for this measure as part of a TEP and a Patient Working Group.

**References:**

Liebs TR, Herzberg W, Gluth J, et al. Using the patient's perspective to develop function short forms specific to total hip and knee replacement based on WOMAC function items. Bone Joint J. 2013; 95-B:239–43

**sp.12. Numerator Statement:** The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows:

- For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and
- For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR).

SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group.

References:

Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.

**sp.14. Denominator Statement:** The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.

**sp.16. Denominator Exclusions:** Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.

---

**Measure Type:** Outcome: PRO-PM

**sp.28. Data Source:**

Claims

Instrument-Based Data

**sp.07. Level of Analysis:**

Facility

---

**IF Endorsement Maintenance – Original Endorsement Date:** 2020-11-20 04:03 PM

**Most Recent Endorsement Date:** 11/20/2020 4:03:23 PM

---

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

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

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

## 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

**1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.**

**[Response Begins]**

No

**[Response Ends]**

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

**Current Submission:**

Updated evidence information here.

**Previous (Year) Submission:**

Evidence from the previous submission here.

**1a.01. Provide a logic model.**

*Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.*

**[Response Begins]**

**[Response Ends]**

**1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.**

*Describe how and from whom input was obtained.*

**[Response Begins]**

**[Response Ends]**

**1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.**

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

The goal of this measure is to improve patient outcomes by providing information to patients, physicians, and hospitals about hospital-level, risk-standardized patient-reported outcomes, such as pain and functional status, following elective primary THA/TKA. Measurement of patient-reported outcomes allows for a broad view of quality of care. Complex and critical aspects of care — such as communication among providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment — all contribute to patient outcomes but are difficult to measure by individual process-of-care measures. As patient outcomes are not only influenced by care given during the time of hospitalization but also by patient status on presentation, outcome measures ideally are risk adjusted for patients' comorbid conditions.

THA/TKA procedures provide a particularly rich test bed for developing quality measures based upon patient-reported experiences. These procedures are commonly performed in older patients who have marked pain and functional limitation preoperatively, and who often experience significant improvements postoperatively. Patients who have undergone THA/TKA procedures have already indicated their support of such outcomes in the published literature (Liebs et al., 2013) and voiced their support for this measure as part of a TEP and a Patient Working Group.

**References:**

Liebs TR, Herzberg W, Gluth J, et al. Using the patient's perspective to develop function short forms specific to total hip and knee replacement based on WOMAC function items. *Bone Joint J.* 2013; 95-B:239–43

**[Response Ends]**

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

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

**[Response Begins]**

Table 1. Mean and Distribution of Hospital-level Risk Standardized Improvement Rates (RSIRs) for Risk Model for SCB Improvement following Elective Primary THA/TKA Performed July 1, 2016 to June 30, 2017 (Hospitals with >25 THA/TKA Patients with PRO Data)

Statistic// THA/TKA Procedures

N (Hospitals)// 123

Mean (SD)// 60.16% (219.58)

Percentile

100% Max// 86.84%

99%// 84.73%

95%// 81.92%

90%// 78.85%

75% (Q3)// 72.51%

50% (Median)// 66.49%

25% (Q1)// 54.36%

10%// 20.94%

5%// 13.42%

1%/ 7.70%

0%/ 6.65%

**[Response Ends]**

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

**[Response Begins]**

As stated previously, THA/TKA procedures are commonly performed in older patients who have marked pain and functional limitation preoperatively, and who often experience significant improvements postoperatively. However, not all patients experience benefit from THA/TKA procedures (National Joint Registry, 2012), and many note that their preoperative expectations for functional improvement have not been met (Ghomrawi et al., 2011; Harris et al., 2013; Jourdan et al., 2012; Suda et al., 2010). While the degree and extent of variation in these outcomes across hospitals in the U.S. is unknown, variation in clinical practice has been well documented in the U.S. (American Academy of Orthopedic Surgeons, 2011; Anderson et al., 2012; Roos, 2003). Readmission and complication rates vary across hospitals (Suter et al., 2013a; Suter et al., 2013b), and international experience documents hospital-level variation in patient-reported outcome measures following THA/TKA. The United Kingdom data demonstrated greater than 15% differences among hospitals in the proportion of patients who improved after surgery (National Health System, 2012; Neuburger et al., 2013); and THA/TKA surgical practices vary broadly (American Academy of Orthopaedic Surgeons, 2011; Anderson et al., 2012). Data from this measure support high variability in hospital performance, as noted above.

**References:**

American Academy of Orthopaedic Surgeons (AAOS). Preventing Venous Thromboembolic Disease in Patients Undergoing Elective Hip and Knee Arthroplasty: Evidence-Based Guideline and Evidence Report. 2011.

Anderson FA, Jr., Huang W, Friedman RJ, Kwong LM, Lieberman JR, Pellegrini VD, Jr. Prevention of venous thromboembolism after hip or knee arthroplasty: findings from a 2008 survey of US orthopedic surgeons. The Journal of arthroplasty. May 2012; 27(5):659-666 e655.

Ghomrawi HM, Franco Ferrando N, Mandl LA, Do H, Noor N, Gonzalez Della Valle A. How Often are Patient and Surgeon Recovery Expectations for Total Joint Arthroplasty Aligned? Results of a Pilot Study. HSS journal: the musculoskeletal journal of Hospital for Special Surgery. Oct 2011; 7(3):229-234.

Harris IA, Harris AM, Naylor JM, Adie S, Mittal R, Dao AT. Discordance between patient and surgeon satisfaction after total joint arthroplasty. The Journal of arthroplasty. May 2013; 28(5):722-727.

Jourdan C, Poiraudau S, Descamps S, et al. Comparison of patient and surgeon expectations of total hip arthroplasty. PloS one. 2012; 7(1):e30195.

National Health System: The Information Centre for Health and Social Care. HESonline Hospital Episode Statistics: Proms Data. <http://www.hesonline.nhs.uk/Ease/ContentServer?siteID=1937&categoryID=1295>, 2012.

National Joint Registry. National Joint Registry for England and Wales 9th Annual Report 2012. Available at [www.njrcentre.org.uk](http://www.njrcentre.org.uk): National Joint Registry; 2012.

Neuburger J, Hutchings A, van der Meulen J, Black N. Using patient-reported outcomes (PROs) to compare the providers of surgery: does the choice of measure matter? Medical Care. Jun 2013; 51(6):517-523.

#3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA), Submission Last Updated: Dec 01, 2022

Roos EM. Effectiveness and practice variation of rehabilitation after joint replacement. Current opinion in rheumatology. Mar 2003; 15(2):160-162.

Suda AJ, Seeger JB, Bitsch RG, Krueger M, Clarius M. Are patients' expectations of hip and knee arthroplasty fulfilled? A prospective study of 130 patients. Orthopedics. Feb 2010; 33(2):76-80.

Suter LG, Grady JN, Lin Z, et al. 2013 Measure Updates and Specifications: Elective Primary Total Hip Arthroplasty (THA) And/OR Total Knee Arthroplasty (TKA) All-Cause Unplanned 30-Day Risk-Standardized Readmission Measure (Version 2.0). March 2013a; Available at: <http://qualitynet.org/>.

Suter LG, Parzynski CS, Grady JN, et al. 2013 Measures Update and Specifications: Elective Primary Total Hip Arthroplasty (THA) AND/OR Total Knee Arthroplasty (TKA) Risk-Standardized Complication Measure (Version 2.0). March 2013b; Available at: <http://qualitynet.org/>.

**[Response Ends]**

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

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

**[Response Begins]**

Disparities data are presented below for the Development Dataset (n=6734 patients). These data show bivariate and multivariate results for the following social risk factors: race (non-White), insurance status (Dual eligibility: Medicare and Medicaid coverage), and socioeconomic status (AHRQ SES Index: Bottom quartile). Chi-square analyses and multivariate analyses reveal no statistically significant association between non-White race or AHRQ SES Index bottom quartile and SCB improvement in our Development Dataset; dual eligibility was borderline significant (p=0.058) at the bivariate level (see Table 2 below), and statistically significant when entered into the risk model (see Table 3 below).

Table 2. Bivariate Associations between Social Risk Factors and Observed SCB Improvement (Development Dataset, N=6734)

Variable // Total: Frequency (%) // Achieved SCB Improvement: Frequency (%) // Did Not Achieve SCB Improvement: Frequency (%) // P-value  
Race: Non-White // 548 (8.14%) // 351 (8.06%) // 197 (8.28%) // 0.7569  
Dual eligibility: Medicare and Medicaid // 206 (3.06%) // 146 (3.35%) // 60 (2.52%) // 0.0580  
AHRQ SES Index: Bottom quartile // 688 (10.22%) // 446 (10.24%) // 242 (10.17%) // 0.9922

Table 3. Adjusted Odds Ratios (ORs) for Social Risk Factors Individually Evaluated in the Risk Model for SCB Improvement (Development Dataset, N=6734)

Variable // Frequency (%) // Estimate (SE) // OR (95% CI) // C Statistic for Model Including Social Risk Factor  
Race: Non-White // 548 (8.14%) // -0.08 (0.10) // 0.93 (0.76, 1.13) // 0.68\*  
Dual eligibility: Medicare and Medicaid // 206 (3.06%) // 0.40 (0.17) // 1.49 (1.07, 2.08) // 0.68\*  
AHRQ SES Index: Bottom quartile // 688 (10.22%) // 0.04 (0.09) // 1.04 (0.87, 1.25) // 0.68\*

\* C-statistic for the risk model for SCB improvement in the Development Dataset without any of the three social risk factors = 0.68

Table 4. Mean and Distribution of Risk Standardized Improvement Rates (RSIRs) Calculated without and with Social Risk Factors in the Risk Model (Development Dataset: Hospitals with >25 THA/TKA Patients with PRO Data)  
Statistic // No Social Risk Factors Included // Race: Non-White // Dual Eligibility // AHRQ SES Index: Bottom Quartile

N (Hospitals) // 94 // 94 // 94 // 94

Mean (SD) // 60.39% (19.85) // 60.36% (19.87) // 60.40% (19.85) // 60.30% (19.86)

Percentile

100% Max // 86.25% // 86.03% // 86.21% // 86.23%

99% // 86.25% // 86.03% // 86.21% // 86.23%

95% // 81.94% // 81.71% // 81.96% // 82.03%

90% // 79.95% // 80.10% // 79.95% // 79.95%

75% (Q3) // 72.37% // 72.45% // 72.38% // 72.33%

50% (Median) // 66.57% // 66.60% // 66.53% // 66.57%

25% (Q1) // 53.22% // 53.26% // 53.23% // 53.22%

10% // 20.07% // 20.04% // 20.08% // 20.06%

5% // 14.47% // 14.43% // 14.49% // 14.50%

1% // 8.47% // 8.42% // 8.48% // 8.46%

0% // 8.47% // 8.42% // 8.48% // 8.46%

Pearson Correlation Coefficient (in association with "No Social Risk Factors")

Race: Non-White: 0.9997

Dual Eligibility: 0.9999

AHRQ SES Index: Bottom Quartile: >0.9999

**[Response Ends]**

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

**[Response Begins]**

N/A

**[Response Ends]**

## 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.

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

**[Response Begins]**

No

**[Response Ends]**

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

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

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

**[Response Begins]**

N/A

**[Response Ends]**

**sp.01. Provide the measure title.**

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

**[Response Begins]**

Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)

**[Response Ends]**

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

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

**[Response Begins]**

This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.

**[Response Ends]**

**sp.04. Check all the clinical condition/topic areas that apply to your measure, below.**

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

*Please do not select:*

- *Surgery: General*

**[Response Begins]**

**[Response Ends]**

**sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.**

**[Response Begins]**

**[Response Ends]**

**sp.06. Select one or more target population categories.**

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

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

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

**[Response Ends]**

**sp.07. Select the levels of analysis that apply to your measure.**

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

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

*Please do not select:*

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

**[Response Begins]**

Facility

**[Response Ends]**

**sp.08. Indicate the care settings that apply to your measure.**

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

**[Response Begins]**

Inpatient/Hospital

**[Response Ends]**

**sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.**

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

**[Response Begins]**

N/A

**[Response Ends]**

**sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

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

**[Response Begins]**

No data dictionary/code table – all information provided in the submission form

**[Response Ends]**

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.13. State the numerator.**

*Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).*

*DO NOT include the rationale for the measure.*

**[Response Begins]**

The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows:

- For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and
- For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR).

SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group.

References:

#3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA), Submission Last Updated: Dec 01, 2022

Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.

**[Response Ends]**

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.14. Provide details needed to calculate the numerator.**

*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

This is a patient-reported outcome-based performance measure (PRO-PM).

Two joint-specific Patient Reported Outcome Measure (PROM) surveys are used to collect the data for calculating the numerator: 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients.

These PROM data and specific risk variable data will be collected 90 to 0 days prior to surgery, and PROM data will be collected again 270 to 365 days following surgery.

Data elements used to define the numerator and for risk adjustment that are collected with PROM data include:

- HOOS, JR or KOOS, JR
- Date of Birth
- Single-Item Literacy Screening (SILS2) Questionnaire
- Body Mass Index (BMI) or Weight (kg) and Height (cm)
- Chronic (>90 Day) Narcotic Use
- Total Painful Joint Count (Patient-Reported in Non-Operative Lower Extremity Joint)
- Quantified Spinal Pain (Patient-Reported Back Pain, Oswestry Index Question)
- PROMIS Global Mental Health Score (calculated with data from the Patient-Reported Outcomes Measurement Information Systems (PROMIS) Global or Veteran's Rand 12-Item Health Survey (VR-12); data from VR-12 is translated to PROMIS Global Mental Health scores using a crosswalk created by Cella et. al for PROsetta® Stone)

(Please note: Data elements listed above are detailed in the Data Dictionary accompanying this NQF submission; see Tabs: Risk Variables with PRO Data; HOOS, JR; KOOS, JR; PROMIS Global; VR-12)

Center for Medicare and Medicaid Services (CMS) administrative data is used to identify eligible THA/TKA procedures for the measure cohort (denominator) and additional risk variables, including patient demographics and clinical comorbidities (ICD-10 codes for eligible THA/TKA procedures identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.)

#3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA), Submission Last Updated: Dec 01, 2022

The numerator is the risk-adjusted proportion of patients undergoing an elective primary THA/TKA that meet or exceed a SCB improvement on the HOOS, JR or KOOS, JR from preoperative to postoperative assessment. SCB improvement is defined as:

- For THA patients, an increase of 22 points or more on the HOOS, JR
- For TKA patients, an increase of 20 points or more on the KOOS, JR

SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, TEP, and Technical Advisory Group.

Further, the measure numerator was defined with extensive patient and clinician input. Among the numerator definitions considered by stakeholders during measure development included:

- Change in PROM score from preoperative to postoperative assessment reported as an average for a hospital's patients;
- Postoperative PROM score reported as an average for a hospital's patients;
- A threshold change in PROM score from preoperative to postoperative assessment reported as a proportion of a hospital's patients meeting the threshold;
- A threshold postoperative PROM score reported as a proportion of a hospital's patients meeting the threshold; and
- A combination of both a minimum threshold change in PROM score from preoperative to postoperative assessment and a minimum threshold for postoperative PROM score.

Clinical experts and patients supported a numerator definition that assessed change in PROM score from preoperative to postoperative assessment over a numerator definition that focused on postoperative PROM score. TEP members and patients noted that patients want to see improvement, and that the numerator definition should reflect change following surgery. Comments against using a numerator definition focusing on the postoperative PROM score included concern that it does not reflect degree of improvement, and may incentivize surgery on patients with less severe disease who have better preoperative scores. This concern about assessment of the postoperative PROM score also led to dislike of the last option noted above, a numerator definition combining threshold change and threshold postoperative PROM score.

Stakeholders also strongly supported a numerator definition assessing a threshold change in PROM score over averaging patient change in PROM scores for hospital reporting. They noted that measurement of a threshold change will highlight lower performing patients, will protect at-risk patients, and is easy to understand as a performance measure. Comments against a reported average change included concern that a hospital whose patients all achieve average results could have a reported average change result that would be very similar to a hospital whose patients achieve either very good or very poor results; an average change numerator could show similar results for hospitals with very different patient outcomes).

The numerator definition of SCB threshold change, supported by patients and clinical experts, provides an easy to understand metric that patients found intuitive. Using a SCB threshold avoids the potential for misleading consumers and patients by averaging patient change scores across a hospital when individual patient outcomes within hospitals may vary considerably (as noted above). Using a SCB incentivizes providers to perform surgery on patients with worse baseline scores, a group that might otherwise not be offered surgery, as patients with poorer baseline PRO scores have more room to improve and thus a greater opportunity to achieve SCB. It also encourages providers to not perform THA/TKA procedures on patients with minimal symptoms, who will not benefit at all from surgery. And, since the SCB was defined with close input from patients and clinicians, it does set a minimum improvement threshold, but not one so large as to cause surgeons to avoid performing THA/TKA procedures on patients who would benefit.

References:

Cella D, Schalet BD, Kallen M, Lai JS, Cook KF, Rutsohn J, Choi SW. PROsetta® Stone Analysis Report Volume 2: A Rosetta Stone for Patient Reported Outcomes, PROMIS Global Health – Mental Component and VR-12 – Mental

Component (Algorithmic Scores). <http://www.prosetastone.org/LinkingTables1/GlobalHealth/Pages/default.aspx>, 2018.

Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.

**[Response Ends]**

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.15. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.

**[Response Ends]**

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.16. Provide details needed to calculate the denominator.**

*All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

The cohort for this measure is Medicare FFS patients 65 years of age and older undergoing an elective primary THA/TKA procedure at a non-federal short-term acute care hospital. Inclusion criteria includes patients:

- Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission
- Discharged alive from a non-federal short-term acute care hospital
- Undergoing only elective primary THA/TKA procedures (patients with fractures and revision procedures or with bone metastases are not included)
- Inclusion criteria are harmonized with CMS's existing measure cohort for the hospital-level 90-day risk-standardized THA/TKA complication measure

Center for Medicare and Medicaid Services (CMS) administrative data is used to identify qualifying THA/TKA procedures for the measure cohort. (ICD-10 codes for eligible THA/TKA procedures are identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.)

**[Response Ends]**

**sp.17. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.

**[Response Ends]**

**sp.18. Provide details needed to calculate the denominator exclusions.**

*All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Patients with staged procedures in the measure period are excluded. A staged procedure is identified if a patient has more than one hospitalization with an eligible, elective primary THA or TKA procedure during the measurement period. ICD-10 codes for eligible, elective primary THA/TKA procedures (listed in the Data Dictionary on “ICD-10 2017-2018” tab) are used to identify all eligible procedures during the measurement period; patients with an ICD-10 code for an eligible elective primary THA or TKA procedure in two or more hospital admissions during the measurement period are identified as having a staged procedure, and the patient, including all procedures, is removed from the measure cohort.

**[Response Ends]**

**sp.19. Provide all information required to stratify the measure results, if necessary.**

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

**[Response Begins]**

N/A

**[Response Ends]**

**sp.20. Is this measure adjusted for socioeconomic status (SES)?**

**[Response Begins]**

**[Response Ends]**

**sp.21. Select the risk adjustment type.**

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

**[Response Begins]**

**[Response Ends]**

**sp.22. Select the most relevant type of score.**

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Rate/proportion

**[Response Ends]**

**sp.23. Select the appropriate interpretation of the measure score.**

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

**[Response Begins]**

Better quality = Higher score

**[Response Ends]**

**sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

*Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.*

**[Response Begins]**

Target population: Medicare FFS patients 65 years and older undergoing an elective primary THA or TKA in a non-federal short-term acute care hospital.

To create the denominator:

Step 1. If the patient is a Medicare FFS patient, go to Step 2. If not, do not include in the denominator.  
Step 2. If the patient is identified in CMS administrative claims data as having undergone an eligible elective primary THA or TKA during the measurement period, go to Step 3. If not, do not include in the denominator.  
Step 3. If the patient is 65 years of age or older, go to Step 4. If not, do not include in the denominator.  
Step 4. If the patient was enrolled in Medicare FFS Part A and Part B for the 12 months prior to index admission, and enrolled in Part A during the index admission, then go to Step 5. If not, do not include in the denominator.  
Step 5. If the patient was discharged alive from the hospital, include in the denominator. If not, do not include in the denominator.  
Step 6. If the patient experienced only one elective primary THA/TKA during the measurement period, or if the patient experience more than one elective primary THA/TKA during a singular hospitalization during the measurement period, + in the denominator. If the patient experienced two elective primary THA/TKA procedures during the measurement period performed during distinct hospitalizations, do not include in the denominator.

To create the numerator:

If the patient has complete PRO data collected during the prescribed preoperative and postoperative time windows, and meets or exceeds the SCB improvement threshold on the joint-specific PROM between the preoperative and postoperative assessment:

- for THA patients, an increase of 22 points on the HOOS, JR
- for TKA patients, an increase of 20 points on the KOOS, JR

then include in the numerator. If not, then do not include in the numerator.

The hospital-level measure result is calculated by aggregating all patient-level results across the hospital. For calculation of measure results, we recommend that hospitals should have a minimum case-volume of 25 elective

primary THA/TKA patients with complete patient-reported outcomes and risk variable data collected 90 – 0 days preoperatively and complete patient-reported outcomes data collected 270 – 365 days postoperatively. Hospital-specific risk-standardized improvement rates (RSIRs) are calculated as the ratio of a hospital’s “predicted” improvement to “expected” improvement multiplied by the overall observed improvement rate. Both predicted improvement and expected improvement are derived based on the output of a hierarchical logistic regression model that adjusts for patient case-mix and applies stabilized inverse probability weighting (IPW) to address potential non-response bias.

**[Response Ends]**

**sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

*Examples of samples used for testing:*

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The [2010 Measure Testing Task Force](#) recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

**[Response Begins]**

This PRO-PM is not based on a sample. The measure will allow for proxy responses from a caregiver and hospitals will report whether the PROM survey responder is the patient or a surrogate.

**[Response Ends]**

**sp.30. Select only the data sources for which the measure is specified.**

**[Response Begins]**

Claims

Instrument-Based Data

**[Response Ends]**

**sp.31. Identify the specific data source or data collection instrument.**

*For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.*

**[Response Begins]**

The PROM surveys used to define the measure outcome are 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are presently available in English, not yet

in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) Global or the Veterans RAND 12 Item Health Survey (VR-12) are used. The PROMIS Global is available in sixteen languages; the VR-12 is available in Spanish, Chinese and German.

[Response Ends]

**sp.32. Provide the data collection instrument.**

[Response Begins]

[Response Ends]

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

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

**Current Submission:**

*Updated testing information here.*

**Previous Submission:**

*Testing from the previous submission here.*

[Response Begins]

[Response Ends]

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

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

**Current Submission:**

*Updated testing information here.*

**Previous Submission:**

*Testing from the previous submission here.*

[Response Begins]

[Response Ends]

**2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?**

[Response Begins]

[Response Ends]

**2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation.**

**This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.**

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

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

**[Response Begins]**

**[Response Ends]**

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

Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.

All required sections must be completed.

For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.

If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.

An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.

Contact NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).

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

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

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

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

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

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that

the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

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

an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

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

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

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

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

### Definitions

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

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

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

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

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

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

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

**Current Submission:**

Updated testing information here.

**Previous (Year) Submission:**

Testing from the previous submission here.

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

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

*Please do not select:*

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

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

*Choose one or both levels.*

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

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

[Response Begins]

[Response Ends]

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

*Examples may include correlations or t-test results.*

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

**2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.**

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

[Response Begins]

[Response Ends]

**2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.**

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

[Response Begins]

[Response Ends]

**2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.**

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

[Response Begins]

[Response Ends]

**2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.**

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

[Response Begins]

[Response Ends]

**2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.**

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

[Response Begins]

[Response Ends]

**2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.**

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

[Response Begins]

[Response Ends]

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

**2b.11. Indicate whether there is more than one set of specifications for this measure.**

[Response Begins]

[Response Ends]

**2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.**

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

[Response Begins]

[Response Ends]

**2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.**

*Examples may include correlation, and/or rank order.*

[Response Begins]

[Response Ends]

**2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.**

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

[Response Begins]

[Response Ends]

**2b.15. Indicate whether the measure uses exclusions.**

[Response Begins]

[Response Ends]

**2b.16. Describe the method of testing exclusions and what was tested.**

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

[Response Begins]

[Response Ends]

**2b.17. Provide the statistical results from testing exclusions.**

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

[Response Begins]

[Response Ends]

**2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.**

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

[Response Begins]

[Response Ends]

**2b.19. Check all methods used to address risk factors.**

[Response Begins]

[Response Ends]

**2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.**

[Response Begins]

[Response Ends]

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

[Response Begins]

[Response Ends]

**2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.**

[Response Begins]

[Response Ends]

**2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.**

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

[Response Begins]

[Response Ends]

**2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.**

[Response Begins]

[Response Ends]

**2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.**

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

[Response Begins]

[Response Ends]

**2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.**

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

[Response Begins]

[Response Ends]

**2b.27. Provide risk model discrimination statistics.**

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

[Response Begins]

[Response Ends]

**2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).**

[Response Begins]

[Response Ends]

**2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.**

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

[Response Begins]

[Response Ends]

**2b.30. Provide the results of the risk stratification analysis.**

**[Response Begins]**

**[Response Ends]**

**2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).**

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

**[Response Begins]**

**[Response Ends]**

**2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.**

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

**[Response Begins]**

**[Response Ends]**

### 3. Feasibility

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.

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

No data elements are in defined fields in electronic sources

[Response Ends]

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

[Response Begins]

Currently, this measure allows hospitals to collect data using a range of methods, including paper and electronic formats. While we strongly support the use of electronic data capture, not all clinicians collect patient-reported outcomes on their patients eligible for and undergoing elective primary THA/TKA procedures and many fewer collect these data in electronic form. In fact, the vast majority of hospitals participating in the Center for Medicare and Medicaid Innovation (CMMI) Comprehensive Care for Joint Replacement (CJR) model submitting PRO data do not use electronic data capture. The rapid and continual advances being made in mobile applications and other modes of electronic PRO data capture support likely feasibility of moving to an electronic format for this measure in the near future in ways that were not available at the time of measure development. Further the specifications are harmonized with eCQM process measures that incentivize collection of the PRO data needed to calculate the measure outcome, making future e-specification less burdensome.

[Response Ends]

**3.04. Describe any efforts to develop an eCQM.**

[Response Begins]

[Response Ends]

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

**[Response Begins]**

**[Response Ends]**

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

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

**Attach the fee schedule here, if applicable.**

**[Response Begins]**

N/A

**[Response Ends]**

## 4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

---

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

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

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

Name of program and sponsor

URL

Purpose

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

Level of measurement and setting

[Response Begins]

Payment Program

Quality Improvement (Internal to the specific organization)

Not in use

[Response Ends]

### 4a.02. Check all planned uses.

[Response Begins]

Payment Program

[Response Ends]

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

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

[Response Begins]

This PRO-PM is being submitted for initial endorsement and is not currently used in any accountability program.

[Response Ends]

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

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

**[Response Begins]**

This PRO-PM will be implemented in to-be-determined federal accountability programs through rulemaking in the future.

**[Response Ends]**

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

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

**[Response Begins]**

This PRO-PM has not been implemented yet and thus measure results have not been shared with the measured entities (hospitals). However, feedback was obtained from a TEP (23 total members, five of which were patients), a Technical Advisory Group (eight members), and a Patient Working Group (six total members). These individuals were selected through a publicly posted call for TEP members on the CMS website or through partnerships with the National Partnership for Women and Families and Rainmakers. Feedback was obtained via teleconference calls and online surveys. Patients engaging in this work were provided with preparation calls that reviewed the meeting materials ahead of the meeting date and debrief calls that allowed them to share any thoughts after the scheduled meeting. All meeting materials were sent in advance to allow individuals time to review the performance results and data. A summary of the feedback is provided in Section 1a.3 (value and meaningfulness) of the NQF Evidence Form.

**[Response Ends]**

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

**[Response Begins]**

The Technical Expert Panel (TEP) has been engaged in measure development since the conceptual stage. They have provided input on cohort, outcome, and risk adjustment decisions. The Technical Advisory Group was consulted on determining an outcome and selection of Patient Reported Outcome Measures (PROMs). The Patient Working Group provided input on measure outcome, risk adjustment, and testing results. Statistical analyses were shared with the TEP and Patient Working Group.

**[Response Ends]**

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

**[Response Begins]**

Feedback was obtained via seven teleconference meetings with the TEP, three teleconference meetings with the Patient Working Group, and one online survey administer to the Technical Advisory Group.

**[Response Ends]**

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

**[Response Begins]**

Measure results have not been shared with hospitals, but the TEP, which had multiple clinicians indicated strong support for a patient-reported outcomes performance measure following elective THA and TKA.

**[Response Ends]**

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

**[Response Begins]**

The Patient Working Group members indicated strong support for a patient-reported outcomes performance measure following elective THA and TKA. Patients expected a significant amount of improvement in pain levels and functional status. Patients noted that the procedure impacted their physical health and their quality of life, and find the measure to be valuable.

**[Response Ends]**

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

**[Response Begins]**

TEP, Technical Advisory Group, and Patient Working Group feedback has been considered in the development of this measure through the selection of a cohort, measure outcome, data collection instruments, and risk adjustment models. Patients provided input on the amount of change they would like to see, which helped define the thresholds for the measure outcome.

**[Response Ends]**

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

**[Response Begins]**

This is a new PRO-PM, not currently used in a quality improvement program, and there are no performance results to assess. A primary goal of the PRO-PM following implementation in a federal accountability program is to provide hospitals with performance information necessary to implement focused quality improvement efforts.

**[Response Ends]**

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

**[Response Begins]**

N/A; this is a new PRO-PM not yet implemented. No unexpected findings were noted during PRO-PM development or testing.

**[Response Ends]**

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

**[Response Begins]**

N/A; this is a new PRO-PM not yet implemented. No unexpected benefits were noted during PRO-PM development or testing.

**[Response Ends]**

## 5. Comparison to Related or Competing Measures

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

---

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

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

**NOTE: If there are no related measures, please select N/A.**

*(Can search and select measures.)*

**[Response Begins]**

**[Response Ends]**

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

**NOTE: If there are no competing measures, please select N/A.**

*(Can search and select measures.)*

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

NQF # 2653: Average change in functional status following total knee replacement surgery

**[Response Ends]**

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

**[Response Begins]**

Yes

**[Response Ends]**

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

**[Response Begins]**

To the extent feasible, we have harmonized with existing, related measures. However, we have prioritized the goal of the measure to assess substantial clinical benefit (SCB) improvement in patient-reported outcomes for elective primary THA/TKA patients with minimal patient and provider burden over harmonization if discrepancies occur.

**[Response Ends]**

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

*Provide analyses when possible.*

**[Response Begins]**

NQF # 2653: Average change in functional status following total knee replacement surgery.

This PRO-PM measure differs from NQF #2653 in attribution, cohort, outcome, and risk adjustment.

Attribution: This PRO-PM is a hospital-level quality measure, whereas NQF #2653 is a clinician-level measure.

Cohort: This PRO-PM includes both THA and TKA procedures, as clinical experts agree that hospital-level processes are shared across these procedures, and includes only primary, not revision, procedures, based upon clinical input that revision procedures are more complicated to perform and patient-reported outcomes may be influenced by the initial surgery. The target population is Medicare FFS beneficiaries 65 years of age and older. NQF #2653 includes only TKA procedures, includes knee replacement revisions as well as primary procedures, and includes all adults 18 years of age and older.

Outcome: This PRO-PM collects PROs with the HOOS, JR for THA patients and the KOOS, JR for TKA patients. Timing of PRO data collection is 90 – 0 days prior to and 270 – 365 days following surgery. The numerator measures SCB improvement for each patient from preoperative to postoperative assessment with a binary outcome (Yes/No), and the measure produces a risk-standardized improvement rate that elucidates for hospitals the risk-adjusted proportion of patients with improvement and those without improvement. In contrast, NQF #2653 collects PRO data with the Oxford Knee Score three months prior to and 9 – 15 months following surgery, and measures average change in knee function score. The outcome definition of SCB, with a defined threshold for change in PROM score, allows patients with poorer baseline PRO scores more room to improve and thus a greater opportunity to achieve SCB. This was identified by our TEP members as a specific benefit of measuring SCB versus average change; measuring SCB incentivizes providers to offer and perform THA/TKA procedures on even those with poor PRO scores. Further stated TEP and Patient Working Group concerns with measuring an average change score included the fact that hospitals with all average outcomes would look similar to hospitals whose patients either did very well or very poorly (bimodal distributed outcomes), thus providing potentially misleading information to consumers and patients.

Risk Adjustment: This risk model for this PRO-PM includes important risk variables supported by technical expert panel (TEP) and other expert clinical consultants including health literacy, other musculoskeletal pain and chronic narcotic use which are not included in NQF #2653; these risk variables were identified and tested based upon input from orthopedic professional societies, including AAHKS and AAOS, through public comment (Centers for Medicare & Medicaid Services, CJR Final Rule 2015, Section III.D.3.A).

This PRO-PM is superior to NQF #2653: 1) it more appropriately provides a signal of hospital quality which reflects outcomes for both THA and TKA recipients since within hospitals, care for patients undergoing THA/TKA procedures is provided by the same providers and hospital staff; 2) it assesses SCB improvement

with a binary outcome that elucidates for hospitals and patients the risk-adjusted proportion of patients with and without improvement (a clear, understandable metric that patients support); 3) it uses a more robust and stakeholder-driven risk model, anticipated to produce a measure with greater face validity with stakeholders; and 4) it is harmonized with related measures including NQF #1550 Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA) and Risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) for Merit-based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups (MUC19-28).

References:

Comprehensive Care for Joint Replacement (CJR) Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services Final Rule, 80 C.F.R. 73273 (Nov 24, 2015).

**[Response Ends]**

## Appendix

**Supplemental materials may be provided in an appendix.:**

## Contact Information

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

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

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

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

Sutton, Lamont, doris.peter@yale.edu

## Additional Information

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

**[Response Begins]**

**[Response Ends]**

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

*Describe the members' role in measure development.*

**[Response Begins]**

Yale New Haven Health Services Corporation/Center for Outcomes Research (YNHHSC/CORE) Measure Team Members

1. Lisa Suter, MD –Associate Director, Quality Measurement Program. Provided experience relevant to clinical content and performance measurement.
2. Kathleen Balestracci, PhD – Lead. Provided experience relevant to performance measurement.
3. Zhenqiu Lin, PhD- Analytic Director. Provided experience relevant to performance measurement.
4. Sarah Zimmerman, MS – Lead Analyst. Provided experience relevant to performance measurement.
5. Yongfei Wang, MS – Supporting Analyst. Provided experience relevant to performance measurement.
6. Sheng Zhou, MD, ScM - Supporting Analyst. Provided experience relevant to performance measurement.
7. Kyaw Sint, PhD - Supporting Analyst. Provided experience relevant to performance measurement
8. Elina Kurkurina, MPH – Research Project Coordinator. Provided experience relevant to performance measurement.
9. Lynette Lines, MST, PMP – Research Project Manager. Provided experience relevant to performance measurement.
10. Julia McMahon, BS – Research Assistant II. Provided experience relevant to performance measurement.
11. Susannah Bernheim, MD, MHS- Director of CMS Hospital Measures, Clinical Investigator. Provided experience relevant to clinical content and performance measurement.
12. Harlan Krumholz, MD, SM- Director of CORE. Provided experience relevant to clinical content and performance measurement.

Technical Expert Panel (TEP) Members

1. Peter G. Allen, MS- Regulatory Scientist/Biomedical Engineer, Food and Drug Administration (FDA). Provided experience relevant to performance measurement.
2. David C. Ayers, MD- Professor of Orthopedics, University of Massachusetts (UMass) Medical School. Provided experience relevant to clinical content and performance measurement.
3. Thomas C. Barber, MD- Vice President of Perioperative Services at UCSF and Professor of Orthopedic Surgery. Provided experience relevant to clinical content and performance measurement.
4. Daniel J. Berry, MD- Chairman of Department of Orthopedic Surgery, Mayo Clinic. Provided experience relevant to clinical content and performance measurement.
5. Vinod Dasa, MD- Associate Professor, Department of Orthopaedic Surgery, Louisiana State University Health Sciences Center. Provided experience relevant to clinical content and performance measurement.
6. Cheryl Fahlman, PhD, MBA, BSP- President, CAF Consulting Solutions. Provided experience relevant to performance measurement.
7. Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN- Association of Rehabilitation Nurses; Associate Professor, University of Massachusetts Amherst School of Nursing. Provided experience relevant to clinical content and performance measurement.

8. Courtland G. Lewis, MD- Director of Orthopedic Surgery, Hartford Hospital. Provided experience relevant to clinical content and performance measurement.
9. Patient – Recipient of elective THA or TKA procedure. Provided patient perspective.
10. Michael H. Perskin, MD- The American Geriatrics Society; Associate Chair of Clinical Affairs and Assistant Professor in the Department of Medicine, New York University School of Medicine. Provided experience relevant to clinical content and performance measurement.
11. Jonathan L. Schaffer, MD, MBA- Managing Director, eCleveland Clinic Information Technology Division of The Cleveland Clinic Foundation. Provided experience relevant to clinical content and performance measurement.
12. John H. Seiverd, MD, MBA- Physical Therapy Center Coordinator of Clinical Education, Orthopaedic and Neurologic PT Residency Program Director, James A. Haley Veterans' Hospital. Provided experience relevant to clinical content and performance measurement.
13. Lyle S. Sorensen, MD- Chief of Orthopedics and Sports Medicine, Virginia Mason Medical Center. Provided experience relevant to clinical content and performance measurement.
14. A. Christopher Strenta, PhD- Recipient of elective THA or TKA procedure; Associate Dean, Finance and Operations, Dartmouth College. Provided patient perspective.
15. Margaret A. VanAmringe, MHS- Vice President, Public Policy and Government Relations, The Joint Commission. Provided experience relevant to performance measurement.
16. Rachel DuPre Brodie – Director of Performance Information, Pacific Business Group on Health. Provided experience relevant to performance measurement.
17. Sandra Geisinger, RN, EdD – Recipient of elective THA or TKA procedure. Provided patient perspective.
18. Cherie Gress - Recipient of elective THA or TKA procedure. Provided patient perspective.
19. William Hamilton, MD – Clinical Instructor, Anderson Orthopaedic Clinic; Chair of the Quality Measures Committee, American Association of Hip and Knee Surgeons. Provided experience relevant to clinical content and performance measurement.
20. Arthur Malkani, MD - Chief of Adult Reconstruction and Clinical Professor, Department of Orthopedic Surgery, University of Louisville. Provided experience relevant to clinical content and performance measurement.
21. Nan Rothrock, PhD – Research Associate Professor. Provided experience relevant to performance measurement.
22. Adolph J. Yates, Jr, MD – Orthopaedic Surgeon/Associate Professor, University of Pittsburgh Medical Center. Provided experience relevant to clinical content and performance measurement.
23. Patient – Recipient of elective THA or TKA procedure. Provided patient perspective.

#### Technical Advisory Group

1. Inder Johnson, MBA, OTR/L – Director of Rehabilitation Services and Occupational Therapist, Adventist Health and Rideout Hospital. Provided experience relevant to clinical content.
2. Sheila Barnett, MD – Associate Professor of Anesthesiology, Beth Israel Deaconess Medical Center. Provided experience relevant to clinical content.
3. Brian Curtin, MD, MS – Orthopedic Surgeon, OrthoCarolina Hip and Knee Center. Provided experience relevant to clinical content.
4. Sarah R. Piva, PT, PhD – Associate Professor of Physical Therapy, University of Pittsburgh. Provided experience relevant to clinical content.
5. Jennifer Lennon, OTR/L – Director of Rehabilitation Services, UPMC Presbyterian, Montefiore, and Western Psychiatric Institute Clinic. Provided experience relevant to clinical content.
6. Brandy N. Wilkins, DTP – Program Coordinator, The Joint Commission. Provided experience relevant to performance measurement.
7. Paula Farrell, BSN, RN, CPHQ – Associate Project Director, The Joint Commission. Provided experience relevant to performance measurement.
8. Daniel Riddle, PT, PhD, FAPTA – Otto D. Payton Professor of Physical Therapy, Virginia Commonwealth University. Provided experience relevant to clinical content.

#### Work Group Member

1. Kevin Bozic, MD, MBA- William R. Murray Professor, Chair of the Department of Surgery and Perioperative Care, and Professor of Orthopaedic Surgery at the University of Texas at Austin Dell Medical School. Provided experience relevant to clinical content and performance measurement.

Technical Advisors

1. Kate Chenok, MBA - President of Chenok Associates. Prepared materials for stakeholder engagement.

**[Response Ends]**

**3. Indicate the year the measure was first released.**

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**