# Patient-Reported Outcomes (PROs) Following Elective Primary Total Hip and/or Total Knee Arthroplasty: Hospital-Level Performance Measure

## Version 1.0 Methodology Report

**Submitted by:**
Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNHHSC/CORE)

**Prepared for:**
Centers for Medicare & Medicaid Services (CMS)

**March 2021**

# Table of Contents

# List of Figures and Tables

# Yale New Haven Health Services Corporation — Center for Outcomes Research and Evaluation (CORE) Project Team

Kathleen MB Balestracci, PhD, MSW* – Project Lead

Lisa G. Suter, MD* – Project Director; Contract Director

Zhenqiu Lin, PhD – Analytic Director

Elina Kurkurina, MPH – Project Coordinator

Sarah Zimmerman, MS – Lead Analyst

Kyaw Sint, PhD* – Lead Analyst

Yongfei Wang, MS* – Supporting Analyst

Sheng Zhou, MD, ScM* – Supporting Analyst

Lynette Lines, MS, PMP – Research Project Manager, New Measure Development

*Yale School of Medicine

# Acknowledgements

We would like to acknowledge the support and input from former team members: Susannah Bernheim, Julia McMahon, Demetri Goutos, Nina Brandi, Nicole Voll, Chi Ngo, Elizabeth George, Steven Susaña-Castillo, Mallory Perez, Michael Araas, Weiwei Zhang, Jaymie Simoes, Rana Searfoss, Elizabeth Drye, and Susan Merkle-Ward. We would also like to acknowledge the ongoing support from our Expert Clinical Consultant, Kevin Bozic, MD, MBA. Dr. Bozic is the Chair of the Department of Surgery and Perioperative Care and Professor of Orthopaedic Surgery at the University of Texas at Austin Dell Medical School. He provided considerable guidance on key clinical and methodological decisions throughout the development and testing of this measure.

Acknowledgment of our collaborators and identification of stakeholder engagement, including our Technical Expert Panel (TEP), Technical Advisory Group (TAG), and Patient Working Group, are offered in Appendix A.

**Disclaimer**: The views, thoughts, and opinions expressed in this report belong solely to the author, and not necessarily to any contributors or consultants, including the Expert Clinical Consultants and TEP members and their affiliated organizations. Acknowledgement of input does not imply endorsement of the methodology and policy decisions.

# Executive Summary

Elective total hip arthroplasty and total knee arthroplasty (THA and TKA, respectively) are important, effective procedures performed on a broad population. They offer significant improvement in quality of life by reducing pain and improving function and mobility for the majority of patients undergoing these procedures. They are costly and frequently performed surgeries, most commonly performed for degenerative joint disease and osteoarthritis, conditions affecting millions of Americans. As such, they are priority areas for patient-reported outcome performance measure (PRO-PM) development.

The development of a hospital-level PRO-PM for evaluating THA/TKA reflects the importance of the care and coordination of multiple providers in the clinical outcomes for patients undergoing surgery. Patients will not receive outstanding results if surgeons perform the surgeries well, but the quality of care delivered by others caring for the patients before, during, and after surgery falls short. The goal of a hospital-level outcome measure is to capture the full spectrum of care in order to incentivize collaboration and shared responsibility for improving patients' health and reducing the burden of their disease.

The Centers for Medicare & Medicaid Services (CMS) contracted with Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE) to develop a hospital-level PRO-PM following primary elective THA and TKA. In this report, we present the development, measure specification, and testing results of this PRO-PM.

## Measure Development

This THA/TKA PRO-PM was developed over a multi-year period. We used a multi-faceted approach to develop measure specifications, including environmental scans and literature reviews; consultations with experts experienced with the collection and use of THA/TKA patient-reported outcomes (PROs) data; prospective data collection and analyses to inform measure development and feasibility; and extensive stakeholder engagement in the form of a national Technical Expert Panel (TEP), a Technical Advisory Group (TAG), a Patient Working Group, and multiple public comments.

Due to the absence of large scale and uniformly collected available PRO data from patients undergoing elective primary THA/TKA, in November 2015 CMS established an incentivized, voluntary PRO data collection opportunity within the Comprehensive Care for Joint Replacement (CJR) model to support measure development. Hospitals participating in the CJR model are incentivized to voluntarily collect and submit PRO data from patients prior to and following an elective THA or TKA. Successful submission of PRO data increases a hospital's composite quality score and can positively affect model reconciliation payments. CMS implemented a five-year data collection plan, commencing in 2016. Requirements for successful submission of PRO data for eligible elective primary THA/TKA procedures were identified by CMS in the 2015 CJR Final Rule.[1] This THA/TKA PRO-PM was developed and tested using PRO and risk variable data collected from and submitted by CJR participant hospitals.

Measure specifications have been finalized, risk models developed, and testing completed. Assessments of social risk factors for inclusion in the risk model and analyses for addressing potential bias due to non-response have been conducted. A more detailed description and rationale for measure decisions are provided in the body of the report.

## Measure Specifications

**Data Sources:** This THA/TKA PRO-PM primarily uses PRO and risk variable data collected from and submitted by hospitals, and administrative claims data for Medicare fee-for-service (FFS) beneficiaries. PRO and additional risk variable data are collected from hospitals preoperatively and PRO data are collected again postoperatively on patients undergoing an elective primary THA/TKA. Claims data are used to identify eligible elective primary THA/TKA procedures for the measure cohort and candidate risk variables, including patient demographics and clinical comorbidities up to 12 months prior to surgery. Three additional data sources provide data for the measure as follows: the Medicare Enrollment Database (EDB) identifies Medicare FFS enrollment and race; the Master Beneficiary Summary File (MBSF) allows for the determination of dual eligibility status; and the American Community Survey data allow for derivation of the Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score. Data from these sources are linked for patients undergoing elective primary THA and TKA procedures for the measurement period.

**Measure Cohort:** The measure cohort includes Medicare FFS patients 65 years of age or older undergoing elective primary THA/TKA procedures. Patients with fractures and revisions are excluded from the measure cohort. The measure cohort is intentionally aligned with CMS's existing Hospital-level Risk-standardized Complication Rate (RSCR) Following Elective Primary THA and/or TKA measure (THA/TKA Complication measure) cohort.

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

The measure outcome will assess patient improvement in PROs following elective primary THA/TKA. Patient improvement will be measured using the joint-specific instruments chosen for CJR PRO data collection:

- The Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR)[2] for THA patients, and
- The Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR)[3] for TKA patients.

We recommend PRO data be collected 90 to zero days prior to surgery and 300 to 425 days following surgery.

- The postoperative data collection period finalized in the CJR model was 270 to 365 days after surgery, and these data were used in the development and testing of this measure. However, CORE received extensive input following measure development from clinical experts strongly recommending a revision to the postoperative data collection period to better align with clinical workflow and typical one-year follow-up scheduling and to allow for better postoperative PRO data capture. On this recommendation, we propose measure specifications with a postoperative PRO data collection period of 300 to 425 days after surgery. We anticipate, based upon extensive stakeholder input, this will result in limited impact to the measure's scientific acceptability while significantly increasing clinical acceptance and potentially response rates.

The measure outcome defines patient improvement as a binary outcome (yes/no) of meeting or exceeding an SCB between preoperative and postoperative assessments on the joint-specific PROMs as follows:

- For THA patients: meeting or exceeding the SCB threshold of 22 points on the HOOS, JR.
- For TKA patients: meeting or exceeding the SCB threshold of 20 points on the KOOS, JR.

**Risk Adjustment:** Through a consensus-based approach, we identified a specific set of candidate clinical risk-adjustment factors. The final risk model was developed with input from the TEP and expert orthopedic consultants and empirical analyses. The preoperative score of the Mental Health subscale from the following global PROMs collected with CJR PRO data is included as one of 19 risk variables in the final model:

- The Patient-Reported Outcomes Measurement Information Systems (PROMIS)-Global,[4] or
- The Veterans Rand 12-Item Health Survey (VR-12).[5]

Using the risk model, hospital-specific risk-standardized improvement rates (RSIRs) are calculated, producing a performance measure per hospital which accounts for patient case mix and represents a measure of the quality of care following primary elective THA and TKA.

**Addressing Potential Response Bias**: Analyses examining the impact of missing PRO data were conducted and a statistical approach to potential bias in measure results was implemented. Stabilized inverse probability weights (IPW) were calculated, applied to the risk model, and reflected in RSIRs.

**Testing**: CORE conducted signal-to-noise measure score reliability testing and compared hospital-level RSIRs to hospital-level THA/TKA RSCRs to validate the measure score (see below).

## Measure Results

Among hospitals submitting complete PRO data for at least 25 THA/TKA procedures (N=123), the mean RSIR was 60.2%, and the median RSIR was 66.5%, with the 25$^{th}$ and 75$^{th}$ quartiles at 54.4% and 72.5%, respectively. Signal-to-noise analysis indicated excellent reliability, with a ratio yielding a median reliability score of 0.96 (range 0.90 – 0.99). Measure score validity was supported by a comparison of RSIRs to RSCRs for the NQF-endorsed hospital-level THA/TKA complication measure, which revealed a stepwise trend indicating that hospitals with "Worse than National Rate" RSCRs had a lower median

RSIR and hospitals with “Better than National Rate” RSCRs had a higher median RSIR. Stabilized IPW to address potential non-response bias had a limited impact of RSIR results in the data but was retained due to the importance of consideration of non-response bias for PRO-based measures.

This report details the development, testing, and specifications for a valid, reliable, and meaningfully patient-centered THA/TKA PRO-PM.

# 1. Measure Introduction

## 1.1 Measure Overview

The Centers for Medicare & Medicaid Services (CMS) contracted with Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE) to develop a hospital-level patient-reported outcome performance measure (PRO-PM) following primary elective total hip arthroplasty (THA) and total knee arthroplasty (TKA). The CORE measure development team consisted of a multidisciplinary panel of clinicians, health service researchers, and analysts with expertise in outcome measure development.

The development of a hospital-level PRO-PM for evaluating THA/TKA reflects the importance of the care and coordination of multiple providers in the clinical outcomes for patients undergoing surgery. Patients will not receive outstanding results if surgeons perform the surgery well, but the quality of care delivered by others caring for the patients before, during, and after surgery falls short. The goal of a hospital-level outcome measure is to capture the full spectrum of care in order to incentivize collaboration and shared responsibility for improving patients' health and reducing the burden of their disease.

## 1.2 Key Terminology

There exist many acronyms related to patient-reported outcomes. Throughout this report, we use the terminology advanced by the National Quality Forum (NQF): a "**PRO**" refers to the concept of a patient-reported outcome; a "**PROM**" refers to a survey instrument that captures patient-reported outcomes; and a "**PRO-PM**" is a performance measure that uses PRO data to define the measure outcome.[6]

## 1.3 THA/TKA PRO-PM as a Measure of Quality

Patient-reported outcomes (PROs) assessing health status as a result of care are a critical type of outcome needed for healthcare quality assessment. The use of patient-reported outcome measures (PROMs), standardized instruments that query patients' self assessments of their health, provide a direct way to capture patients' experience of care and the results of that care. PROMs can assess multiple health domains, including physical health, emotional well-being, and social functioning through measuring outcomes relevant to each domain, such as symptoms, functional status, and mental status. As a result, they provide rich information on how care affects multiple dimensions of patients' well-being. Currently, only PROMs assessing patients' experience with the healthcare system are widely used as performance measures; less frequently used are PROMs that ask patients how the care they received affected their health.[7] However, there is a strong interest in PROMs for performance measurement outlined in the National Quality Strategy (NQS) and the quality domains identified by the National Academy of Medicine (previously known as the Institute of Medicine [IOM]).[8,9] Patient-centeredness is one of the ten principles of the NQS and one of the IOM's quality domains.

Many scientifically sound and well-tested PROMs exist. They fall into two broad categories:

- Specific instruments (specific to a condition, procedure, or anatomic location) are developed for use in groups of patients with specific conditions or undergoing specific interventions. These instruments may focus on multiple domains of health or be more narrowly focused on a single domain, such as functional status. In either case, these instruments address outcomes that are more specific to the condition or the procedure, such as considering only lower extremity pain and function following hip or knee surgery.
- Generic or global tools assess general health-related quality of life. These instruments can be used to assess the health status of healthy people or patients with specific or multiple health conditions, but they are more general in nature and often assess overall health-related quality of life. They typically cover multiple outcome domains.

PROMs can provide timely information on patient health status, function, and symptoms over time that can be used to improve patient-centered care and inform clinical decision making.[10] Nevertheless, the use of PROMs in clinical practice is still limited. The use of PROMs for national performance measurement will require broad adoption of clinical guidelines supporting PRO collection and integration into medical decision making.[11]

We decided to measure PROs following THA/TKA because they are important, effective procedures performed on a broad population, and the patient outcomes for these procedures (such as pain, mobility, and quality of life) can be measured in a scientifically sound way,[12-25] are influenced by a range of improvements in care,[26-33] and demonstrate hospital-level variation even after patient case mix adjustment.[34,35] Further, THA/TKA procedures are specifically intended to improve function and reduce pain, making PROs the most meaningful outcome metric to assess. Thus, PRO-PMs for THA/TKA can meet NQF's measure criteria of importance, scientific acceptability, feasibility, and usability. Importance, performance, measurement gap, and feasibility and usability are addressed in the sections that follow; scientific acceptability will be addressed in the testing section of this report.

### 1.3.1 Importance

Elective THA/TKAs are most commonly performed for degenerative joint disease or osteoarthritis, which affects more than 30 million Americans.[36] Osteoarthritis accounts for more than half of all arthritis-related hospitalizations,[37] and in 2013 there were approximately 1,023,000 hospitalizations for osteoarthritis.[38] Knee osteoarthritis is one of the leading causes of disability among non-institutionalized adults,[39] and roughly 80% of patients with osteoarthritis have some limitation in mobility.[40] THA and TKA offer a significant improvement in the quality of life by decreasing pain and improving function in a majority of patients, without conferring a high risk of complications and/or death.[41-44] As the goal of the procedures is to improve quality of life, THA and TKA are ideal candidates for assessing PROs.[45]

Due to their frequency and cost, THA and TKA are also priority areas for outcome measure development. Approximately six million adults aged 65 or older suffer from osteoarthritis.[46] Between July 2014 and June 2017, there were a total of 1,021,359 THA and TKA procedures performed on Medicare fee-for-service (FFS) patients 65 years and older.[47] Estimates place the annual insurer cost of

osteoarthritis in the United States (US) at $149 billion, with Medicare direct payments to hospitals performing THA/TKA exceeding $15 billion annually.[48]

Administrative claims-based elective primary THA/TKA risk-standardized complication and readmission measures have been publicly reported since 2013, assessing outcomes important to patients and clinicians.[47] However, neither of these measures capture the reasons for which patients undergo elective THA and TKA: Will I have less pain and more mobility after surgery? In short, will my quality of life be improved after undergoing the procedure? Therefore, a quality measure based upon PRO data provides both patients and providers with a unique and critical perspective on care.

### 1.3.2 Performance

THA/TKA procedures provide a particularly rich testbed for developing quality measures based upon patient-reported experiences and piloting performance measures based upon PROMs. These procedures are commonly performed in older patients who often experience significant improvements in pain and physical function postoperatively. However, not all patients experience these benefits.[49] Many patients note that their preoperative expectations for functional improvement have not been met.[50-53] PRO data from the first few years for the Comprehensive Care for Joint Replacement (CJR) model reveal hospital-level variation in these outcomes across US hospitals, although the full degree and extent of variation is unknown. In addition, clinical practice variation has been well documented in the US,[54-56] readmission and complication rates vary across hospitals,[57,58] and international experience documents wide hospital-level variation in PROMs following THA/TKA.[59] United Kingdom (UK) data demonstrates a > 15% difference in the proportion of patients improved after surgery across hospitals.[60,61] This evidence supports examining PROs following THA/TKA.

### 1.3.3 Measurement Gap

This THA/TKA PRO-PM fills an important measurement gap in the assessment of the quality of care given to THA/TKA recipients. There are other PRO-PMs addressing total joint replacement, including several by Focus on Therapeutic Outcomes, Inc. (FOTO) and one developed by the Minnesota Community Measurement group that is in use in the state of Minnesota (NQF #2653). However, the FOTO measures use proprietary software and focus on individual joints and not specifically on THA/TKA patients. NQF #2653 was developed for TKA recipients only, limiting the measure's scope, and assesses an average change score for all eligible patients within an orthopedic practice. In determining the measure outcome definition for the THA/TKA PRO-PM, we heard stakeholder concerns about averaging the changes in PROM scores across patients, which can make a hospital with all patients experiencing an average improvement appear the same as a hospital where half of the patients do very well while the other half do very poorly.

In addition, the THA/TKA PRO-PM, through the CMS's prospective data collection, benefited from the ability to consider clinically critical risk variables to develop a model that adequately adjusts for patient case mix, and reflects a more robust and stakeholder-driven risk model. This measure includes key clinical risk variables for a PRO-PM identified by clinical experts and supported by orthopedic professional societies, such as health literacy, back pain, and contralateral leg pain. These ensure

accurate assessment of the index THA/TKA procedure and account for concomitant comorbidities that can interfere with PROM interpretation. Furthermore, this measure accounts for non-response bias, a critical potential threat to the validity of PRO-PMs; failure to account for it may lead to worsening disparities.

### 1.3.4 Feasibility and Usability

THA/TKA procedures provide an opportunity for initiating public reporting of PRO-PMs because there are already multiple initiatives expanding their use within the US. There are several efforts led by orthopedic surgeons and their professional societies to create regional and national patient registries. In addition, the Office of the National Coordinator for Health Information Technology (ONC) and CMS have included a process electronic clinical quality measure (eCQM) indicating the use of a THA/TKA PROM in the Promoting Interoperability programs and have developed eCQMs to promote THA/TKA PROM data collection for the Merit-based Incentive Payment System (MIPS). These initiatives are driven both by an interest in improving clinical care and by the need to evaluate long-term device safety, further prompted by recent publicized orthopedic device failures.[62]

In addition to the fact that orthopedics is advanced in its development and use of validated PROMs for research, an elective procedure such as THA/TKA provides a clear time zero (a reference time) for measurement: the date of the surgery. This allows the use of a standardized measurement timeframe across hospitals.

## 1.4 Measure Use

This measure is intended to measure hospital performance for patients undergoing elective primary THA/TKA procedures. It is important to acknowledge that optimal clinical outcomes depend not just on the surgeon performing the procedure, but on the entirety of the team's efforts in the care of that patient. Care coordination across provider groups and specialties has an important effect on clinical outcomes.[63,64] Even the very best surgeon will not get outstanding results if there are gaps in the quality of care provided by others caring for the patient before, during, and after surgery. The goal of a hospital-level outcome measure is to capture the full spectrum of care to incentivize collaboration and shared responsibility for improving patients' health and reducing the burden of their disease. THA and TKA procedures provide a suitable environment for optimizing care, as there are many studies indicating how providers can improve outcomes of their patients by addressing aspects of pre-, peri-, and postoperative care.[26-31]

## 1.5 Approach to Measure Development

In preparation for measure development, CORE conducted literature reviews examining THA/TKA and PROM use to identify and define the technical decisions to be made in building this measure. The team also performed a systematic literature review of multivariable risk models predicting THA/TKA PROs to inform risk model development. We interviewed experts involved in implementing PROMs for quality

assessment, including international experts experienced in national public reporting and US experts launching PROMs as part of THA/TKA registries. CORE enlisted a national clinical leader in the field of orthopedics, Dr. Kevin Bozic, who served as a consultant to provide clinical input throughout the development and testing of this PRO-PM.

In 2013, concurrent to our work, the ONC contracted with another organization to develop eligible clinician-level PRO-PMs based solely on electronic health record (EHR) data to assess outcomes following THA/TKA. CORE collaborated with this contractor on early measure development decisions to harmonize our work to reduce provider and patient reporting burden and improve the feasibility and usability of these measures. Together, we convened a Technical Expert Panel (TEP) for input on the measure concept, measure cohort (the patients that will be included in the measure), data collection, measure outcome definitions, and risk variables for case mix adjustment. This TEP was reconvened in 2014 for continued input on measure development work.

CORE then paused the development of this measure to support CMS's initiative to collect PRO data from patients following an elective primary THA/TKA. With the recognition that there was no available source for uniformly collected PRO data in numbers large enough to support measure development, in November 2015 CMS established an incentivized, voluntary PRO data collection opportunity within the CJR model. Data analyses on PRO data collected, cleaned, and matched to date (during the first three of five performance years [PYs]) through the CJR model have informed measure development.

CORE returned to measure development work on the PRO-PM following submission of PRO data in the first few CJR PYs. CORE conducted a preliminary analysis of CJR PRO data, reconvened a TEP with returning and new members, and engaged a Patient Working Group and a Technical Advisory Group (TAG) comprised of interested clinicians and other experts for further stakeholder input. Overall, this measure development process informed the PROM data collection, development of the measure specifications (measure cohort, measure outcome including the timing of PROM data collection, calculation of the measure outcome, and risk adjustment), and measure testing.

### 1.5.1 Early and Interim Development Work

Prior to initiation of prospective data collection, CORE worked with multiple partners to gain data and analytic insight into the development of a THA/TKA PRO-PM consistent with stakeholder input. This included data from a medical record review examining risk variable documentation completed by seven orthopedic practices nationwide, and analytic work to support validation of novel short form THA/TKA-specific PROMS to capture variation in hospital-level performance.

#### 1.5.1a Medical Record Review

To evaluate the feasibility, uniformity, and reliability of clinical data elements prioritized by orthopedists for use in risk adjustment following elective primary THA/TKA, CORE, with assistance from national orthopedic professional societies, solicited input from orthopedic practices. Seven practices across the country participated. The medical record review consisted of 30 standardized medical record abstractions at each practice, requiring the abstractor to collect data on the presence or absence of risk

variables in each preoperative record. The abstraction included 27 to 30 individual questions regarding 11 to 12 specific risk variables, depending upon whether the patient was undergoing THA or TKA, respectively. The seven sites provided summary data on 210 total THA/TKA patients (95 THAs, 115 TKAs). Of the seven orthopedic practices surveyed, five were affiliated with hospitals or hospital systems and the mean number of THA/TKA procedures performed annually was 2,329 (range: 158 – 7,578). The results of this review can be found in Appendix B.

#### 1.5.1b Validation Testing of HOOS, JR and KOOS, JR for Hospital Performance

In response to public comment regarding concern about the burden of PRO data collection with the Hip dysfunction and Osteoarthritis Outcome Score (HOOS) and Knee injury and Osteoarthritis Outcome Score (KOOS) surveys, CORE worked with the developer of the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR)[2] and Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR)[3] surveys, Stephen Lyman, PhD, at the Hospital for Special Surgery Research Institute. In collaboration with Dr. Patricia Franklin, Principal Investigator for the Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) research registry, Dr. Lyman further tested these newly validated surveys to evaluate their ability to assess hospital-level performance. These shorter versions greatly reduced data collection burden (from > 40 questions to six – seven), were non-proprietary, and were developed with the close input of THA/TKA patients with the intention of measuring THA/TKA outcomes. This work demonstrated both high correlations between pain and function, supporting the fact that these attributes are strongly correlated among THA/TKA patients, and hospital-level variation in HOOS, JR and KOOS, JR scores. This finding supports their use in a measure of hospital performance.

### 1.5.2 Stakeholder Input

#### 1.5.2a TEP Engagement in 2013 and 2014

CORE convened a national TEP through a public process in 2013 and 2014. The TEP included clinical and methodological experts from several relevant disciplines as well as two patient experts. Four TEP meetings were jointly held with Booz Allen Hamilton, the original contractor developing the harmonized EHR data-based eligible clinician-level measure under contract from ONC. The purpose of this collaboration was to harmonize early measure development decisions for THA/TKA PRO-PMs between CORE's hospital-level and the clinician-level measure(s). The TEP was asked for feedback on a list of valid, reliable, and responsive PROMs for patients undergoing THA/TKA identified through literature reviews and supported by clinical and methodological experts, and the optimal timeframe for collecting PRO data pre and postoperatively. We sought input from the TEP on narrowing a list of candidate risk variables for case mix adjustment based on whether the risk variables were evidence based, valid, reliable, and feasible for patients and/or providers to report. The TEP was also presented with several candidate outcome definitions and asked for feedback on the usability and interpretability (for patients and surgeons) of outcomes measuring postoperative change and measuring postoperative symptom state.

Key input from the TEP included:

- Strong advocacy for minimizing patient and provider burden (shorter PROMs over longer PROMs and non-proprietary over proprietary instruments);
- Recommendations of two procedure-specific and two global PROMs each for data collection:
    - HOOS and KOOS (procedure-specific PROMs), and
    - Patient-Reported Outcomes Measurement Information Systems (PROMIS) -Global and Veterans Rand 12-Item Health Survey (VR-12) (global PROMs);
- Support by clinicians for the use of separate PRO assessments of pain and function to inform clinical decision making;
- Support for a three-month preoperative assessment timeframe and a postoperative data collection timeframe as nine to 12 months post-surgery; and
- Support for an outcome measure capturing improvement in PROMs.

#### 1.5.2b Public Comments in 2014 and 2015

CORE held an interim public comment from March 24, 2014 to April 18, 2014 to gain stakeholder input on preliminary measure specifications. In addition, a public comment period was held from July 9,2015 to September 8, 2015 for input on CMS's proposal of an incentivized, voluntary PRO data collection opportunity as part of the initial CJR model proposed rule. CMS received robust feedback, including a consensus statement from the American Association of Orthopaedic Surgeons (AAOS), the American Association of Hip and Knee Surgeons (AAHKS), the American Joint Replacement Registry, The Hip Society, and The Knee Society recommending a more parsimonious set of PRO data elements and prioritized clinical risk variables for incentivized data collection in CJR. Feedback from both public comments highlighted the burden of collecting the data elements recommended by the TEP. Specifically, the length of the HOOS/KOOS surveys and the number of proposed candidate risk variables were identified as concerning. Several orthopedic leaders asked that we consider recently developed and validated short-form versions of the HOOS/KOOS surveys.

#### 1.5.2c TEP, TAG, and Patient Working Group Engagement in 2018 Through 2020

CORE reconvened a national TEP, and engaged a TAG (composed of those not selected for the TEP) and a Patient Working Group, in 2018. The TEP, comprised of participants continuing their involvement from the earlier TEP for this measure and newly selected TEP members to replace prior members not returning, was convened to help finalize measure specifications and provide input on testing for the hospital-level measure. Due to the considerable number of nominations for the TEP, CORE had the opportunity to create a TAG which, like the TEP, was composed of expert clinicians, methodologists, and researchers. The Patient Working Group was engaged with the assistance of the National Partnership for Women and Families. Members of this group were patient experts with experience undergoing THA(s) or TKA(s).

Key input from stakeholders included:

- Strong TEP support for capturing improvement in the measure outcome using a threshold approach;
- TEP support for using pre-defined substantial clinical benefit (SCB) thresholds of improvement on the HOOS, JR and the KOOS, JR from preoperative to postoperative assessments;
- Mixed support from the TEP and the TAG for consideration of a measure outcome using only joint-specific PROMs versus one that also includes a global PROM;
    - Mixed support for the relative importance of the PROMIS-Global Physical Health subscale or the PROMIS-Global Mental Health subscale (or equivalent VR-12 subscales);
    - Some suggestions of including mental health as a risk factor rather than an element of the outcome; and
    - A few TEP members voiced concern over whether the PROMIS-Global PROM was the appropriate instrument for capturing change in the quality of life following elective primary THA/TKA due to its focus on physical and mental health, rather than a formal health-related quality-of-life assessment.
- TEP and TAG recommendations for specific assessment of outcome disparities and the impact of social risk;
- TEP voiced concerns about combining the THA and TKA cohorts for one measure; and
- TEP support for the risk model and for accounting for social risk factors in the analytic approach to non-response bias.

Key input from the Patient Working Group included:

- Strong support for the inclusion of a global PROM measuring change in quality of life following THA/TKA in the outcome;
- Support for the defined SCB thresholds for the HOOS, JR and KOOS, JR;
- Support for combining the THA and TKA cohorts; and
- Support for the risk model and for accounting for social risk factors in analytic approach to non-response bias.

**1.5.2d Stakeholder Feedback Following Measure Development and Testing in 2020 and 2021**

In Spring 2020, CORE convened an Orthopedic TEP and an Orthopedic Clinical Working Group to provide input across CMS's elective primary THA/TKA measures. The Orthopedic TEP was comprised of clinical experts, measure development experts, and THA/TKA patients. The Orthopedic Clinical Working Group was comprised of four members, each nominated by one of the orthopedic professional societies (AAOS; AAHKS; The Hip Society; and The Knee Society). Dr. Kevin Bozic, our CORE Expert Clinical Consultant, regularly participated in meetings with both the Orthopedic TEP and the Orthopedic Clininical Working Group..

During engagement with these groups, CORE received input regarding the experience of PROM data collection through CJR. Clinical experts expressed concern that the postoperative window established for CJR PRO data collection (270 – 365 days following THA/TKA surgery), set to align with typically

scheduled one-year post-surgery appointments, required the capture of PRO data no later than 365 days or one year. Appointments scheduled any later than one year, or missed appointments rescheduled beyond one year, occurred too late for eligible PRO data collection. Clinical experts strongly recommended extending or shifting the postoperative data collection timeframe to better align with the clinical workflow and follow-up scheduling and to increase postoperative PRO data capture.

# 2. Methods

## 2.1 Overview

The principal data for the development and testing of this measure were PRO data and patient- and provider-reported risk variable data collected and submitted by CJR participant hospitals for elective primary THA and TKA procedures. Patients with complete preoperative and postoperative PRO and risk variable data were included in the dataset used for the development and testing of this measure. We randomly split the dataset into the **Development** (60%) and **Validation** (40%) **Datasets** for risk model development and validation as well as the measure calculation and testing. Next in this report we present detailed data sources, measure specifications, measure development methods, and rationale for measure development decisions.

## 2.2 Data Sources

The principal data for this THA/TKA PRO-PM are PROM and risk variable data collected from hospitals and derived from CMS administrative claims data. PRO and additional risk variable data are collected for patients by CJR participant hospitals prior to an elective primary THA/TKA, and PRO data are collected again following surgery. Eligible THA/TKA procedures for the measure cohort and candidate risk variables, including patient demographics and clinical comorbidities for 12 months prior to surgery, are identified using the CMS administrative claims data to align with the THA/TKA complication measure.

Three additional data sources were used: the Medicare Enrollment Database (EDB) was used to assess Medicare FFS enrollment and race, the Master Beneficiary Summary File (MBSF) was used to determine dual eligibility status, and the Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score was derived from American Community Survey data. Data from these sources were linked for patients undergoing elective THA and TKA procedures during the measurement period. Patients with complete preoperative and postoperative PRO and risk variable data were included in the dataset used for the development and testing of this measure.

### 2.2.1 Collection of PRO Data with CMS's CJR Model

In the absence of a data source for THA/TKA PROs that was uniformly and consistently collected across many hospitals, contained stakeholder-prioritized risk variables, and could be generalizable to US Medicare beneficiaries, CMS incentivized the collection of PRO data by participant hospitals within the CJR model to provide data on a sample of patients undergoing elective primary THA/TKA for PRO collection. Per the 2015 Final Rule, CMS determined that this approach would provide hospital-level PRO data representative of THA/TKA procedures conducted in hospitals, from hospitals representative of the case mix experienced at various hospitals nationwide. Data would be consistently collected at the hospital level, containing risk variables identified by orthopedists as important for risk-adjustment consideration, and allow for the assessment of a set of "parsimonious" data elements so as to minimize

the burden for patients, surgeons, and hospitals. These data would be used for the development of a THA/TKA PRO-PM.

Timing for preoperative PROM and risk variable data collection was set for 90 to zero days prior to the elective primary THA/TKA procedure, and the timing of postoperative PROM data was set for 270 to 365 days following the procedure. Patient identifiers (to allow matching of pre- and postoperative PRO data and to match PRO data to claims data) were likewise collected preoperatively and postoperatively.

CMS incentivized submission of these PRO data by awarding CJR participant hospitals extra points towards their composite quality score, which could positively affect model reconciliation payments. CMS implemented a five-year data collection plan, commencing in 2016. Requirements for successful submission of PRO data for eligible elective primary THA/TKA procedures as identified by CMS in the 2015 Final Rule are provided in Table 1.

**Table 1. Minimum Case Requirements for Eligible Procedures in Each PY for Successful PRO and Risk Variable Data Collection in CJR**

| Performance Year | Eligible THA/TKA Procedure Timeframe | Submission Requirements |
|---|---|---|
| 1 | July 1, 2016 – August 31, 2016 | ≥ 50% or ≥ 50 eligible procedures |
| 2 | September 1, 2016 – June 30, 2017 | ≥ 60% or ≥ 75 eligible procedures |
| 3 | July 1, 2017 – June 30, 2018 | ≥ 70% or ≥ 100 eligible procedures |
| 4 | July 1, 2018 – June 30, 2019 | ≥ 80% or ≥ 200 eligible procedures |
| 5 | July 1, 2019 – June 30, 2020 | ≥ 80% or ≥ 200 eligible procedures |

Complete data specifications for the development data collection as part of CJR are presented in Appendix C. Data are entered by CJR participant hospitals into a standardized macro-enabled template for uniform data submission. The template and accompanying data entry instruction materials guide hospital users in the collection and submission of required data elements. The use of short-form PROMs (HOOS, JR or KOOS, JR, plus VR-12 or PROMIS-Global) minimize provider and patient burden.

## 2.3 Measure Cohort

The cohort for this measure is harmonized with CMS's Hospital-level Risk-standardized Complication Rate (RSCR) Following Elective Primary THA and/or TKA Measure. Detailed measure specifications for the existing administrative claims-based complication measures are publicly available in the Hospital-level Risk-standardized Complication Rate Following Elective Primary THA and/or TKA Measure Methodology Report[65,66] on the CMS Measure Methodology webpage. The inclusion and exclusion criteria are summarized below.

### 2.3.1 Inclusion Criteria

- 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.
- Aged 65 or older.

- Discharged alive from non-federal short-term acute care hospital.
- Elective primary THA/TKA procedures only (patients with fractures and revisions, malignant neoplasms, or mechanical complications are not included)

These inclusion criteria are harmonized with CMS's existing measure cohort for the 90-day hospital-level RSCR measure.[47]

**Rationale for Elective Primary THA/TKA Procedures**

Patients who undergo an elective primary THA or TKA procedure tend to differ in health status from patients requiring non-elective lower joint procedures. Patients who present urgently for non-elective THA or TKA procedures, such as those with hip fractures, tend to be frailer and carry a greater comorbidity burden than their peers undergoing elective primary THA/TKA procedures. As such, they represent a distinct clinical population not appropriate for inclusion in a performance measure of elective primary THA/TKA. Furthermore, these patients cannot easily complete preoperative PROMs, a critical piece of information to gauge response to surgery. Similarly, patients undergoing non-primary procedures, such as revision THA/TKA, have experienced complications from their primary procedure and are receiving a salvage procedure.

Also, a PRO-PM will likely be most useful for patients who are undergoing an elective primary THA/TKA and who consequently have the greatest freedom to choose the hospital at which to have the surgery. As elective primary THA/TKA are preference-sensitive procedures, shared decision making is an important part of the care process. Patients and providers can decide when to conduct the surgery and what steps patients need to engage in before and after surgery to optimize outcomes.

Therefore, like the measure cohorts for CMS's THA/TKA complication and readmission measures, we recommend excluding all patients presenting with a relevant anatomic lower extremity or a pelvic fracture or bony metastasis, those undergoing revision procedures, those requiring removal of hardware, and those requiring THA/TKA surgery due to a prior mechanical complication. This cohort will best reflect the care provided by the hospital performing the elective primary THA/TKA procedure as well as ensure appropriate risk adjustment across a more homogeneous group of patients.

### 2.3.2 Exclusion Criteria

- 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.
- Patients who were discharged against medical advice.

**Rationale for Exclusion of Staged Procedures**

The overlapping recovery periods for staged procedures occurring within one year of each other has two consequences that set patients experiencing staged THA/TKA procedures apart from patients experiencing unilateral or bilateral procedures: 1) the recovery from one procedure may negatively impact recovery from the other procedure; and 2) it may be challenging to fully distinguish the recovery

for either of the procedures from the other with postoperative PRO data. For these reasons, patients with staged procedures during the measurement period are excluded from the cohort.

### 2.3.3 Measuring THA and TKA PROs Together

An additional consideration for the THA/TKA PRO measure cohort was whether to combine THA and TKA procedures into one combined measure or to report them separately. During measure development of the claims-based THA/TKA measures, CORE found that hospital readmission and complication rates were similar for the two procedures, and, at many hospitals, the staff involved in the care of these two patient groups were the same. For these reasons, and to ensure adequate case volume to allow calculation and reporting of reliable hospital-level performance, CMS combines THA and TKA procedures in the readmission and complication measures.

During measure development, our clinical experts advised us that the recovery course differs for patients undergoing THA compared with TKA. However, although rehabilitation following TKA is more complicated and lengthier than recovery following THA,[67-70] these clinical experts agreed that time to the postoperative data collection window allowed both THA and TKA patients to realize a complete recovery.

We examined separate and combined risk models for THA and TKA procedures using patient demographic and clinical risk variables. We found that model performance and risk prediction were similar or better in the combined model as compared to the THA- or TKA-only models.

## 2.4 Measure Outcome

The measure defines patient-level improvement as a binary outcome (yes/no) of whether or not the change in PROM score between preoperative and postoperative assessment meets or exceeds the SCB on the joint-specific PROMs, defined as follows:

- An increase of 22 points or more on the HOOS, JR for THA patients, or
- An increase of 20 points or more on the KOOS, JR for TKA patients.

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

The PRO preoperative and postoperative data collection periods are as follows:

- Preoperative PRO data collection: PROM and risk variable data will be collected between 90 and zero days before surgery.
- Postoperative PRO data collection: PROM data will be collected between 300 to 425 days following surgery.

While the postoperative data collection period finalized in the CJR model was 270 to 365 days after surgery, CORE received extensive input from clinicians on the Orthopedic TEP and the Orthopedic Clinical Working Group recommending a revision to the postoperative data collection period to better

align with clinical workflow and follow-up scheduling and to allow for better postoperative PRO data capture. Suggestions included the extension of the postoperative data collection window by one or two months, and/or shifting the postoperative data collection window to provide two months prior to and following the typical one-year post-surgery follow up to allow for improved capture of PRO data. CORE is recommending the latter suggestion of postoperative PRO data collection following elective primary THA/TKA from 300 to 425 days (10 – 14 months) following surgery. This minor alteration better aligns with clinical practice patterns and allows for scheduling flexibility around the one-year surgical anniversary. Therefore, while the results presented in this report represent data collected between 270 to 365 days following surgery, we propose measure specifications with a postoperative PRO data collection period of 300 to 425 days after surgery.

**Rationale for Selection of PROM Instruments**

In order to select PROMs suitable for use in hospital-level performance measures, we performed an environmental scan and literature review to identify existing PROMs[23] and assess their performance characteristics in THA/TKA patients based upon published criteria.[10,71,72] We also consulted a range of clinical and orthopedic quality measurement experts regarding their experience collecting PROM data from THA/TKA patients. A list of candidate PROMs identified as valid, reliable, and responsive assessments of patient-reported outcomes in patients undergoing THA/TKA was assessed, including:

- PROMIS-Global;
- PROMIS-29;
- EuroQOL-5D (EQ-5D);
- Medical Outcomes Study (MOS) Short Form (SF)-36 and Veterans Rand (VR)-36;
- MOS Short Form (SF)-12 and VR-12;
- MOS Short Form (SF)-8;
- Continuity Assessment Record and Evaluation (CARE) Item Set C – Section II: D Pain and E Mobility;
- Oxford Hip and Knee Scores (OHS and OKS);
- HOOS and KOOS; and
- The Western Ontario and McMaster Universities Arthritis Index (WOMAC).

Clinical experts, the TEP, and representatives from industry, academia, clinical orthopedics, and rehabilitation at a 2012 Food and Drug Administration (FDA) public meeting recommended that both a global PROM and procedure-specific PROMs be used to capture the full spectrum of relevant risk variables (such as mental well-being) and outcomes (such as functional status).[73] Joint-specific PROMs, the HOOS and the KOOS, were selected. The HOOS and the KOOS include 40 and 42 questions, respectively, regarding hip- or knee-related symptoms and pain, physical function, and quality of life that are specific to the experience of patients with hip or knee pain, respectively.[74] For global health assessment, the PROMIS-Global and the VR-12 were initially considered. The PROMIS-Global is a ten-question survey that addresses physical, mental, social, and global health domains.[4] The VR-12 is a 12 question survey that assesses similar domains and summarizes the score using Physical Health Summary

Measure and Mental Health Summary Measure scores.[5] These surveys are considered "global" PROMs in that they assess general aspects of health and well-being and are not specific to THA/TKA patients.

In addition, the TEP highlighted that the ideal combination of instruments should consider the needs of individuals with lower levels of education, English language skills, literacy, and numeracy.[75] The PROMs selected by the TEP represent validated, non-proprietary PROMs that have either been tested in patients undergoing THA/TKA or, in the case of the PROMIS-Global, undergone rigorous testing during development with plans to test in patients undergoing THA/TKA.

Following initial PROM selection, CMS received feedback from stakeholders that the 40 or more questions required to complete the HOOS or KOOS instruments were too burdensome for national adoption. Both an AAHKS-convened PRO Summit for Total Joint Arthroplasty and the public comment for CMMI's CJR model revealed broad stakeholder support for shifting to less burdensome instruments. The HOOS, JR and KOOS, JR instruments were proposed by a consensus group of stakeholders through public comment as alternatives. Based on this broad public input and the CJR Federal Rule, finalized either the HOOS pain and function, daily living subscales or the KOOS stiffness, pain, and function, daily living subscales or the HOOS, JR and KOOS, JR forms as the joint-specific PROM, along with either the PROMIS-Global or VR-12 PROMs, to be submitted in order to successfully meet the criteria for voluntary PRO and risk variable data submission for the CJR model.

More recent stakeholder input and data analysis have led the measure developers to recommend a measure outcome using the joint-specific PROMs only at this time. While the Patient Working Group supported inclusion of PROMs in the measure outcome to assess improvement in overall quality of life, the TEP and TAG could not reach a consensus on the inclusion of a global PROM in addition to a joint-specific PROM in the measure outcome, and several stakeholders across groups questioned the selection of the PROMIS-Global as the best measure for assessing the quality of life following elective primary THA/TKA. In addition, preliminary data analysis suggested that the PROMIS-Global Physical Health and Mental Health subscales were not as responsive and demonstrated smaller improvements than the joint-specific PROMs. Thus, the global health instruments were not used for outcome measurement. We do, however, include the preoperative Mental Health subscale score derived from these global health instruments as a risk variable (see Section 2.6).

**Rationale for Measure Outcome Definition and SCB Thresholds**

In selecting a measure outcome definition, CORE considered stakeholder input, methodological concerns, anticipated approaches to publicly reporting hospital performance, and potential unintended consequences. The result must provide stakeholders with a usable, understandable metric for evaluating hospital quality that can also provide the level of detail needed by physicians and hospitals to optimize patient-centered decisions at the point-of-care, capture variation in patient outcomes among hospitals that reflects differences in care quality, and support ongoing quality improvement efforts in a timely manner. All consulted experts and the TEP strongly endorsed a measure that assessed improvement between the preoperative and postoperative health states (versus postoperative health status) as critical to creating a meaningful outcome measure for this patient population. With TEP support, an

improvement threshold approach to the outcome (over averaging change among patients) was chosen for the following reasons:

- It measures improvement only and discourages surgeons from performing THA/TKA procedures on patients with milder symptoms, as patients with high preoperative PROM scores inherently cannot meet the improvement threshold;
- It equally rewards hospitals performing THA/TKA on patients with moderate and severe symptoms, as it does not define an "end state" that patients must achieve, only substantive improvement from where they started;
- Avoids creating what is known as a ceiling effect, where many patients can meet the outcome criteria, that decreases the ability of the measure to identify performance variation; and
- It has less risk of unintended consequences. Specifically, we were concerned that requiring patients to meet a postoperative minimum symptom state would encourage hospitals and their surgeons to avoid offering THA/TKA surgery to anyone with severe pain and/or limited function, the people most in need of surgery.

CORE evaluated improvement thresholds for the HOOS, JR and KOOS, JR. One approach considered for defining a minimum change threshold was to use 0.5 of the standard deviation (SD). In this method, 0.5 SD of the mean preoperative PROM score for all patients in the sample is calculated. Improvement is identified as a postoperative PROM score which is equal to or greater than the sum of the preoperative score plus the 0.5 SD. Another approach considered was a fixed threshold among those identified by the developers of the HOOS, JR and KOOS, JR, who identify an anchor-based minimal clinically important difference (MCID) (18 points for HOOS, JR and 14 points for KOOS, JR) and an anchor-based SCB (22 points for HOOS, JR and 20 points for KOOS, JR) for both PROMs using a similar population of THA/TKA recipients, although with a two-year follow-up instead of a one-year follow-up.[76] The MCID reflects the smallest change a patient would identify as reflecting a change in their health, while the SCB is the amount of change a patient would identify as "great improvement" following the procedure.[77] As the PROM data collected for the THA/TKA PRO-PM did not include an anchor question, in order to minimize patient and provider burden, these anchor-based thresholds presented valid alternatives.

We calculated patient-level change in PROM scores between the preoperative and the postoperative PROM scores using the 5% SD and the MCID and SCB thresholds proposed by Lyman and colleagues. Improvement was defined as a binary outcome of meeting or exceeding a threshold (yes) or not (no) following the THA or TKA procedure. We then calculated the percentage of patients meeting or exceeding improvement at each threshold at the hospital level for a review of the distribution of unadjusted observed hospital outcomes. Preliminary data analysis indicated that most patients met the 0.5 SD threshold with limited variation among hospitals. Also, because the "amount" of change is calculated using in-sample statistics, the target threshold can and likely will change over time which complicates quality improvement efforts by hospitals. Likewise, the MCID thresholds did not appear to capture the substantial change that patients in our Patient Working Group told us patients expected for such a procedure.

Based on our analyses of published literature and measure development data and with considerable stakeholder input, CORE determined that SCB thresholds for the HOOS, JR (22-point change) and the KOOS, JR (20-point change) best met the criteria: 1) understood by patients, providers, and stakeholders; 2) clinically meaningful to patients; and 3) capture variation in patient outcomes among hospitals that reflects differences in care quality among hospitals.

**Rationale for 90 to Zero Day Preoperative PROM Data Collection**

Clinical experts agree that preoperative PROM data provides not only a baseline assessment of the patient's preoperative health status but also can provide critical risk variables for predicting how that patient will respond to surgery. Many US registry-based efforts to collect PROM data do not specify that preoperative PROM data be collected within a defined window. Clinical experts noted that PROM data collected too far in advance of THA/TKA may not accurately reflect patients' baseline statuses before surgery, although there are limited data to inform the selection of an appropriate preoperative data collection window. PROMs collected prior to elective primary THA/TKAs — which are most often performed for osteoarthritis, a chronic, slowly progressive, degenerative joint process — are unlikely to vary much in the immediate preoperative time period.[78]

In collaboration with our TEP, we evaluated two potential timeframes for collecting preoperative data for this measure: within 90 days before surgery and within 30 days before surgery. We selected these specific options because 1) there is precedent for a 90-day data collection timeframe and 2) clinical experience demonstrates higher response rates and greater physician agreement with a shorter preoperative PROM data collection timeframe.[79] Clinical experts noted that a 30-day window corresponds to The Joint Commission requirement for a history and physical examination within 30 days of surgery.[80] However, the TEP recommended a 90-day preoperative window for data collection. All surgeons on the TEP believed that elective primary THA and TKA candidates were unlikely to have significant changes in preoperative PROM scores within 90 days of surgery, and indicated the additional time to collect data would increase response rates, particularly as the precise data collection mechanism has yet to be specified and is likely to vary across hospitals and surgical practices.

**Rationale for 300 to 425 Day Postoperative PROM Data Collection**

Postoperative PROM collection ranges from as early as three months after surgery to several years after surgery.[81-83] Regional and national registries most commonly collect PROM data within six to 12 months following THA/TKA, and clinical experts and published literature indicate that full clinical rehabilitation is often not reached until one year after surgery.[84,85]

We performed a systematic literature review to examine the differences in PROM results at different time points after surgery: three-, six-, and 12-months post-surgery. In total, we identified seven articles that collected PROM data at both three and six months postoperatively.[84-90] Of these seven, only one found both statistically and clinically meaningful differences between preoperative and three-month postoperative PROM assessments; the remaining studies demonstrated continued clinical improvements between three and six months after surgery. We identified five articles that collected PROM data at six and 12 months after THA/TKA.[84-86,90,91] Johansson et al. found that WOMAC scores at

six months after THA were no different than those at 12 and 24 months after THA. Three studies of TKA recipients demonstrated continued improvement at three, six, and 12 months after surgery.[84,85,90] PROM data from the New Zealand Joint Registry demonstrate six-month PROM results are highly correlated with revision rates within two years following THA/TKA,[81] indicating that PROM results collected as early as six months following THA/TKA may be adequate for the purposes of quality measurement, even if the patients have not reached their final functional statuses.

There has been a consensus among the surgeons on the TEP that THA patients improve more quickly postoperatively than TKA patients, although a TEP member noted that total joint replacement data support that many THA and TKA patients show clinically significant improvement at six months. Some surgeons advocate early PROM assessment after surgery to inform clinical decision making and the need for further surgical intervention. Others advocate assessments at six and/or 12 months after surgery to assess clinical improvement. A majority of clinicians providing input on the measure identified a one-year follow-up anniversary for long-term assessment after THA/TKA surgery.

With regards to data collection, both published literature[90] and anecdotal experience of THA/TKA surgeons collecting PROM data indicate that response rates decline over time and the losses to follow-up three to six months after THA/TKA are not negligible. In addition, these losses may be greater among under-resourced patient populations. This finding differs from the UK experience, where response rates were higher among patients living in lower socioeconomic areas, but is consistent with experience in clinical areas outside orthopedics collecting PROMs.[92-95]

During early measure development in 2013 and 2014, input from TEP members, public comments, literature, and registry experiences informed the initial decision to define the postoperative PROM data collection timeframe as between 270 days and 365 days following THA/TKA. This time period was finalized for voluntary CJR PRO and risk variable data collection and the measure testing results in this report reflect these data. The TEP concurred with this data collection window.

More recent stakeholder input, following several years' experience of PRO data collection through the CJR model, reflects concern that the postoperative data collection window established for PRO data collection in CJR limits postoperative data collection. Clinical experts on the Orthopedic TEP and the Orthopedic Clinical Working Group convened by CORE in 2020 expressed concern that the 365-day upper limit on postoperative data capture missed patients who schedule "one-year follow-up" appointments beyond 12 months after surgery. They strongly advocated for revising the postoperative data collection window to better align with clinical practice. Therefore, the final measure specifications for this measure define the postoperative PROM data collection timeframe as between 300 days and 425 days following THA/TKA. We believe this timeframe will capture patients' full improvement and align with one-year follow-up visits.

## 2.5 Attribution

This THA/TKA PRO-PM is a hospital-level measure, intended to attribute patient improvement to the inpatient facility which performed the procedure. This attribution reflects the importance of the care

and coordination of multiple providers in the clinical outcomes for THA/TKA patients. The goal of this measure is to capture the full spectrum of care in order to incentivize collaboration and shared responsibility for improving patients' health and reducing the burden of their disease.

## 2.6 Risk Adjustment

Risk adjustment is applied to outcome measurement to account for patient characteristics and other medical conditions that are clinically relevant, related to the outcome, and outside of the control of the hospital. Risk adjustment puts measured entities on a level playing field when comparing performance across providers. Accounting for these risk factors allows differences in quality of care to emerge.

We identified an extensive list of risk variables for consideration in the development of the risk model(s) through a systematic literature review and environmental scan, as well as from orthopedists surveyed about what risk variables they consider important in predicting THA/TKA outcomes. In consultation with our expert clinical consultant and the TEP, and through detailed public comments from specialty societies, we focused on candidate risk-adjustment variables of interest that were clinically relevant and had an evidence-based relationship with clinical outcomes following elective primary THA or TKA. Likewise, we considered several potential data sources, including administrative claims, registry- or clinician-provided data, and patient-reported sources.

We used the comprehensive list of candidate risk variables obtained through expert and public input to survey our TEP on their assessment of each risk variable's priority (survey results in Appendix D). In addition, as noted above, we collaborated with orthopedic societies and individual orthopedic practices to conduct a medical record review to evaluate the feasibility, uniformity, and reliability of clinical data elements prioritized by orthopedists by performing a medical record review at seven practices across the country (Appendix B).

The candidate risk variables were categorized and prioritized according to the following criteria:

- **Availability:** Estimate of availability of data based upon existing data sources and/or results of pilot feasibility assessment;
- **Importance:** Estimate of the clinical importance of risk variable based upon published literature and clinical expert input, including the TEP;
- **Ease of Collection:** Estimate of time and effort (patient and provider) to collect risk variable *beyond* the additional burden of de novo collection of the PROM surveys;
- **Reliability:** Estimate of reproducibility of risk variable across providers and patients; and
- **Validity:** Estimate of how well the risk variable captures the intended clinical assessment.

Not all candidate variables prioritized in our stakeholder survey were included in CJR. The AAHKS and AAOS held a joint in-person PROs Summit for Total Joint Arthroplasty on August 31, 2015, to create a consensus-based parsimonious list of prioritized risk variables for inclusion in CJR that eliminated some high-priority variables considered less feasible and/or reliable at this time. This was submitted as part of the public comment on the proposed CJR model.

The final list of risk variables selected for CJR PRO data collection is provided in Appendix C. In addition to these risk variables, all diagnostic codes from administrative claims during the 12 months prior to the THA/TKA procedure or secondary diagnosis codes during the index admission, except those associated with potential complications during the index admission, were evaluated for possible inclusion in the risk model (see Appendix E).

The final risk variables included in the risk model are as follows:

- Age, in years
- Male sex
- BMI, in kg per $m^2$
- Health literacy (assessed by response to Single Item Literacy Screener questionnaire, "Comfort Filling Out Medical Forms by Yourself")
- Back pain at preoperative assessment (Quantified Spinal Pain: Patient-reported Back Pain, Oswestry Disability Index question)[96,97]
- Pain in non-operative lower extremity joint (Total painful joint count: Patient-reported in Non-operative Lower Extremity Joint)[97]
- Narcotic use for >90 days
- Baseline PROMIS-Global Mental Health subscale score
- Severe infection; other infectious diseases (CC 1, 3-7)
- Liver disease (CC 27-31)
- Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)
- Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)
- Depression (CC 61)
- Other psychiatric disorders (CC 63)
- Coronary atherosclerosis or angina (CC 88-89)
- Vascular or circulatory disease (CC 106-109)
- Renal failure (CC 135-140)

The baseline PROMIS-Global Mental Health subscale score, noted in the list above, was included following the suggestion by several stakeholders that this subscale might be a better fit for risk adjustment than for outcome measurement. Published literature supports mental health as a predictor of THA/TKA outcomes. This mental health subscale score is measured using one of the two following global PROMs:

- The PROMIS-Global, or
- The VR-12 Health Survey.

Global PROM data collected using the VR-12 is linked to PROMIS-Global scores using a crosswalk created by PROsetta Stone.[98] This allows for aggregating global PROM results regardless of the global PROM used for collecting PRO data.

**Rationale for Risk Adjustment**

The principles underlying the assessment of individual risk variables in the context of risk model development are summarized below:

- The goal of risk adjustment is to account for patient characteristics that are reasonably beyond the control of the hospital. Therefore, risk variables must represent clinically important risk predictors; that is, they must be predictive of the outcome (in this case, the change in PROs after THA/TKA) and reasonably beyond hospital control.
    - The goal is not perfect risk prediction — this would imply that the hospital has no effect on clinical outcomes (that is, all variation is entirely explained by patient characteristics and healthcare providers have no effect on clinical outcomes). We know this is not true — providers can improve care and outcomes by active quality improvement efforts (such as patient education; adjustments to patient care before, during, and after surgery; etc.).
- Risk variables must be feasible to collect and report. If a variable creates a data collection burden to patients, surgeons, hospitals, or the healthcare system, the incremental value of including the variable in the risk model should significantly outweigh the burden.
    - The definition of burden is subjective. This measure can only be implemented by requiring that hospitals, surgeons, and patients collect the PROM and relevant risk variables data both before and after the THA/TKA. The TEP recommended that we collect both a global PROM (the PROMIS-Global or VR-12) and a hip- or knee-specific PROM (the HOOS, JR or KOOS, JR). It is our goal to minimize any *additional* data collection requirements beyond the PROM surveys, if possible.
- Risk variables must be reliably and consistently defined so the risk variables carry the same information across all patients and hospitals.

### 2.6.1 Social Risk Factors

Our goal in selecting risk factors for adjustment was to develop a parsimonious model that included clinically relevant variables strongly associated with achieving the SCB following an index procedure. We used a two-stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, and then considering the potential addition of social risk factors. While health literacy also reflects social risk, our patient and technical experts strongly supported including health literacy in the risk model for a PRO-based measure, due to the nature of PRO data requiring patients to complete survey instruments as part of the measurement. For this reason, we included it in the candidate risk variable list during the initial stage of risk variable selection and in the final risk model.

To explore the impact of social risk on the measure outcome, we tested associations of dual eligibility and the AHRQ SES index lowest quartile (low SES) with the achievement of the SCB improvement

following elective primary THA/TKA. Likewise, we included race in our social risk factor analyses based upon literature specifically documenting racial and ethnic disparities in THA/TKA offer and acceptance rates, as well as outcomes.[99,100]

We examined the associations of dual eligibility, low SES, and race with the measure outcome using the **Development Dataset** with bivariate and multivariate analyses. Bivariate and multivariate analyses showed no statistically significant association between non-white race and SCB improvement, or low SES and SCB improvement; dual eligibility was borderline significant at the bivariate level and statistically significant in the risk model (see Section 3.3.1, Table 6 and Table 7). However, accounting for these risk factors in the model had little effect on hospital-specific risk-standardized improvement rates (RSIRs). When RSIRs calculated with no social risk factors in the risk model were compared to RSIRs calculated with each of the three risk factors individually included in the risk model, correlation coefficients indicated near perfect correlation in our data (see Section 3.3.1, Table 8).

Additional analyses of hospital proportion of dual eligible patients by hospital RSIRs (see Section 3.3.1, Figure 2) indicate that hospitals with the lowest proportion of dual eligible patients and those hospitals with the highest proportion of dual eligible patients have similar RSIR distributions. These data do not provide evidence of significant differences in RSIRs due to the proportion of a hospital's patients with dual eligibility.

Given these results, we have not included dual eligibility, AHRQ SES index, or non-white race in the risk model(s) at this time. However, given the associations between social risk factors and response in our data and supported by published literature,[101-103] we have included these social risk variables in the statistical approach to addressing non-response bias (see Section 2.7.1 below). In addition, due to existing disparities in access to and outcomes for THA/TKA, the TEP and other stakeholders strongly urged CMS to consider accounting for social risk in measure implementation (such as through stratification) to avoid unintended consequences. We will continue to assess the impact of social risk on this measure moving forward.

## 2.7 Statistical Approach to Model Development

The total dataset for measure development and testing included data from the 238 CJR participant hospitals that submitted complete preoperative and postoperative PRO and risk variable data for at least one elective primary THA/TKA procedure performed from July 1, 2016 through June 30, 2017. Complete PRO and risk variable data were defined as the submission of preoperative PROM and risk variable data with no missing or out-of-range values for required data elements and that could be matched to postoperative PROM data with no missing or out-of-range values for an elective primary THA/TKA procedure identified in claims data for the measurement period. The number of patients with complete PRO data for an elective primary THA or TKA procedure (excluding patients with staged elective primary THA/TKA procedures during the measurement period, defined as two or more procedures performed during separate inpatient admissions) was 11,270. These data were randomly divided 60%/40% into a **Development Dataset** and a **Validation Dataset**.

- The **Development Dataset**, used for risk model development, included 230 hospitals and 6,734 patients, of whom 2,252 had THA procedures and 4,482 had TKA procedures.
- The **Validation Dataset**, used for model validation, included 219 hospitals and 4,536 patients, of whom 1,560 had THA procedures and 3,006 had TKA procedures.
- A **Combined Dataset** of 123 hospitals including only those hospitals with at least 25 THA and TKA patients with complete PRO data submission was used for reliability and validity testing. A case-volume cut-off of 25 was selected as it provided high measure-result reliability and was consistent with volume thresholds used for public reporting of claims-based measures with which this measure was intentionally harmonized.

In the building of risk models for claims-based measures, CORE has previously used the strength of association between the risk variable and the measure outcome to empirically guide risk variable selection. When expert input deems it appropriate, we force in additional risk variables, such as those that indicate frailty, that might have an important influence on the measure outcome and yet might not be selected for the model based purely on statistical considerations. In this way, CORE's risk models have always reflected both empirical data and clinical input. This approach has produced robust risk models that have been repeatedly and successfully validated against medical record data.

For this measure, we applied the same principles, but recognize that PRO-PM development, particularly that based upon a voluntary data sample, may require a greater reliance on clinical input to select risk variables than traditional claims-based outcome measures. Therefore, for this measure, we conducted analyses to evaluate two approaches to risk model development for each PROM outcome — one used a purely data-driven approach (referred to as the empirically derived model) and another used candidate risk variable selection based on empirical findings in the literature, review of data-driven risk factors, and iterative TEP and clinical expert input and ranking of importance and feasibility of risk variables for a THA/TKA PRO-PM (referred to as a clinically derived model).

For the empirically derived model development, CORE identified appropriate risk variables as follows:

- Determined which candidate variables were consistently and strongly associated with a change in PROM scores.
- Examined model performance while including potential risk variables.

We conducted bivariate analyses to examine the association of each potential candidate variable with the outcome to determine the final list of candidate variables. We then conducted bootstrap analyses (1,000 iterations) using logistic regression to assess which of the risk variables were more consistently identified as those associated with outcome in the development cohort. We used the bootstrapping approach to ensure we identified a consistent and stable risk model.

We then created preliminary risk models for THA and TKA separately, consisting of variables more frequently identified as being significantly associated with an outcome. For purposes of measure specification, we included the risk variables that were selected in at least 30% of the models via bootstrapping for evaluation and forced age and sex in the final risk models. We developed a logistic

regression model of improvement on the PROM score using the approach described in the publicly reported 30-day hospital-level risk-standardized mortality or 30-day hospital-level risk-standardized readmission measures and selected the best model using the logistic regression model with the stepwise selection method based on 1,000 bootstrapping samples.[104,105] We examined the direction of variables for clinical sensibility. We also tested the models by examining the model performance (C-statistics), model calibration (lack of fit), model discrimination in terms of predictivity (range of observed outcomes among deciles of predicted outcomes), and distribution of model residuals. We calculated the model estimates, the coefficients, and 95% confidence intervals (CIs) for risk-adjustment variables in the final model in the development and validation samples.

For the clinically derived model, we identified candidate risk variable selection based on empirical findings through a thorough literature review, exploration of data-driven risk factors, and iterative TEP and expert input including a survey of the TEP that asked individuals to rank the importance and feasibility of clinical variables for use in a PRO-PM risk model and input from clinical consultants to finalize a list of clinically relevant and important risk variables for risk adjustment of a THA/TKA PRO-PM. As with the empirically derived models, we assessed model performance by examining the model performance (C-statistics), model calibration (lack of fit), model discrimination in terms of predictivity (range of observed outcomes among deciles of predicted outcomes), and distribution of model residuals. We calculated the model estimates, the coefficients, and 95% CIs for risk-adjustment variables in the clinically derived model in the **Development** and **Validation Datasets**.

**Rationale for Finalizing the Clinically Derived Risk Model**

Given that risk model developed with the clinically derived approach illustrated equivalent performance to and greater face validity than the empirically derived approach, we finalized the clinically derived model. We recognize that PRO-PM development, particularly that based upon a voluntary data sample, may require a greater reliance on clinical input to select risk variables than the development of traditional claims-based outcome measures for which there is abundant, comprehensive data for each patient. TEP feedback indicated support for this approach.

CORE developed the finalized risk model using risk variables endorsed and prioritized by the TEP and by expert orthopedic consultants, to appropriately risk adjust for the patient clinical comorbidities most relevant in THA and TKA. Model performance statistics were highly comparable to those based upon the empirically derived approach, and scatterplot analyses comparing HOOS, JR- and KOOS, JR-specific models demonstrated a high correlation of hospital-specific RSIRs regardless of the model approach and for both THAs and TKAs (see Figure F1 and Figure F2 in Appendix F). The empirically derived approach used the strength and prevalence of risk variables to statistically select the most frequently and strongly associated variables with the measure outcome in bootstrapped iterations. Both approaches — clinical and empirical — produced risk models with similar performance characteristics and predictive ability. Both the clinical and empirical risk models for SCB on the HOOS, JR and KOOS, JR included health literacy, back pain, pain in the non-operative joint, and preoperative global mental health variables. However, the empirically derived models contained a greater proportion of risk variables with less face

validity. For example, while the empirical model included risk variables clinically relevant to PROs following elective primary THA or TKA, such as disorders of the vertebrae and spinal discs, these models also included risk variables with less clinical relevance, such as ear, nose, and throat disorders. This likely represents the overall greater health of a patient population undergoing these elective procedures. It also reflects that we drew data from the prior 12 months of inpatient and outpatient claims for risk adjustment.

### 2.7.1 Response Bias

Due to the voluntary nature of and the burden of novel data collection for PRO data, we understand that accounting for potential non-response bias is important for this measure. Also, the fact that poorly or incompletely collected data may be asymmetrically distributed across lower socioeconomic or disadvantaged populations has the potential to directly affect measure scores. With a thorough literature search, we identified several approaches for missingness (covariates adjustment in regression, submission score adjustment in regression, and stabilized inverse propensity score weighted regression). Following consultation with a statistical expert, we decided to address potential response bias using stabilized inverse probability weighting (IPW), as it would not modify the risk model, and would not assume the form of a relationship between submission score and outcome (as suggested by Garrido[106] and by Thoemmes and Ong[107]).

For this approach, we performed the following steps:

1) All eligible THA/TKA procedures performed during the measurement period at the 238 hospitals submitting complete PRO and risk variable data for at least one of these procedures were identified via CMS claims data (N=39,356 procedures).
2) These eligible THA/TKA procedures were categorized into one of three PRO response groups:
   a) Procedures for which complete PRO and risk variable preoperative data and complete PRO postoperative data were submitted ("complete PRO submission," N=11,270).
   b) Procedures for which incomplete PRO and risk variable data were submitted (including submissions with missing data elements and submissions of only preoperative PRO data or only postoperative PRO data ("incomplete PRO submission," N=10,133).
   c) Procedures for which no PRO data were submitted ("no response," N=17,953).
3) We compared patient characteristics and clinical comorbidities across the three PRO response groups and determined there were statistical differences in case mix.
4) We conducted a literature review and identified the following variables associated with unit non-response to PROM survey data that were also available in our data: age, sex, race, low SES, and postoperative complication following hip or knee procedures.[101-103,108]
5) Additional variables associated with PRO submission in our data were identified through multinomial logistic stepwise regression.
6) Propensity scores were calculated using a multinomial logistic regression where the outcome was 1) complete PRO submission, 2) incomplete PRO submission, and 3) no response.

7) Stabilized IPWs were calculated for each of the three groups. For the complete responders, the stabilized weights were calculated using the following formula: $\frac{P(Z=1)}{P(Z=1|x)}$ where $(Z = 1)$ represents the complete responders. Stabilized weights produce estimates with smaller variance and less extreme values compared to using the standard non-stabilized weights calculated in the following way: $\frac{1}{P(Z=1|x)}$. (See Section 3.7, Table 12, for the distribution of the stabilized weights with mean 1.00 and standard deviation of 0.26.)
8) The stabilized IPWs were incorporated into the hierarchical risk-adjustment model for SCB improvement following elective primary THA/TKA and used in the calculation of the risk-adjusted and bias-adjusted RSIRs.

Incorporating the stabilized weights in the calculation of the RSIRs helps to reduce bias due to non-response by giving higher weight to patients who were less likely to respond and deflating the weight of patients who were more likely to respond, based on patient characteristics. Weighting the responders based on their likelihood of response, given their patient characteristics, helps reduce non-response bias in our RSIR measure.

Among the 238 hospitals submitting at least one complete PRO submission for an eligible THA/TKA procedure during the measurement period, 389 (0.89%) patients died before having the opportunity to complete postoperative PRO data. Given the small number of deaths, we excluded those who died within nine months of the procedure from the propensity score model.

## 2.8 Calculation of Measure Score

With the list of variables for both the clinically derived and empirically derived risk models identified, we estimated the hospital-specific RSIR using a hierarchical logistic regression model (hierarchical model). This strategy accounts for within-hospital correlation of the observed outcome among patients and accommodates the assumption that underlying differences in the quality of care across hospitals lead to systematic differences in patient outcomes. This approach models the log odds of patient improvement on the PROM as a function of patient demographics and clinically relevant comorbidities with an intercept for the hospital-specific random effect.

We then calculate the hospital-specific RSIRs, which were calculated as the ratio of a hospital's "predicted" number of improvements to "expected" number of improvements multiplied by the overall observed improvement rate. The expected number of improvements for each hospital (denominator) was estimated using its patient mix and the average hospital-specific intercept (i.e., the average intercept among all hospitals in the sample). The predicted number of improvements for each hospital (numerator) was estimated given the same patient mix but an estimated hospital-specific intercept. Operationally, the expected number of improvements for each hospital was obtained by summing the expected improvements for all patients in the hospital. The expected improvement for each patient was calculated via the hierarchical model, which applies the estimated regression coefficients to the observed patient characteristics and adds the average of the hospital-specific intercept. The predicted

number of improvements for each hospital was calculated by summing the predicted improvements for all patients in the hospital. The predicted improvement for each patient was calculated through the hierarchical model, which applies the estimated regression coefficients to the patient characteristics observed and adds the hospital-specific intercept.

More specifically, we used a hierarchical model to account for the natural clustering of observations within hospitals. The model employs a logit link function to link the risk factors to the outcome with a hospital-specific random effect:

Let $Y_{ij}$ denote the outcome (equal to one if the patient has an improvement, zero otherwise) for patient *i* at hospital *j*; $\boldsymbol{Z}_{ij}$ denotes a set of risk factors for patient $i$ at hospital $j$; and $n_j$ is the number of index admissions to hospital $j$. We assume the outcome is related linearly to the covariates via a logit function:

Logistic Regression Model

$$logit\big(\text{Prob}(Y_{ij}=1)\big) = \alpha + \boldsymbol{\beta Z}_{ij} \qquad (1)$$

and $\boldsymbol{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \dots, Z_{pij})$ is a set of $p$ patient-specific covariates.

To account for the natural clustering of observations within hospitals, we estimate a hierarchical logistic regression model that links the risk factors to the same outcomes and a hospital-specific random effect.

Hierarchical Logistic Regression Model

$$logit\big(\text{Prob}(Y_{ij}=1)\big) = \alpha_j + \boldsymbol{\beta Z}_{ij} \qquad (2)$$

$$\text{where } \alpha_j = \mu + \omega_j;\ \omega_j \sim N(0, \tau^2) \qquad (3)$$

where $\alpha_j$represents the hospital-specific intercept, $\boldsymbol{Z}_{ij}$is defined as above, μ is the adjusted average intercept over all hospitals in the sample, $\omega_j$ is the hospital-specific intercept deviation from $\mu$, and $\tau^2$ is the between-hospital variance component. This model separates within-hospital variation from between-hospital variation. Both the hierarchical logistic regression model and the logistic regression model are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the logistic regression model described in Equation (1) with covariates selected for the clinically derived risk model. We then calculated stabilized IPW from a propensity score analysis using multinomial logistic regression to model three PRO data response groups: complete PRO submission, incomplete PRO submission, and no response (see Section 2.7.1 for a detailed description of the analytic approach to addressing potential response bias). Next, we fit the hierarchical logistic regression model described in Equations (2) and (3) to the corresponding parameters along with the stabilized IPW to adjust for response bias when we ran the model. Lastly, we calculated the risk-standardized improvement rates for each hospital.

## 2.9 Testing

Below we report our approach to evaluate the model performance. It is conducted using the development and validation samples.

### 2.9.1 Model Performance

We assessed model performance in the **Development Dataset** by examining the model performance (C-statistics), model calibration (lack of fit), model discrimination in terms of predictive ability (range of observed outcomes among deciles of predicted outcomes), and distribution of model residuals. We calculated the model estimates as well as the coefficients and 95% CIs for the risk-adjustment variables for the best-performing model in the **Development Dataset**. We assessed risk factors in THA-specific and TKA-specific cohorts to ensure risk prediction for a combined THA/TKA cohort was consistent with that for THA- and TKA-specific cohorts. We compared measure results and risk model performance for the THA-specific, TKA-specific, and the combined THA/TKA cohorts. We then repeated the assessment of model performance for the final combined THA/TKA cohort in the **Validation Dataset**.

We computed discrimination and calibration statistics for assessing model performance[109] for the clinically derived models, including:

1. The area under the receiver operating characteristic (ROC) curve (the C-statistic [also called ROC] is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model can distinguish between a patient with and without an outcome);
2. Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; good discrimination is indicated by a wide range between the lowest decile and highest decile); and
3. Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the **Development Dataset** but fails to provide valid predictions in new patients).

### 2.9.2 Measure Results

Meaningful differences in performance measure scores are assessed by calculating the distribution of hospital-level RSIRs. Variation in hospital-level RSIRs indicates a clinically meaningful quality gap in the delivery of care to patients undergoing elective THA/TKA, as some hospitals can achieve substantially higher rates than the average performer, while other hospitals perform much worse than an average performer.

In addition, statistically significant differences were assessed using a median odds ratio (MOR).[110] The MOR represents the median increase in odds of the patient outcome (an SCB improvement in PROM score from preoperative to postoperative assessment) if a procedure on a single patient was performed by a higher-performing hospital compared to a lower-performing hospital. It is calculated by taking all

possible combinations of hospitals (N=238 hospitals in the total dataset), always comparing the higher-performing hospitals to the lower performing hospitals. The MOR is interpreted as a traditional odds ratio would be.

## 2.9.3 Reliability

### 2.9.3a Data Element Reliability

Data element reliability is evidenced by reliability testing conducted during the development and validation of the joint-specific PROMs on which this THA/TKA PRO-PM is based.

**HOOS, JR Reliability**

**Internal consistency:** The developers of the HOOS, JR[2] assessed internal consistency reliability using the Person Separation Index (PSI). The PSI was used in two data samples, the Hospital for Special Surgery (HSS) cohort and the FORCE-TJR, a nationally representative joint replacement registry. A higher value on the PSI indicates a greater ability to differentiate patients with varying levels of ability, which in turn provides evidence of good internal consistency. For testing internal consistency for the HOOS, JR, a PSI value > 0.7 was considered acceptable.[2] The developers also conducted principal component analysis on the standardized residuals to assess HOOS, JR items.

**Test-Retest Reliability:** Test-retest reliability was not tested by developers of the HOOS, JR as it had already been tested in the HOOS in several validation studies.[111-114] Intra-class correlation coefficients (ICCs) between dimensions (Pain, Symptoms, Activities of Daily Living, Sport and Recreation Function, and Quality of Life) were used to determine test-retest reproducibility.

**KOOS, JR Reliability**

**Internal Consistency:** The developers of the KOOS, JR[3] assessed the internal consistency reliability of using the PSI. The PSI was used in two data samples, the HSS cohort and the FORCE-TJR, a nationally representative joint replacement registry. A higher value on the PSI indicates a greater ability to differentiate patients with varying levels of ability, which in turn provides evidence of good internal consistency. For testing internal consistency for the KOOS, JR, a PSI value > 0.7 was considered acceptable.[3] The developers also conducted principal component analysis on the standardized residuals to assess KOOS, JR items.

**Test-Retest Reliability:** Test-retest reliability was not tested by developers of the KOOS, JR as it had already been tested in the KOOS.[74] To examine test-retest reliability, the KOOS was administered to patients twice prior to surgery within a nine-day period. ICCs between dimensions (Pain, Symptoms, Activities of Daily Living, Sport and Recreation Function, and Quality of Life) were used to determine test-retest reproducibility.

#### 2.9.3b Measure Result Reliability

The reliability of measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. Using the **Combined Dataset (Development** and **Validation Datasets)**, we identified the hospitals with ≥ 25 THA/TKA patients with PRO data during the measurement period and assessed signal-to-noise reliability to describe how well the measure can distinguish the performance of one hospital from another.[115,116] The signal is the proportion of the variability in measured performance that can be explained by real differences in performance. Scores can range from zero to one. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to a real difference in performance.

### 2.9.4 Validity

#### 2.9.4a Data Element Validity

Data element validity is evidenced by validity testing conducted during the development and testing of the joint-specific PROMs on which this THA/TKA PRO-PM is based. All validity testing for the HOOS, JR and KOOS, JR instruments was conducted by the PROM developers.[2,3]

#### HOOS, JR Validity

**Responsiveness:** Responsiveness of the HOOS, JR to changes following a THA was evaluated using standardized response means, and then examined against other previously validated PROMs (HOOS domains, WOMAC domains) in the HSS cohort and the FORCE-TJR registry at two years after a THA procedure.[2] A standardized response mean > 0.8 was considered large.[117]

**External Validity**: External construct validity was evaluated using Spearman's correlations between the HOOS, JR, the HOOS, and the WOMAC. A Spearman's correlation coefficient ≥ 0.8 was considered very high external validity.[118] External correlations were assessed using a scatterplot overlying a contour plot based on bivariate kernel density estimation between the HOOS, JR and HOOS domains.[2]

**Floor and Ceiling Effects**: Floor and ceiling effects (percent at worst possible score preoperatively and best possible score postoperatively) were evaluated against the HOOS and the WOMAC instruments.[2]

#### KOOS, JR Validity

**Responsiveness:** Responsiveness of the KOOS, JR to changes following TKA was evaluated using standardized response means, and then examined against other validated PROMs (KOOS domains, WOMAC domains) in the validation cohort.[3] A standardized response mean > 0.8 was considered large.[117]

**External Validity**: External construct validity was evaluated using Spearman's correlations between the KOOS, JR, the KOOS, and the WOMAC. A Spearman's correlation coefficient ≥0.8 was considered very

high external validity.[118] External correlations were assessed using a scatterplot overlying a contour plot based on bivariate kernel density estimation between the KOOS, JR and KOOS domains.[3]

**Floor and Ceiling Effects**: Floor and ceiling effects (percent at worst possible score preoperatively and best possible score postoperatively) were evaluated against the KOOS and the WOMAC instruments.[3]

#### 2.9.4b Empirical Measure Score Validity

To assess empirical measure score validity, we compared the THA/TKA PRO-PM RSIRs to the NQF endorsed THA/TKA Complication Measure (NQF #1550: Hospital-level risk-standardized complication rate [RSCR] following elective primary total hip arthroplasty [THA] and/or total knee arthroplasty [TKA]) This measure estimates the risk-adjusted rate that patients who have experienced an elective primary THA/TKA experience at least one of eight complications within 90 days of the procedure. The RSCR is categorized into three groups: worse than national average, same as national average, and better than national average. Data for the hospital RSCRs from April 1, 2015 through March 31, 2018 were compared to RSIRs for procedures performed July 1, 2016 through June 30, 2017. We examined the distribution of THA/TKA PRO-PM RSIRs by THA/TKA RSCR national categories within hospitals submitting complete PRO data for at least 25 THA/TKA procedures.

Like the THA/TKA PRO-PM we are seeking to validate, NQF #1550 measures outcomes of elective primary THA/TKA procedures. While the outcomes of these two measures are not clinically expected to be perfectly correlated, they both reflect the hospital-level quality of care for patients experiencing elective primary THA/TKA surgery. It is clinically accepted that certain THA/TKA complications, specifically mechanical complications that require further surgical interventions, revision or even removal of the prosthetic joint, are associated with worse PROs such as pain and function. Therefore, we would anticipate that hospitals with higher overall risk-adjusted complication rates in this population (patients undergoing elective primary THA/TKA procedures) would see overall worse PROs. Higher complication rates (RSCRs) would lead to worse clinical outcomes such as increased pain and decreased function, resulting in lower percentages of patients achieving an SCB (lower RISRs). Thus, in terms of the measure results, we expect an inverse association between the RSIRs of the THA/TKA PRO-PM and the RSCRs of THA/TKA complication measure.

Our examination of this association between RSIRs and RSCRs used categories of hospital performance on the THA/TKA complication measure:

- Hospitals whose performance is worse than the national rate (those with higher RSCR),
- Hospitals whose performance is no different than the national rate, and
- Hospitals whose performance is better than the national rate (those with lower RSCR).

With this approach, we expect that hospitals within the RSCR category “Worse than the National Rate” will have lower RSIRs and that hospitals within the RSCR category “Better than National Rate” will have higher RSIRs.

# 3. Results

## 3.1 Measure Cohort

Characteristics of the 6,734 patients in the **Development Dataset** and the 4,536 patients in the **Validation Dataset** are presented in Table 2.

**Table 2. Patient Characteristics in Development and Validation Datasets**

| Variable | Development Dataset N (%) | Validation Dataset N (%) |
|---|---|---|
| Total N | 6,734 | 4,536 |
| Age in years *(Mean, SD)* | 73.63 (5.75) | 73.74 (5.84) |
| Sex: Male | 2,442 (35.97%) | 1,660 (36.60%) |
| Race: Black, non-Hispanic | 254 (3.77%) | 160 (3.53%) |
| White, non-Hispanic | 6,200 (92.07%) | 4,205 (92.70%) |
| Hispanic | 178 (2.64%) | 98 (2.16%) |
| Other | 102 (1.51%) | 73 (1.61%) |
| Bilateral procedure: Yes (vs. No) | 31 (0.46%) | 35 (0.77%) |
| Health literacy (Comfort Filling Out Medical Forms by Yourself): None | 1,000 (14.85%) | 663 (14.62%) |
| A little bit | 518 (7.69%) | 352 (7.76%) |
| Somewhat | 775 (11.51%) | 524 (11.55%) |
| Quite a bit | 1,192 (17.70%) | 853 (18.81%) |
| Extremely | 3,249 (48.25%) | 2,144 (47.27%) |
| Patient-reported back pain (Oswestry index question): None | 2,562 (38.05%) | 1,754 (38.67%) |
| Very mild | 1,661 (24.67%) | 1,074 (23.68%) |
| Moderate | 1,706 (25.33%) | 1,156 (25.49%) |
| Fairly severe | 570 (8.46%) | 391 (8.62%) |
| Very severe/Worst imaginable | 235 (3.49%) | 161 (3.55%) |
| Patient-reported pain in non-operative lower extremity joint: None | 2,298 (34.13%) | 1,552 (34.22%) |
| Mild | 1,640 (24.35%) | 1,125 (24.80%) |
| Moderate | 1,727 (25.65%) | 1,079 (23.79%) |
| Severe | 856 (12.71%) | 635 (14.00%) |
| Extreme | 213 (3.16%) | 145 (3.20%) |
| BMI *(Mean, SD)* | 30.39 (6.01) | 30.46 (6.03) |
| Narcotic use for >90 days | 1,224 (18.18%) | 787 (17.35%) |
| PROMIS-Global Mental Health score *(Mean, SD)* | 49.71 (8.10) | 49.70 (8.05) |
| Severe infection; other infectious diseases | 1,258 (18.68%) | 842 (18.56%) |
| Diabetes or diabetes complications | 1,735 (25.76%) | 1,217 (26.83%) |
| Liver disease | 1,794 (26.64%) | 1,229 (27.09%) |

| Variable | Development Dataset N (%) | Validation Dataset N (%) |
|---|---|---|
| Rheumatoid arthritis and inflammatory connective tissue disease | 750 (11.14%) | 457 (10.07%) |
| Depression | 1,047 (15.55%) | 698 (15.39%) |
| Other psychiatric disorders | 1,105 (16.41%) | 714 (15.74%) |
| Coronary atherosclerosis or angina | 1,622 (24.09%) | 1,138 (25.09%) |
| Vascular or circulatory disease | 1,279 (18.99%) | 862 (19.00%) |
| Renal failure | 905 (13.44%) | 621 (13.69%) |

## 3.2 Attribution

Characteristics of the 230 hospitals in the **Development Dataset**, the 219 hospitals in the **Validation Dataset**, and the 123 hospitals in the **Combined Dataset** of hospitals with ≥25 THA/TKA patients with PRO data (used for reliability and validity testing, and for response bias analyses) are presented in Table 3.

**Table 3. Characteristics of Hospitals in Development and Validation Datasets and Hospitals with ≥ 25 THA/TKA Patients with PRO Data during the Measurement Period**

| | Hospitals in Development Dataset | Hospitals in Validation Dataset | Hospitals in Combined Dataset for Reliability and Validity Testing and Response Bias Analyses (with ≥ 25 THA/TKA Patients with PRO Data) |
|---|---|---|---|
| **Total Hospitals, N** | 230 | 219 | 123 |
| **Median # of Elective Primary THA/TKA Procedures Performed (Q1, Q3)** | 121 (56, 244) | 123 (54, 250) | 209 (114, 300) |
| **Mean % of Patients on Medicaid (SD)** | 18.3% (10.3) | 18.0% (0.1) | 20.4% (11.4) |
| **Region, %** | | | |
| West | 24.8% | 25.1% | 27.6% |
| Midwest | 28.7% | 31.1% | 34.2% |
| Northeast | 23.5% | 21.9% | 17.9% |
| South | 23.0% | 21.9% | 20.3% |
| **Teaching Status, %** | | | |
| Teaching | 46.1% | 44.8% | 48.8% |
| Non-teaching | 53.9% | 55.2% | 51.2% |

## 3.3 Risk Model Performance and Testing

Testing results using the **Validation Dataset** of the risk-adjusted model for SCB improvement following elective primary THA/TKA are presented in Table 4. Risk variable odds ratios (ORs) are adjusted for other risk variables in the model. As previously noted, the SCB outcome allows patients with poor baseline

PRO scores to improve, so some risk variables that might be traditionally considered as predictors of worse outcomes are positively associated with achieving an SCB.

**Table 4. Final Risk Model Variables and Adjusted ORs (Hierarchical Linear Model): Validation Dataset (Patient N = 4,536, Hospital N = 219)**

| Variable | Frequency | OR (95% CI) |
|---|---|---|
| Age in years (Mean, SD) | 73.74 (5.84) | 1.00 (0.98, 1.01) |
| Sex: Male | 1,660 (36.60%) | 0.76 (0.65, 0.88) |
| Procedure: THA | 1,530 (33.73%) | 1.41 (1.20, 1.66) |
| Bilateral procedure | 35 (0.77%) | 1.51 (0.60, 3.83) |
| Health literacy (Comfort Filling Out Medical Forms by Yourself): Not at all *(Reference)* | 663 (14.62%) | -- |
| A little bit | 352 (7.76%) | 1.24 (0.88, 1.76) |
| Somewhat | 524 (11.55%) | 1.96 (1.43, 2.68) |
| Quite a bit | 853 (18.81%) | 2.28 (1.71, 3.03) |
| Extremely | 2,144 (47.27%) | 2.17 (1.68, 2.81) |
| Back Pain: None *(Reference)* | 1,754 (38.67%) | -- |
| Very mild | 1,074 (23.68%) | 0.84 (0.69, 1.01) |
| Moderate | 1,156 (25.49%) | 0.84 (0.69, 1.01) |
| Fairly severe | 391 (8.62%) | 1.01 (0.76, 1.34) |
| Very severe/Worst imaginable | 161 (3.55%) | 1.68 (1.05, 2.68) |
| Pain in non-operative lower extremity joint: None *(Reference)* | 1,552 (34.22%) | -- |
| Mild | 1,125 (24.80%) | 0.84 (0.69, 1.02) |
| Moderate | 1,079 (23.79%) | 0.88 (0.72, 1.08) |
| Severe | 635 (14.00%) | 1.33 (1.03, 1.72) |
| Extreme | 145 (3.20%) | 2.18 (1.31, 3.62) |
| BMI (Mean, SD) | 30.46 (6.03) | 1.00 (0.99, 1.01) |
| Narcotic use for >90 days | 787 (17.35%) | 0.96 (0.78, 1.18) |
| Baseline PROMIS-Global Mental Health score (Mean, SD) | 49.70 (8.05) | 0.99 (0.98, 0.996) |
| Severe infection; other infectious diseases (CC 1, 3-7) | 842 (18.56%) | 0.93 (0.77, 1.12) |
| Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123) | 1,217 (26.83%) | 0.85 (0.60, 1.20) |
| Liver disease (CC 27-31) | 1,229 (27.09%) | 1.21 (0.85, 1.72) |
| Rheumatoid arthritis and inflammatory connective tissue disease (CC 40) | 457 (10.07%) | 0.81 (0.64, 1.02) |
| Depression (CC 61) | 698 (15.39%) | 0.91 (0.72, 1.13) |
| Other psychiatric disorders (CC 63) | 714 (15.74%) | 0.94 (0.75, 1.16) |
| Coronary atherosclerosis or angina (CC 88-89) | 1,138 (25.09%) | 0.74 (0.62, 0.88) |
| Vascular or circulatory disease (CC 106-109) | 862 (19.00%) | 0.93 (0.77, 1.12) |
| Renal failure (CC 135-140) | 621 (13.69%) | 1.09 (0.88, 1.36) |

Model performance statistics of the risk model for meeting or exceeding the SCB improvement threshold are provided in Table 5. In the **Development Dataset**, the C-statistic for the risk model is 0.68 and the predictive ability from the lowest to highest decile is 26% – 82%. In the **Validation Dataset**, the C-statistic for the risk model is 0.69 and the predictive ability from the lowest to highest decile is 26% – 81%. The *calibration indices* (γ0, γ1) used to assess the risk model for meeting or exceeding SCB improvement for the **Validation Dataset** are (-0.08, 1.02).

Results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics. The calculated C-statistic was 0.68 using the **Development Dataset** and 0.69 using the **Validation Dataset** and indicates adequate model discrimination across the cohort models. With both the **Development** and **Validation Datasets**, the model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

The calibration values which are consistently close to zero at one end and close to one at the other end indicates good calibration of the model. If the γ0 in the model performance using the **Validation Dataset** is substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration values of close to zero at one end and close to one on the other end indicate good calibration of the model between the **Development** and **Validation Datasets**.

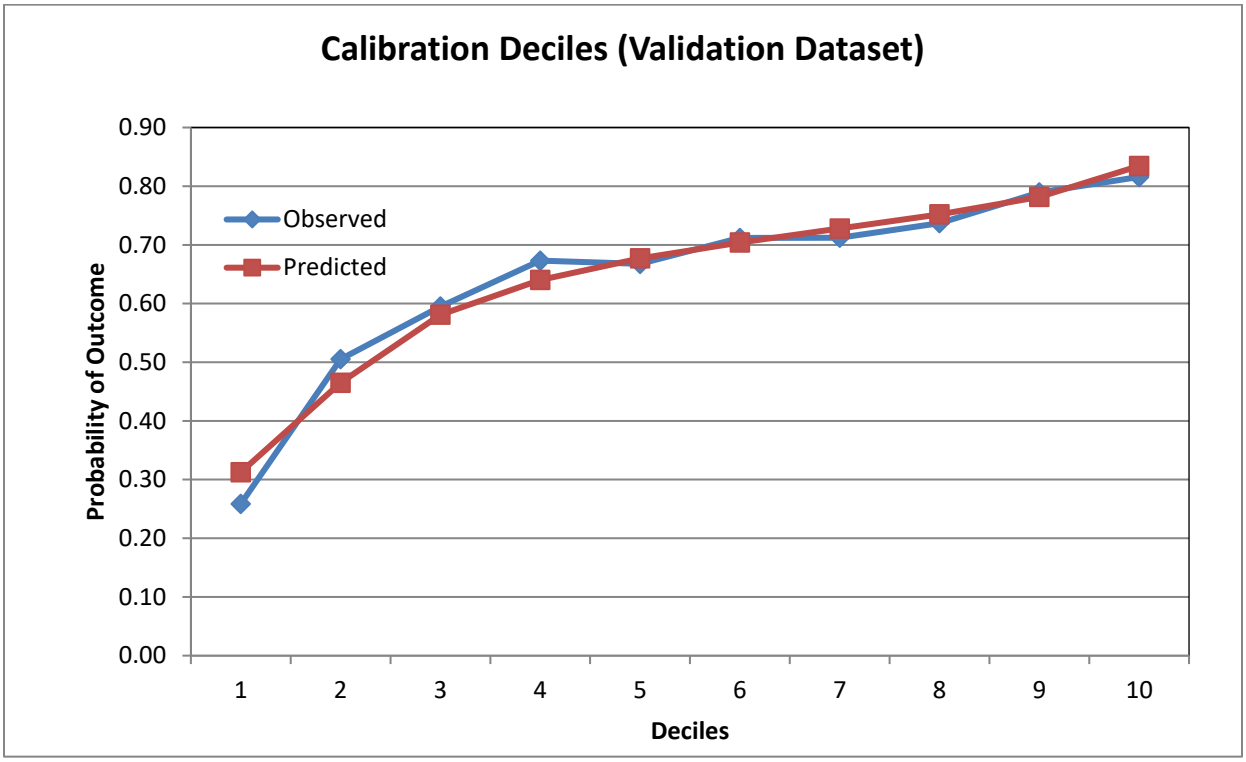
In the risk decile plot (Figure 1), higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. The plot indicates good discrimination of the model and good predictive ability.

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

**Table 5. Model Performance of Risk-Adjusted Model of SCB Improvement Following THA/TKA**

| Model Performance Statistic | Development Dataset | Validation Dataset |
|---|---|---|
| C-statistic | 0.68 | 0.69 |
| Calibration (γ0, γ1) | 0.00, 1.00 | -0.08, 1.02 |
| Predictive ability | 26% – 82% | 26% – 81% |

**Figure 1. Calibration Deciles for the Validation Dataset**




### 3.3.1 Evaluation of Social Risk Factors

Bivariate and multivariate analyses conducted with the **Development Dataset** showed no statistically significant association between AHRQ SES index lowest quartile and SCB improvement, or between non-white race and SCB improvement; dual eligibility was borderline significant (p=0.058) at the bivariate level (Table 6), and statistically significant when entered into the risk model (OR=1.49, 95% CI=1.07, 2.08), indicating that patients with dual eligibility had higher odds of achieving SCB improvement (Table 7). Table 8 provides the mean and range of hospital-specific RSIRs with no social risk factors included in the risk model, with dual eligibility, and with AHRQ SES index lowest quartile individually included in the risk model. Correlation coefficients between RSIRs calculated without social risk factors and RSIRs calculated individually for each of the social risk factors indicates a near perfect or perfect correlation in our data. This was also true when comparing RSIRs calculated without social risk factors with RSIRs calculated including non-white race. The lack of association and effect of these factors may be due to lower case selection in these groups for these elective primary procedures.

**Table 6. Bivariate Associations of Social Risk Factors and Race with SCB Improvement: Development Dataset (Patient N = 6,734, Hospital N = 230)**

| Variable | Frequency (%) of Social Risk Factor among Patients in the Development Dataset | Frequency (%) of Social Risk Factor among Patients Achieving SCB Improvement | Frequency (%) of Social Risk Factor among Patients Not Achieving SCB Improvement | P-value |
|---|---|---|---|---|
| Total | 6,734 (100%) | 4,354 (100%) | 2,380 (100%) | -- |
| Dual eligibility | 206 (3.06%) | 146 (3.35%) | 60 (2.52%) | 0.0580 |
| AHRQ SES index: Lowest quartile | 688 (10.22%) | 446 (10.24%) | 242 (10.17%) | 0.9222 |
| Race: Non-white | 548 (8.14%) | 351 (8.06%) | 197 (8.28%) | 0.7569 |

**Table 7. Adjusted ORs for Social Risk Factors and Race Individually Evaluated in the Risk Model for SCB Improvement: Development Dataset (Patient N = 6,734, Hospital N = 230)**

| Variable | Frequency (%) | Estimate (Standard Error) | OR (95% CI) | C-Statistic for Model Including Social Risk Factor |
|---|---|---|---|---|
| Dual eligibility | 206 (3.06%) | 0.40 (0.17) | 1.49 (1.07, 2.08) | 0.68* |
| AHRQ SES index: Lowest quartile | 688 (10.22%) | 0.04 (0.09) | 1.04 (0.87, 1.25) | 0.68* |
| Race: Non-white | 548 (8.14%) | -0.08 (0.10) | 0.93 (0.76, 1.13) | 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*

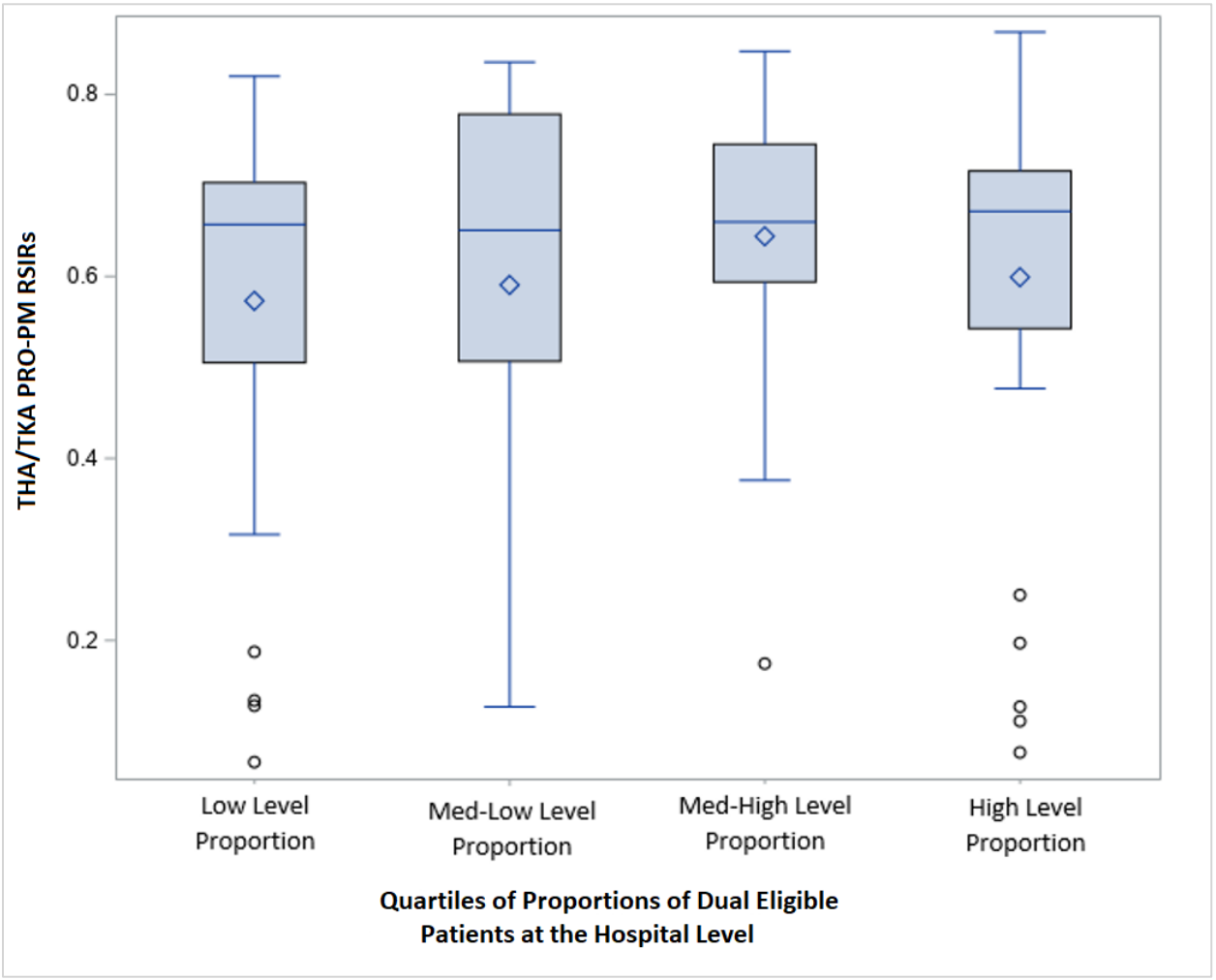
**Table 8. Mean and Distribution of RSIRs Calculated without and with Social Risk Factors and Race in the Risk Model (Development Dataset: Hospitals with ≥25 THA/TKA Patients with PRO Data)**

| | No Risk Factors Included | Dual Eligibility | AHRQ SES Index: Lowest Quartile | Race: Non-White |
|---|---|---|---|---|
| **N (Hospitals)** | 94 | 94 | 94 | 94 |
| **Mean (SD)** | 60.39% (19.85) | 60.40% (19.85) | 60.30% (19.86) | 60.36% (19.87) |
| **Percentile** | -- | -- | -- | -- |
| 100% Max | 86.25% | 86.21% | 86.23% | 86.03% |
| 99% | 86.25% | 86.21% | 86.23% | 86.03% |
| 95% | 81.94% | 81.96% | 82.03% | 81.71% |
| 90% | 79.95% | 79.95% | 79.95% | 80.10% |
| 75% (Q3) | 72.37% | 72.38% | 72.33% | 72.45% |
| 50% (Median) | 66.57% | 66.53% | 66.57% | 66.60% |
| 25% (Q1) | 53.22% | 53.23% | 53.22% | 53.26% |
| 10% | 20.07% | 20.08% | 20.06% | 20.04% |

| | No Risk Factors Included | Dual Eligibility | AHRQ SES Index: Lowest Quartile | Race: Non-White |
|---|---|---|---|---|
| 5% | 14.47% | 14.49% | 14.50% | 14.43% |
| 1% | 8.47% | 8.48% | 8.46% | 8.42% |
| 0% Min | 8.47% | 8.48% | 8.46% | 8.42% |
| **Pearson Correlation Coefficient (With "No Social Risk Factors")** | | 0.9999 | >0.9999 | 0.9997 |

Analysis of hospital proportion of dual eligible patients by hospital RSIRs is provided in Figure 2. The results indicate that hospitals with the lowest proportion of dual eligible patients and those hospitals with the highest proportion of dual eligible patients have similar RSIR distributions.

**Figure 2. THA/TKA PRO-PM RSIRs by Quartiles of Hospitals Grouped by Proportion of Dual Eligible Patients**



## 3.4 Measure Results (RSIRs)

Table 9 provides the mean and distribution of hospitals' RSIRs, weighted for non-response bias. The RSIRs ranged from 6.65% to 86.84% (median: 66.49%). The variation in the RSIRs suggests that there are meaningful differences in performance measure scores across hospitals. The interquartile range represents a difference of 18 percentage points, and the difference between the 10th and 90th percentiles (20.94% and 78.85%, respectively) is nearly 58 percentage points. This variation indicates an important quality gap among hospitals.

**Table 9. Mean and Distribution of RSIRs for Risk Model of SCB Improvement Following Elective Primary THA/TKA (Hospitals with ≥25 THA/TKA Patients with PRO Data)**

| | RSIRs (Combined Dataset) |
|---|---|
| **N (Hospitals)** | 123 |
| **Mean (SD)** | 60.16% (19.58) |
| **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% Min | 6.65% |

Results of the analyses to examine the MOR were 3.44, with upper and lower 95% confidence bands of 3.385 and 3.485. This MOR suggests significant and substantial increases in the likelihood of SCB improvement by higher-performing hospitals compared to lower-performing hospitals. At the hospital level, the MOR value indicates that a patient is 3.44 times more likely to achieve SCB improvement if their elective primary THA/TKA procedure was performed by a higher-performing hospital than by a lower-performing hospital.

## 3.5 Reliability

### 3.5.1 Data Element Reliability Results

Data element reliability results are reported for reliability testing conducted during the development and testing of the joint-specific PROMs on which this THA/TKA PRO-PM is based.

**HOOS, JR Reliability**

**Internal Consistency:** The developers of the HOOS, JR[2] assessed the internal consistency reliability of using the PSI. Internal consistency of the HOOS, JR on the PSI were 0.86 in the HSS cohort and 0.87 in the FORCE-TJR cohort. Results of a principal component analysis conducted on the standardized residuals indicated that the six HOOS, JR items existed in a single dimension.[2]

**Test-Retest Reliability:** Test-retest reliability was not tested by developers of the HOOS, JR as it had already been tested in the HOOS in several validation studies.[111-114] ICCs were used to determine test-retest reproducibility and ranged from 0.75 to 0.97 in the validation studies. Specifically, the Pain and Activity of Daily Living domains, from which HOOS, JR pain and functioning questions are drawn, had ICCs of 0.83 – 0.89 (Pain sub-scale) and 0.86 – 0.94 (Activity of Daily Living sub-scale).

**KOOS, JR Reliability**

**Internal Consistency:** The developers of the KOOS, JR[3] assessed the internal consistency reliability of using the PSI. Internal consistency of the KOOS, JR on the PSI were 0.84 in the HSS cohort and 0.85 in the FORCE-TJR cohort. Results of a principal component analysis conducted on the standardized residuals indicated that the seven KOOS, JR items existed in a single dimension.[3]

**Test-Retest Reliability:** Test-retest reliability was not tested by developers of the KOOS, JR as it had already been tested in the KOOS.[74] ICCs were used to determine test-retest reproducibility and ranged from 0.75 to 0.93. Specifically, the Pain, Activity of Daily Living and Symptom domains, from which KOOS, JR pain, functioning and stiffness questions are drawn, had ICCs of 0.85 (Pain sub-scale), 0.75 (Activity of Daily Living sub-scale), and 0.93 (Symptoms).

The reliability results from the literature demonstrate that the HOOS, JR and the KOOS, JR PROM instruments are sufficiently reliable and exceed accepted norms for reliability testing. The results assessing internal consistency indicated PSI values of 0.86 – 0.87 for the HOOS, JR[2] and 0.84 – 0.85 for the KOOS, JR[3], indicating the ability of the instruments to differentiate patients with varying levels of pain and functioning, which in turn provides evidence of good internal consistency. Test-retest reliability results for the HOOS domains from which HOOS, JR questions were drawn (Pain and Activity of Daily Living domains) revealed high ICCs. Likewise, test-retest reliability for the KOOS domains from which the KOOS, JR questions were drawn (ICCs of 0.75 – 0.93) provided evidence good reliability.

### 3.5.2 Measure Score Reliability Results

The signal-to-noise ratio (Table 10) yielded a median reliability score of 0.959 (range: 0.896 – 0.992). Interquartile range was 0.037.

**Table 10. Signal-to-Noise-Reliability, Hospitals with Volume ≥25 THA/TKA Patients with PRO Data**

| | # Hospitals | Mean (SD) | Median | Minimum | Maximum | Interquartile Range | |
|---|---|---|---|---|---|---|---|
| | | | | | | Q1 | Q3 |
| **Signal-to-Noise** | 123 | 0.952 (0.263) | 0.959 | 0.896 | 0.992 | 0.935 | 0.972 |

The signal-to-noise reliability of 0.952 indicates excellent reliability. Our interpretation of these results is based on standards established by Landis and Koch:[119]

- <0 = Less than chance agreement
- 0 – 0.2 = Slight agreement
- 0.21 – 0.39 = Fair agreement
- 0.4 – 0.59 = Moderate agreement
- 0.6 – 0.79 = Substantial agreement
- 0.8 – 0.99 = Almost Perfect agreement
- 1 = Perfect agreement

## 3.6 Validity

### 3.6.1 Data Element Validity Results

Data element validity results are reported for validity testing conducted during the development and testing of the joint-specific PROMs on which this THA/TKA PRO-PM is based.

**HOOS, JR Validity**

**Responsiveness**: Standardized response means for the HOOS, JR, relative to other PROMs measuring postoperative hip improvement, were 2.38 (95% CI, 2.27–2.49) in the HSS data and 2.03 (95% CI, 1.84–2.22) in the FORCE registry data.[2]

**External validity:** Correlations between the HOOS, JR and HOOS Pain domain were 0.87 (95% CI, 0.86–0.89) in the HSS data and 0.87 (95% CI, 0.84–0.90) in the FORCE registry data. Correlations between the HOOS, JR and HOOS Activity of Daily Living domain were 0.94 (95% CI, 0.93–0.95) in the HSS data and 0.94 (95% CI, 0.93–0.96) in the FORCE registry data. Likewise, correlations between the HOOS, JR and the WOMAC Pain domain was 0.84 (95% CI, 0.81–0.86) in the HSS data and 0.85 (95% CI, 0.81–0.88) in the FORCE registry data; between HOOS, JR and WOMAC Functioning were 0.94 (95% CI, 0.93–0.95) in the HSS data and 0.94 (95% CI, 0.93–0.96) in the FORCE registry data; and between the HOOS, JR and WOMAC Stiffness domain were 0.64 (95% CI, 0.58–0.71) in the HSS data and 0.65 (95% CI, 0.61–0.68) in the FORCE registry data.[2]

**Floor and ceiling effects:** Floor effects for the HOOS, JR were 0.6%–1.9% (percent at worst possible score preoperatively) and ceiling effects were 37%–46% (percent at best possible score postoperatively), comparable to or better than HOOS domains and the WOMAC.[2]

**KOOS, JR Validity**

**Responsiveness**: Standardized response means for the KOOS, JR, relative to other PROMs measuring postoperative knee improvement, were 1.79 (95% CI, 1.70–1.88) in the HSS data and 1.70 (95% CI, 1.54–1.86) in the FORCE registry data.

**External validity:** Correlations between the KOOS, JR and KOOS Pain domain were 0.89 (95% CI, 0.88–0.91) in the HSS data and 0.91 (95% CI, 0.90–0.93) in the FORCE registry data. Correlations between the KOOS, JR and KOOS Activity for Daily Living domain were 0.87 (95% CI, 0.85–0.88) in the HSS data and 0.84 (95% CI, 0.81–0.87) in the FORCE registry data. Correlations with the Symptoms domain were 0.59 (95% CI, 0.55–0.64) in the HSS data and 0.69 (95% CI, 0.64–0.74) in the FORCE registry data. Similarly, correlations between the KOOS, JR and WOMAC Pain were 0.80 (95% CI, 0.77–0.82) in the HSS data and 0.82 (95% CI, 0.79–0.86) in the FORCE registry data; between KOOS, JR and WOMAC Function were 0.87 (95% CI, 0.85-0.88) in the HSS data and 0.84 (95% CI, 0.81–0.87) in the FORCE registry data; and between KOOS, JR and WOMAC Stiffness were 0.72 (95% CI, 0.69–0.75) in the HSS data and 0.76 (95% CI, 0.72–0.80) in the FORCE registry data.[3]

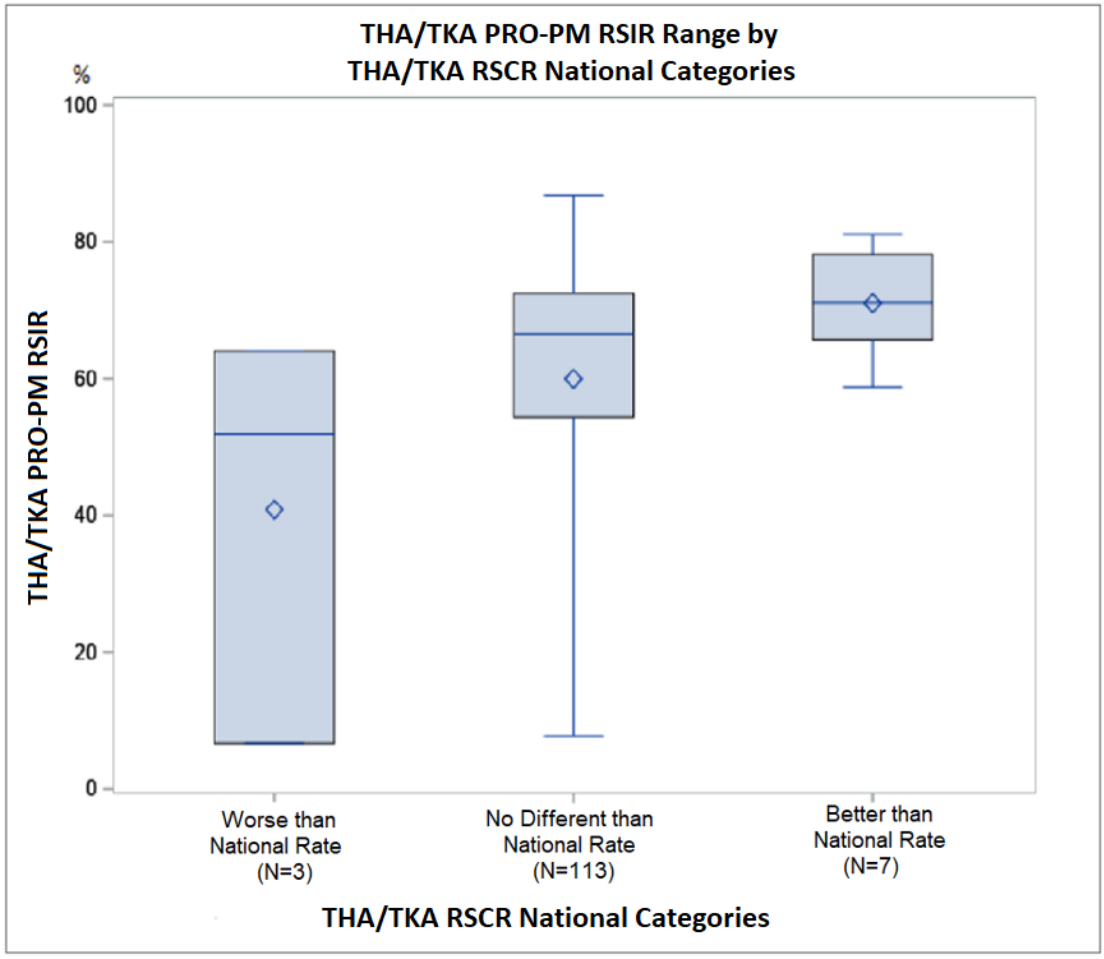
**Floor and ceiling effects:** Floor effects for the KOOS, JR were 0.4% – 1.2%, and the ceiling effects were 18.8% – 21.8%.[3]

The validity results from the literature demonstrate the HOOS, JR and the KOOS, JR PROM instruments are valid and meaningful measures for assessing patient-reported outcomes following THA/TKA procedures. The HOOS, JR and the KOOS, JR showed very high responsiveness, well beyond the 0.8 standardized response mean value considered “very large.”[117] Spearman correlation values between the HOOS, JR and the HOOS domains from which the HOOS, JR questions were drawn (Pain and Activity of Daily Living domains) were high; likewise, Spearman correlation values between the KOOS, JR and the KOOS Pain and Activity of Daily Living domains were high, and were moderate between the KOOS, JR and the KOOS Symptom domain. Floor effects were small; ceiling effects for the HOOS, JR were 37%–46%, but were comparable to or better than HOOS domains and the WOMAC (Lyman et al, 2016a; Lyman et al, 2016b).

### 3.6.2 Empirical Measure Score Validity Results

Comparison of THA/TKA PRO-PM RSIRs to RSCR categories indicates an increasing stepwise trend. Those hospitals in the RSCR “Worse than National Average” category have lower median RSIRs (51.87%) than the median RSIR (66.49%) of hospitals in the RSCR “No different than National Rate" category, which is lower than the median RSIR (71.13%) of hospitals in the RSCR “Better than National Average" category. The hospitals with lower RSCRs have higher THA/TKA RSIRs (Figure 3).

**Figure 3. Range of THA/TKA PRO-PM RSIRs by THA/TKA RSCR National Categories within Hospitals Submitting Complete PRO Data (Hospitals with ≥25 THA/TKA Procedures, N=123)**




As these outcomes are not clinically expected to be perfectly correlated but do reflect hospital-level care and processes affecting the quality of care for patients experiencing elective primary THA/TKA surgery, we interpret the increasing monotonic trend between RSIRs and RSCR national categories as reflective of empirical measure validity. Empirical validation of novel outcome measures is challenging as there is rarely, if ever, a "gold standard" against which to compare the measure. As perfect or even high correlations are not expected given the different time periods and cohorts, we sought to show the conceptual relationship between the two outcomes through actual hospital-level measure results (RSIRs) grouped by statistically significant categories of performance of RSCRs.

## 3.7 Response Rates and Response Bias Adjustment

Response rates for all hospitals and for hospitals with ≥ 25 THA/TKA patients with PRO data is presented in Table 11.

**Table 11. Mean and Distribution of Hospital Response Rates (for Complete PRO and Risk Variable Data, Combined Dataset)**

| | Response Rates (All Hospitals) | Response Rates (Hospitals with ≥25 THA/TKA Patients with PRO Data) |
|---|---|---|
| **N (Hospitals)** | 238 | 123 |
| **Mean (SD)** | 30.62% (22.79) | 43.17% (20.52) |
| **Percentile** | -- | -- |
| 100% Max | 100.00% | 90.50% |
| 99% | 84.78% | 89.66% |
| 95% | 74.29% | 79.64% |
| 90% | 61.45% | 69.66% |
| 75% (Q3) | 46.23% | 60.58% |
| 50% (Median) | 27.88% | 40.85% |
| 25% (Q1) | 9.68% | 28.34% |
| 10% | 3.70% | 17.74% |
| 5% | 2.06% | 11.49% |
| 1% | 0.72% | 5.65% |
| 0% Min | 0.24% | 5.00% |

Propensity scores calculated using a multinomial logistic regression where the outcomes 1) complete PRO submission, 2) incomplete PRO submission, and 3) no response were used to calculate stabilized IPW for each of these three groups. Table 12 provides the distribution of these stabilized IPWs applied to patients with complete PRO data submission.

**Table 12: Distribution of Stabilized IPWs Applied to Patients with Complete PRO Submission (Responders)**

| Summary Statistics | Stabilized IPWs |
|---|---|
| Mean (SD) | 1.00 (0.26) |
| Percentile | -- |
| 100% Max | 4.74 |
| 99% | 1.77 |
| 95% | 1.29 |
| 90% | 1.09 |
| 75% (Q3) | 1.01 |
| 50% (Median) | 0.95 |
| 25% (Q1) | 0.91 |
| 10% | 0.88 |
| 5% | 0.85 |
| 1% | 0.82 |
| 0% Min | 0.73 |

**Stabilized Inverse Probability Weighting**

We assessed the non-response bias using the Pearson correlation between the residuals of the hierarchical outcome model with only clinical risk factors and the probability of response. This correlation is 0.00194 (p-value=0.84). This indicates that there is not an association between the residuals and the probability of response based on our model.

We examined the correlation between the residuals of the stabilized IPW hierarchical model and the submission probability finding it to be 0.00492 (p-value=0.60) suggesting that there is not an association between the residuals weighting for non-response and probability of response.

The comparison of hospital RSIRs for the risk-adjusted model of SCB improvement with stabilized IPW and without stabilized IPW (Table 13) suggests that the results are not sensitive to our weighting adjustment. However, due to the high proportion of non-responders, we considered it important to account for the differences in characteristics of responders and non-responders found in the literature and empirically in our data. We expect that non-response bias will be a factor for the THA/TKA PRO-PM measure due to associations with non-response including SES and health status. We therefore retained response bias adjustment for the measure's results.

**Table 13. Mean and Distribution of Hospital RSIRs for Risk-Adjusted Model of SCB Improvement with and without Stabilized IPW for Potential Non-Response Bias (Combined Dataset, Hospitals with ≥25 THA/TKA Patients with PRO Data)**

| | RSIR (No Weighting) | RSIR (Weighted for Non-Response) |
|---|---|---|
| **N (Hospitals)** | 123 | 123 |
| **Mean (SD)** | 60.21% (19.57) | 60.16% (19.58) |
| **Percentile** | -- | -- |
| 100% Max | 86.66% | 86.84% |
| 99% | 85.34% | 84.73% |
| 95% | 81.69% | 81.92% |
| 90% | 78.98% | 78.85% |
| 75% (Q3) | 72.77% | 72.51% |
| 50% (Median) | 66.18% | 66.49% |
| 25% (Q1) | 54.63% | 54.36% |
| 10% | 21.70% | 20.94% |
| 5% | 13.19% | 13.42% |
| 1% | 7.79% | 7.70% |
| 0% Min | 6.89% | 6.65% |

# References

1. Centers for Medicare Medicaid Services. Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services. Final rule. *Federal register.* 2015;80(226):73273.
2. Lyman S, Lee Y-Y, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: A Short-form Hip Replacement Survey. *Clinical Orthopaedics and Related Research®.* 2016;474(6):1472-1482.
3. Lyman S, Lee Y-Y, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: A Short-form Knee Arthroplasty Outcomes Survey. *Clinical Orthopaedics and Related Research®.* 2016;474(6):1461-1471.
4. National Institutes of Health (NIH). (Patient Reported Outcomes Measurement Information Systems) PROMIS Instrument Details. http://www.nihpromis.org/measures/instrumentdetails. Accessed September 18, 2017.
5. Iqbal US, Rogers W, Selim A, et al. The Veterans Rand 12 Item Health Survey (VR-12): What It Is and How It Is Used. http://www.hosonline.org/globalassets/hos-online/publications/veterans_rand_12_item_health_survey_vr-12_2007.pdf. Accessed September 18, 2017.
6. National Quality Forum. Patient Reported Outcomes (PROs) in Performance Measurement. January 10, 2013; http://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Patient-Reported_Outcomes.aspx, 2013.
7. Agency for Healthcare Research and Quality (AHRQ). CAHPS Surveys and Tools to Advance Patient-Centered Care. 2012; http://cahps.ahrq.gov/. Accessed December 6, 2012.
8. Department of Health & Human Services. Report to Congress: National Strategy for Quality Improvement in Health Care. March 2011; http://www.ahrq.gov/workingforquality/nqs/nqs2011annlrpt.pdf. Accessed August, 2013.
9. Committee on Quality Health Care in America, Institute of Medicine (IOM). Crossing the Quality Chasm: A New Health System for the 21st Century. In: Washington, D.C.: National Academy Press; 2001: http://books.nap.edu/openbook.php?record_id=10027&page=R1. Accessed 2013.
10. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N. *Methodological Issues In The Selection, Administration And Use Of Patient-Reported Outcomes In Performance Measurement In Health Care Settings: Final Manuscript.* Prepared for the National Quality Forum Consensus Development Project on Patient-Reported Outcomes. September 28, 2012.
11. International Consortium for Health Outcomes Measurement Inc. Hip & Knee Osteoarthritis. https://www.ichom.org/portfolio/hip-knee-osteoarthritis/. Accessed March 11, 2019.
12. Alviar MJ, Olver J, Brand C, Hale T, Khan F. Do patient-reported outcome measures used in assessing outcomes in rehabilitation after hip and knee arthroplasty capture issues relevant to patients? Results of a systematic review and ICF linking process. *Journal of Rehabilitation Medicine.* 2011;43(5):374-381.
13. Alviar MJ, Olver J, Brand C, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. *Journal of Rehabilitation Medicine.* 2011;43(7):572-583.
14. Bauman S, Williams D, Petruccelli D, Elliott W, de Beer J. Physical activity after total joint replacement: a cross-sectional survey. *Clin J Sport Med.* 2007;17(2):104-108.
15. Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a "good" measure. *Clin Geriatr Med.* 2012;28(3):367-394.

16. Jones CA, Beaupre LA, Johnston DW, Suarez-Almazor ME. Total joint arthroplasties: current concepts of patient outcomes after surgery. *Rheum Dis Clin North Am.* 2007;33(1):71-86.
17. Jones CA, Pohar S. Health-related quality of life after total joint arthroplasty: a scoping review. *Clin Geriatr Med.* 2012;28(3):395-429.
18. Lau RL, Gandhi R, Mahomed S, Mahomed N. Patient satisfaction after total knee and hip arthroplasty. *Clin Geriatr Med.* 2012;28(3):349-365.
19. Montin L, Leino-Kilpi H, Suominen T, Lepisto J. A systematic review of empirical studies between 1966 and 2005 of patient outcomes of total hip arthroplasty and related factors. *J Clin Nurs.* 2008;17(1):40-45.
20. Papalia R, Del Buono A, Zampogna B, Maffulli N, Denaro V. Sport activity following joint arthroplasty: a systematic review. *Br Med Bull.* 2012;101:81-103.
21. Rolfson O, Rothwell A, Sedrakyan A, et al. Use of patient-reported outcomes in the context of different levels of data. *The Journal of bone and joint surgery American volume.* 2011;93 Suppl 3:66-71.
22. Thorborg K, Roos EM, Bartels EM, Petersen J, Holmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *BJSM online.* 2010;44(16):1186-1196.
23. Suter LG, Potteiger J, Cohen DB, Lin Z, Drye EE, Bernheim SM. Environmental Scan/Literature Review: Total Hip and Total Knee Arthroplasty Patient-Reported Outcome Measure. Report prepared for Centers for Medicare & Medicaid Services. 2012.
24. Liebs TR, Herzberg W, Ruther W, Russlies M, Hassenpflug J, Multicenter Arthroplasty Aftercare Project M. Quality-adjusted life years gained by hip and knee replacement surgery and its aftercare. *Archives of physical medicine and rehabilitation.* 2016;97(5):691-700.
25. White D, Master H. Patient Reported Measures of Physical Function in Knee Osteoarthritis. *Rheum Dis Clin North Am.* 2016;42(2):239-252.
26. Monticone M, Ferrante S, Rocca B, et al. Home-based functional exercises aimed at managing kinesiophobia contribute to improving disability and quality of life of patients undergoing total knee arthroplasty: a randomized controlled trial. *Archives of physical medicine and rehabilitation.* 2013;94(2):231-239.
27. Brown K, Topp R, Brosky JA, Lajoie AS. Prehabilitation and quality of life three months after total knee arthroplasty: a pilot study. *Perceptual and motor skills.* 2012;115(3):765-774.
28. Choong PF, Dowsey MM, Stoney JD. Does accurate anatomical alignment result in better function and quality of life? Comparing conventional and computer-assisted total knee arthroplasty. *The Journal of arthroplasty.* 2009;24(4):560-569.
29. Galea MP, Levinger P, Lythgo N, et al. A targeted home- and center-based exercise program for people after total hip replacement: a randomized clinical trial. *Archives of physical medicine and rehabilitation.* 2008;89(8):1442-1447.
30. McGregor AH, Rylands H, Owen A, Dore CJ, Hughes SP. Does preoperative hip rehabilitation advice improve recovery and patient satisfaction? *The Journal of arthroplasty.* 2004;19(4):464-468.
31. Moffet H, Collet JP, Shapiro SH, Paradis G, Marquis F, Roy L. Effectiveness of intensive rehabilitation on functional ability and quality of life after first total knee arthroplasty: A single-blind randomized controlled trial. *Archives of physical medicine and rehabilitation.* 2004;85(4):546-556.
32. Walters M, Chambers M, Sayeed Z, Anoushiravani A, El-Othmani M, Saleh K. Reducing Length of Stay in Total Joint Arthroplasty Care. *Orthopedic Clinics of North America.* 2016;47(4):653-660.

33. Kim K, Anoushiravani A, Chen K, et al. Perioperative Orthopedic Surgical Home: Optimizing Total Joint Arthroplasty Candidates and Preventing Readmission. *Journal of Arthroplasty.* 2019;34(7):S91-S96.
34. Mäkelä KT, Peltola M, Sund R, Malmivaara A, Häkkinen U, Remes V. Regional and hospital variance in performance of total hip and knee replacements: a national population-based study. *Annals of medicine.* 2011;43(sup1):S31-S38.
35. Bozic KJ, Grosso LM, Lin Z, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. *JBJS.* 2014;96(8):640-647.
36. Centers for Disease Control and Prevention (CDC). Osteoarthritis (OA). https://www.cdc.gov/arthritis/basics/osteoarthritis.htm. Accessed March 8, 2019.
37. Levit K, Stranges E, Ryan K, Elixhauser A. HCUP Facts and Figures, 2006: Statistics on Hospital-based Care in the United States. 2008. http://www.hcup-us.ahrq.gov/reports.jsp.
38. Torio CM, Moore BJ. National inpatient hospital costs: the most expensive conditions by payer, 2013. HCUP statistical brief# 204. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs Rockville, MD, Agency for Healthcare Research and Quality Available at: https://www hcup-us ahrq gov/reports/statbriefs/sb204-Most-Expensive-Hospital-Conditions jsp Accessed August.* 2016;17.
39. Guccione AA, Felson DT, Anderson JJ, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *American journal of public health.* 1994;84(3):351-358.
40. Michaud CM, McKenna MT, Begg S, et al. The burden of disease and injury in the United States 1996. *Population health metrics.* 2006;4:11.
41. Rissanen P, Aro S, Slatis P, Sintonen H, Paavolainen P. Health and quality of life before and after hip or knee arthroplasty. *The Journal of arthroplasty.* 1995;10(2):169-175.
42. Wiklund I, Romanus B. A comparison of quality of life before and after arthroplasty in patients who had arthrosis of the hip joint. *The Journal of bone and joint surgery American volume.* 1991;73(5):765-769.
43. Laupacis A, Bourne R, Rorabeck C, et al. The effect of elective total hip replacement on health-related quality of life. *The Journal of bone and joint surgery American volume.* 1993;75(11):1619-1626.
44. Ritter MA, Albohm MJ, Keating EM, Faris PM, Meding JB. Comparative outcomes of total joint arthroplasty. *The Journal of arthroplasty.* 1995;10(6):737-741.
45. Liebs T, 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):239–243.
46. Arthritis Foundation. Arthritis By the Numbers Book of Trusted Facts and Figures. In:2018: https://www.arthritis.org/Documents/Sections/About-Arthritis/arthritis-facts-stats-figures.pdf. Accessed March 8, 2019.
47. Suter LG, Parzynski CS, Grady JN, et al. 2018 Procedure-Specific Measures Updates and Specifications Report: Hospital-Level 30-Day Risk Standardized Readmission Measures - Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 7.0). March 2018; Available at: http://qualitynet.org/.
48. Miller DC, Gust C, Dimick JB, Birkmeyer N, Skinner J, Birkmeyer JD. Large variations in Medicare payments for surgery highlight savings potential from bundled payment programs. *Health affairs (Project Hope).* 2011;30(11):2107-2115.
49. National Joint Registry. *National Joint Registry for England and Wales 9th Annual Report 2012.* available at www.njrcentre.org.uk: National Joint Registry;2012.

50. 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.* 2010;33(2):76-80.
51. 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.* 2011;7(3):229-234.
52. 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.* 2013;28(5):722-727.
53. Jourdan C, Poiraudeau S, Descamps S, et al. Comparison of patient and surgeon expectations of total hip arthroplasty. *PloS one.* 2012;7(1):e30195.
54. Roos EM. Effectiveness and practice variation of rehabilitation after joint replacement. *Current opinion in rheumatology.* 2003;15(2):160-162.
55. 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.* 2012;27(5):659-666 e655.
56. 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.
57. 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 2013.
58. 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 2013; Available at: http://qualitynet.org/.
59. Rolfson O. Patient-reported Outcome Measures and Health-economic Aspects of Total Hip Arthroplasty: A study of the Swedish Hip Arthroplasty Register. 2010. https://gupea.ub.gu.se/handle/2077/23722. Accessed July 20, 2013.
60. National Health System: The Information Cenre for Health and Social Care. HESonline Hospital Episode Statistics: Proms Data. http://www.hesonline.nhs.uk/Ease/ContentServer?siteID=1937&categoryID=1295, 2012.
61. 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.* 2013;51(6):517-523.
62. Zuckerman D, Brown P, Nissen S. Medical Device Recalls and the FDA Approval Process. *Archives of Internal Medicine.* 2011;171:1006-1011.
63. Feng J, Novikov D, Anoushiravani A, Schwarzkopf R. Total knee arthroplasty: improving outcomes with a multidisciplinary approach. *J Multidiscip Healthc.* 2018;11:63-73.
64. Saufl N, Owens A, Kelly I, Merrill B, Freyaldenhouen L. A multidisciplinary approach to total joint replacement. . *Journal of Perianesthesia Nursing.* 2007;22(3):195.
65. Dorsey K, Grady J, Suter L, et al. *2016 Procedure-Specific Measure Updates and Specifications Report Hospital-Level Risk-Standardized Complication Measure Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) – Version 5.0.* Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE);2016.
66. Suter L, Parzynski C, Searfoss R, et al. *2014 Procedure-Specific Readmission Measure Updates and Specifications Report Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee*

*Arthroplasty (TKA) – Version 3.0.* Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE);2014.

67. Schneider M, Kawahara I, Ballantyne G, et al. Predictive factors influencing fast track rehabilitation following primary total hip and knee arthroplasty. *Archives of orthopaedic and trauma surgery.* 2009;129(12):1585-1591.
68. Salmon P, Hall GM, Peerbhoy D, Shenkin A, Parker C. Recovery from hip and knee arthroplasty: Patients' perspective on pain, function, quality of life, and well-being up to 6 months postoperatively. *Archives of physical medicine and rehabilitation.* 2001;82(3):360-366.
69. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. *The Journal of rheumatology.* 2000;27(7):1745-1752.
70. Singh JA, Jensen MR, Harmsen WS, Gabriel SE, Lewallen DG. Cardiac and thromboembolic complications and mortality in patients undergoing total hip and total knee arthroplasty. *Annals of the rheumatic diseases.* 2011;70(12):2082-2088.
71. Deutsch A, Smith L, Gage B, Kelleher C, Garfinkel D. *Patient-Reported Outcomes in Performance Measurement: Commissioned Paper on PRO-Based Performance Measures for Healthcare Accountable Entities.* Prepared for the National Quality Forum Consensus Development Project on Patient-Reported Outcomes. October 22, 2012.
72. Butt Z, Reeve B. Enhancing the Patient's Voice: Standards in the Design and Selection of Patient-Reported Outcomes Measures (PROMs) for Use in Patient-Centered Outcomes Research. Report submitted to the Patient Centeredness Workgroup of the PCORI Methodology Committee. March 30, 2012; http://www.pcori.org/assets/Enhancing-the-Patients-Voice-Standards-in-the-Design-and-Selection-of-Patient-Reported-Outcomes-Measures-for-Use-in-Patient-Centered-Outcomes-Research.pdf. Accessed December 5, 2012.
73. Minimum Clinically Important Difference (MCID): Defining Outcome Metrics for Orthopaedic Devices - Public Meeting, November 27-28, 2012. Paper presented at: Minimum Clinically Important Difference (MCID): Defining Outcome Metrics for Orthopaedic Devices; November 27-28, 2012; Silver Springs, MD.
74. Roos E. Knee Injury and Osteoarthritis Outcome Score. http://www.koos.nu/index.html, 2013.
75. Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation. *Technical Expert Panel (TEP) Meeting Minutes: Total Hip and Total Knee Arthroplasty Patient-Reported Outcome Measure.* (TEP Call #1 on Tuesday, July 23, 2013) 2013. Report prepared for Centers for Medicare & Medicaid.
76. Lyman S, Lee Y-Y, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the hoos and koos and jr versions after total joint replacement? *Clinical Orthopaedics and Related Research®.* 2018;476(12):2432-2441.
77. Lyman S, Lee Y-Y, McLawhorn A, Islam W, MacLean C. What are the Minimal and Substantial Improvements in the HOOS and KOOS and JR Versions After Total Joint Replacement? *Clinical orthopaedics and related research.* 2018(00):1-10.
78. Vuorenmaa M, Ylinen J, Kiviranta I, et al. Changes in pain and physical function during waiting time and 3 months after knee joint arthroplasty. *Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine.* 2008;40(7):570-575.
79. Melanie Mastanduno. BSN, MPH, Director, Population Health Measurement and Instructor at the Dartmouth Institute. In. Personal Communication, August 16, 2013.
80. The Joint Commission. E-dition Release 5.1 of the Joint Commission Comprehensive Accreditation and Certification Manual. In:July 1, 2013.

81. New Zealand Joint Registry. *The New Zealand Joint Registry Thirteen Year Report: January 1999 to December 2011.* Available at: http://njr.cdhb.health.nz October 2012.
82. Collette Pitzen. RN, BSN, CPQH, Clinical Measure Development Manager, Minnesota Community Measurement. In. Personal Communication, July 12, 2013.
83. Diane Mayberry. RN, MHA, Director Business Development, Minnesota Community Measurement. In. Personal Communication, July 12, 2013.
84. Brandes M, Ringling M, Winter C, Hillmann A, Rosenbaum D. Changes in physical activity and health-related quality of life during the first year after total knee arthroplasty. *Arthritis care & research.* 2011;63(3):328-334.
85. Ilfeld BM, Shuster JJ, Theriaque DW, et al. Long-term pain, stiffness, and functional disability after total knee arthroplasty with and without an extended ambulatory continuous femoral nerve block: a prospective, 1-year follow-up of a multicenter, randomized, triple-masked, placebo-controlled trial. *Regional anesthesia and pain medicine.* 2011;36(2):116-120.
86. Naylor JM, Harmer AR, Heard RC. Severe other joint disease and obesity independently influence recovery after joint replacement surgery: an observational study. *The Australian journal of physiotherapy.* 2008;54(1):57-64.
87. Pan WM, Li XG, Tang TS, Qian ZL, Zhang Q, Zhang CM. Mini-subvastus versus a standard approach in total knee arthroplasty: a prospective, randomized, controlled study. *The Journal of international medical research.* 2010;38(3):890-900.
88. Liebs TR, Herzberg W, Roth-Kroeger AM, Ruther W, Hassenpflug J. Women recover faster than men after standard knee arthroplasty. *Clinical orthopaedics and related research.* 2011;469(10):2855-2865.
89. Nerhus TK, Heir S, Svege I, et al. Time-dependent improvement in functional outcome following Oxford medial unicompartmental knee arthroplasty. A prospective longitudinal multicenter study involving 96 patients. *Acta orthopaedica.* 2012;83(1):46-52.
90. Papakostidou I, Dailiana ZH, Papapolychroniou T, et al. Factors affecting the quality of life after total knee arthroplasties: a prospective study. *BMC musculoskeletal disorders.* 2012;13:116.
91. Johansson HR, Bergschmidt P, Skripitz R, Finze S, Bader R, Mittelmeier W. Impact of preoperative function on early postoperative outcome after total hip arthroplasty. *Journal of orthopaedic surgery (Hong Kong).* 2010;18(1):6-10.
92. Klein DJ, Elliott MN, Haviland AM, et al. Understanding nonresponse to the 2007 Medicare CAHPS survey. *The Gerontologist.* 2011;51(6):843-855.
93. Zaslavsky AM, Zaborski LB, Cleary PD. Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. *Medical care.* 2002;40(6):485-499.
94. McHorney CA, Kosinski M, Ware JE, Jr. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Medical care.* 1994;32(6):551-567.
95. Kahn KL, Liu H, Adams JL, et al. Methodological challenges associated with patient responses to follow-up longitudinal surveys regarding quality of care. *Health services research.* 2003;38(6 Pt 1):1579-1598.
96. Fairbank JCP, Paul B,. The Oswestry disability index. *Spine.* 2000;25(22):2940-2953.
97. Ayers DC, Li W, Oatis C, Rosal MC, Franklin PD. Patient-reported outcomes after total knee replacement vary on the basis of preoperative coexisting disease in the lumbar spine and other nonoperatively treated joints: the need for a musculoskeletal comorbidity index. *The Journal of bone and joint surgery American volume.* 2013;95(20):1833.
98. PROsetta Stone. http://www.prosettastone.org/LinkingTables1/GlobalHealth/Pages/default.aspx, 2018.

99. Irgit K, Nelson CL. Defining Racial and Ethnic Disparities in THA and TKA. . *Clinical Orthopaedics and Related Research®.* 2011;469(7):1817-1823.
100. Kerman HM, Smith SR, Smith KC, et al. Disparities in Total Knee Replacement: Population Losses in Quality-Adjusted Life-Years Due to Differential Offer, Acceptance, and Complication Rates for African Americans. *Arthritis care & research.* 2018;70(9):1326-1334.
101. Hutchings A, Neuburger J, Frie K, Black N, van der Meulen J. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. *Health and Quality of Life Outcomes.* 2012;10(34).
102. de Rooij B, Ezendam N, Mols F, et al. Cancer survivors not participating in observational patient-reported outcome studies have a lower survival compared to participants: the population-based PROFILES registry. *Quality of Life Research.* 2018;27:3313-3324.
103. Schamber E, Takemoto S, Chenok K, Bozic K. Barriers to completion of patient reported outcome measures. *The Journal of arthroplasty.* 2013;28:1449-1453.
104. Dorsey K, Grady J, Desai N, et al. *2018 Condition-Specific Measures Updates and Specifications Report: Hospital-Level 30-Day Risk-Standardized Mortality Measures; Acute Myocardial Infarction – Version 12.0, Chronic Obstructive Pulmonary Disease – Version 7.0, Heart Failure – Version 12.0, Pneumonia – Version 12.0, and Stroke – Version 7.0.* Center for Outcomes Research and Evaluation;2018.
105. Dorsey K, Grady J, Desai N, et al. *2018 Condition-Specific Measures Updates and Specifications Report: Hospital-Level 30-Day Risk-Standardized Readmission Measures; Acute Myocardial Infarction – Version 11.0, Chronic Obstructive Pulmonary Disease – Version 7.0, Heart Failure – Version 11.0, Pneumonia – Version 11.0, and Stroke – Version 7.0.* Center for Outcomes Research and Evaluation;2018.
106. Garrido MM. Covariate Adjustment and Propensity Score. *JAMA.* 2016;315(14).
107. Thoemmes F, Ong AD. A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerging Adulthood.* 2015;4(1):40-59.
108. Patel J, Lee J, Zhongmin L, SooHoo N, Bozic K, Huddleston J. Predictors of low patient-reported outcomes response rates in the California Joint Replacement Registry. *The Journal of arthroplasty.* 2015;30:2071-2075.
109. Harrell FE, Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. In: *Stat Med 15* 1996:361–387.
110. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health.* 2006;60(4):290-297.
111. de Groot IB, Reijman M, Terwee CB, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis and Cartilage.* 2007;15:104-109.
112. Klässbo M, Larsson E, E. M. Hip disability and osteoarthritis outcome score: An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. . *Scandinavian Journal of Rheumatology.* 2003;23(1):46-51.
113. Nilsdotter A, A. B. Measures of hips function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity of Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. . *Arthritis care & research.* 2011;63(S11):S200-S207.
114. Ornetti P, Parratte S, Gossec L, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthritis and Cartilage.* 2010;18:522-529.

115. Adams J, Mehrota A, Thoman J, McGlynn E. Physician cost profiling – reliability and risk of misclassification. *NEJM.* 2010;362(11):1014-1021.
116. Yu H, Mehrota A, J. A. Reliability of utilization measures for primary care physician profiling. . *Healthcare.* 2013;1:22-29.
117. Steiner DL, GR N. *Health Measurement Scales: A Practical Guide to Their Development and Use.* London, UK: Oxford University Press; 2003.
118. Wechsler S. *Statistics at Square One.* 9th ed. London, UK: BMJ Publishing Group; 1996.
119. Landis J, G. K. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(159 -174).

# Appendix A: Collaboration and Stakeholder Engagement

CORE Expert Clinical Consultant:

**Kevin Bozic, MD, MBA**
Chair of the Department of Surgery and Perioperative Care and Professor of Orthopaedic Surgery at the University of Texas at Austin Dell Medical School Austin.

We would like to acknowledge Dr. Black who provided key empirical analyses of the United Kingdom's THA/TKA PRO data to assist with our measure development work, as well as Dr. Lyman and Dr. Franklin for their assistance with the HOOS, JR/KOOS, JR instruments.

**Dr. Nick Black, MD, FFPH, FRCS, DRCOG, DCH**
Professor of Health Services Research, London School of Hygiene and Tropical Medicine

**Dr. Stephen Lyman, PhD**
Director of the Healthcare Research Institute, Hospital for Special Surgery

**Dr. Patricia Franklin, MD, MBA, MPH**
Adjunct Professor Orthopedics and Physical Rehabilitation, University of Massachusetts Medical School, and Executive Director of Research, Musculoskeletal Center of Excellence at the University of Massachusetts Medical School in Worcester, MA

We would also like to thank the members of a combined TEP called by CORE and Booz Allen Hamilton under contract to CMS and ONC, respectively, and those that participated in the TEP in 2018 through 2020.

Measure Development Technical Expert Panel Members (2013 – 2014 and/or 2018 – 2020):

**Peter G. Allen, MS**
Food and Drug Administration (FDA) (Regulatory Scientist/Biomedical Engineer); Silver Spring, MD
*(TEP member 2013 – 2014, 2018 – 2019)*

**David C. Ayers, MD**
University of Massachusetts (UMass) Medical School (Professor of Orthopedics); Worcester, MA
*(TEP member 2013 – 2014, 2018 – 2020)*

**Thomas C. Barber, MD**
University of California San Francisco (Vice President of Perioperative Services and Professor of Orthopedic Surgery); San Francisco, CA
*(TEP member 2013 – 2014, 2018 – 2020)*

**Daniel J. Berry, MD**
Mayo Clinic (Chairman of Department of Orthopedic Surgery); Rochester, MN
*(TEP member 2013 – 2014)*

**Vinod Dasa, MD**
Louisiana State University Health Sciences Center (Associate Professor, Department of Orthopaedic Surgery); New Orleans, LA
*(TEP member 2013 – 2014, 2018 – 2020)*

**Rachel DuPre Brodie**
Pacific Business Group on Health (Director, Performance Information); San Francisco, CA
*(TEP member 2018 – 2020)*

**Cheryl Fahlman, PhD, MBA, BSP**
CAF Consulting Solutions (President); Gaithersburg, MD
*(TEP member 2013 – 2014, 2018 – 2020)*

**Sandra Geisinger, RN, EdD**
Knee Replacement Patient Expert; Schenectady, NY
*(TEP member 2018 – 2020)*

**Cherie Gress**
Hip Replacement Patient Expert; Vancouver, WA
*(TEP member 2018)*

**William G. Hamilton, MD**
Clinical Instructor, Anderson Orthopaedic Clinic; Chair of the Quality Measures Committee, American Association of Hip and Knee Surgeons; Alexandria, VA
*(TEP member 2018 – 2020)*

**Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN**
Association of Rehabilitation Nurses; University of Massachusetts Amherst School of Nursing (Associate Professor); Greenfield, MA
*(TEP member 2013 – 2014, 2018 – 2020)*

**Benita Lattimore**
Patient Expert; Chicago, IL
*(TEP member 2019 – 2020)*

**Courtland G. Lewis, MD**
Hartford Hospital (Director of Orthopedic Surgery); Farmington, CT
*(TEP member 2013 – 2014)*

**Arthur L. Malkani, MD**
University of Louisville (Chief of Adult Reconstruction and Clinical Professor, Department of Orthopedic Surgery); Louisville, KY
*(TEP member 2018 – 2020)*

***Patient***
Recent recipient of THA or TKA
*(TEP member 2013 – 2014)*

**Michael H. Perskin, MD**
The American Geriatrics Society; New York University School of Medicine (Associate Chair of Clinical Affairs and Assistant Professor in the Department of Medicine); New York, NY
*(TEP member 2013 – 2014, 2018 – 2020)*

**Nan Rothrock, PhD**
Northwestern University (Research Associate Professor); Chicago, IL
*(TEP member 2018 – 2020)*

**Jonathan L. Schaffer, MD, MBA**
The Cleveland Clinic Foundation (Managing Director, eCleveland Clinic Information Technology Division); Cleveland, OH
*(TEP member 2013 – 2014, 2018 – 2020)*

**John H. Seiverd, PT, DPT, CCCE**
James A. Haley Veterans' Hospital (Physical Therapy Center Coordinator of Clinical Education, Orthopaedic and Neurologic PT Residency Program Director); Tampa, FL
*(TEP member 2013 – 2014)*

**Lyle S. Sorensen, MD**
Virginia Mason Medical Center (Chief of Orthopedics and Sports Medicine); Seattle, WA
*(TEP member 2013 – 2014)*

**A. Christopher Strenta, PhD**
Dartmouth College (Associate Dean, Finance and Operations); Hanover, NH; recent recipient of THA or TKA
*(TEP member 2013 – 2014)*

**Margaret A. VanAmringe, MHS**
The Joint Commission (Vice President, Public Policy and Government Relations); Washington, DC
*(TEP member 2013 – 2014, 2018 – 2020)*

**Adolph J. Yates, Jr., MD**
University of Pittsburgh Medical Center (Orthopaedic Surgeon/Associate Professor); Pittsburgh, PA
*(TEP member 2018 – 2020)*

Technical Advisory Group Members (2018-2019):

**Sheila Barnett, MD**
Beth Israel Deaconess Medical Center (Associate Professor of Anesthesiology); Boston, MA

**Rachel Clark Sisodia, MD**
Partners Healthcare (Medical Director for Patient Reported Outcomes); Boston, MA

**Brian Curtin, MS, MD**
OrthoCarolina Hip and Knee Center (Orthopedic Surgeon); Charlotte, NC

**Hiren Dave, MBA, MSc, MCP, LSSMBB, PMP, MSF**
Envision Healthcare (Director, Quality Payment Programs); Nashville, TN

**Paula R. Farrell, BSN, RN, CPHQ**
The Joint Commission (Associate Project Director); Oakbrook Terrace, IL

**Andrew Gordon, MD, PhD**
Metro Orthopedics and Sports Therapy (Director of Sports Concussions and Non-Operative Spine); Silver Spring, MD

**Inder Johnson, MBA, OTR/L**
Adventist Health and Rideout Hospital (Director of Rehabilitation Services and Occupational Therapist); Roseville, CA

**Jennifer Lennon, OTR/L**
University of Pittsburgh Medical Center Presbyterian, Montefiore, and Western Psychiatric Institute Clinic (Director of Rehabilitation Services); Pittsburgh, PA

**Daniel A. Mendelson, MD, MS, FACP, CMS, AGSF**
University of Rochester Medical Center (Konar Professor of Geriatrics & Palliative Medicine); Rochester, NY

**Ryan Nunley, MD**
Washington University (Associate Professor of Orthopaedic Surgery); St. Louis, MO

**Sara R. Piva, PT, PhD**
University of Pittsburgh (Associate Professor of Physical Therapy); Pittsburgh, PA

**Daniel L. Riddle, PT, PhD, FAPTA**
Virginia Commonwealth University (Otto. D. Payton Professor of Physical Therapy); Richmond, VA

**Brandy N. Wilkins, DPT**
The Joint Commission (Program Coordinator); Oakbrook Terrace, IL

Patient Working Group (2018 – 2020):

The Patient Working Group was assembled with the support of the National Partnership for Women & Families, with the goal of including patient perspectives in decision making during key measure development decisions. The Patient Working Group consists of seven patients with experience undergoing THA and TKA procedures. CORE recognizes the value of partnering with patients to discuss the development of this measure.

Orthopedic Technical Expert Panel (2020 – 2021)
* indicates individual was also a member of Measure Development Technical Expert Panel, listed above

**David C. Ayers, MD***
University of Massachusetts (UMass) Medical School (Professor of Orthopedics); Worcester, MA

**Thomas C. Barber, MD***
University of California San Francisco (Vice President of Perioperative Services and Professor of Orthopedic Surgery); San Francisco, CA

**Phyllis Bass**
Patient Expert; Cypress, TX

**Vinod Dasa, MD***
Louisiana State University Health Sciences Center (Associate Professor, Department of Orthopaedic Surgery); New Orleans, LA

**Rachel DuPre Brodie***
Pacific Business Group on Health (Director, Performance Information); San Francisco, CA

**Cheryl Fahlman, PhD, MBA, BSP***
CAF Consulting Solutions (President); Gaithersburg, MD

**William G. Hamilton, MD***
Anderson Orthopaedic Clinic (Clinical Instructor); American Association of Hip and Knee Surgeons (Chair of the Quality Measures Committee); Alexandria, VA

**Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN***
Association of Rehabilitation Nurses; University of Massachusetts Amherst School of Nursing (Associate Professor); Greenfield, MA

**Craig T. Miller, PT**
Home Care Therapy (Director); Rivetus Rehabilitation (Senior PT); Macomb, MI

**Michael H. Perskin, MD***
The American Geriatrics Society; New York University School of Medicine (Associate Chair of Clinical Affairs and Assistant Professor in the Department of Medicine); New York, NY

**Benita Lattimore***
Patient Expert; Chicago, IL

**Patricia Walker**
Patient Expert; South Holland, IL

**Nan Rothrock, PhD***
Northwestern University (Research Associate Professor); Chicago, IL

**Jonathan L. Schaffer, MD, MBA***
The Cleveland Clinic Foundation (Managing Director, eCleveland Clinic Information Technology Division); Cleveland, OH

**Adam Schwartz, MD, MBA**
Mayo Clinic (Consultant of the Department of Orthopaedic Surgery, Associate Professor of Orthopaedic Surgery) American Academy of Orthopaedic Surgeons; Phoenix, AZ

**Robert Sterling, MD**
Association of Hip and Knee Surgeons, Johns Hopkins University (Orthopaedic Surgeon, Associate Professor of Orthopaedic Surgery); Baltimore, MD

**Christine Von Raesfeld**
Patient Expert; Santa Clara, CA

**Kevin Woodward, PA-C, MMS**
American Academy of Physician Assistants, John Hopkins University (Physician Assistant of Orthopaedic Surgery); Baltimore, MD

**Margaret A. VanAmringe, MHS***
The Joint Commission (Vice President, Public Policy and Government Relations); Washington, DC

**Adolph J. Yates, Jr., MD***
University of Pittsburgh Medical Center (Orthopaedic Surgeon/Associate Professor); Pittsburgh, PA

<u>Orthopedic Clinical Working Group Members (2020 – 2021):</u>

**James I. Huddleston, III, MD**
Stanford University Medical Center (Associate Professor and Chief of Arthritis Service, Department of Orthopaedic Surgery), Chair of the Advocacy Council for the American Association of Hip and Knee Surgeons; Stanford, CA

**Jay R. Lieberman, MD, FAOOS**
Keck School of Medicine at the University of Southern California (Professor and Chair, Department of Orthopaedic Surgery, Director of Institute of Orthopaedics), Viterbi School of Engineering at the University of Southern California (Professor of Biomedical Engineering), Second Vice President, The Hip Society; Los Angeles, CA

**Mary I. O'Connor, MD**
Yale School of Medicine (Professor of Orthopaedics and Rehabilitation), The Knee Society; New Haven, CT

**Kathryn (Kristy) Schabel, MD, FAAOS**
Oregon Health and Science University (Associate Professor of Orthopaedic Surgery, Adult Reconstruction), American Academy of Orthopaedic Surgeons; Portland, OR

# Appendix B: Medical Record Review Results

In 2014, we performed a medical record review at seven orthopedic practices across the country to evaluate the feasibility, uniformity, and reliability of clinical data elements prioritized by orthopedists for use in risk adjustment for PRO-PMs for elective primary THA/TKA. This review consisted of 30 standardized medical record abstractions. For each record, the abstractor collected data on the presence or absence of risk variables in the preoperative record. The abstraction included 27 to 30 individual questions regarding 11 to 12 specific risk variables, depending upon whether the patient was undergoing TKA or THA, respectively. Table B1 presents findings of this review with regards to documentation of risk variable information in patient medical records and in electronic health records.

**Table B1. Percentage of Patients for Whom Recommended Candidate Clinical Risk Variables Were Documented within the Medical Record and EHR, Respectively**

| Question | Patients with "Yes" Response (%) | Patients with "Yes, in EHR Standardized Field" Response (%) |
|---|---|---|
| Did this patient receive an elective primary THA or TKA or both? | N/A[a] | N/A[a] |
| Is live-in home support (spouse or domestic partner or live-in family) documented in preoperative surgical notes (or elsewhere in medical record)? | 67.6 | 49.5 |
| Is marital status documented in preoperative surgical notes (or elsewhere in medical record)? | 88.6 | 72.9 |
| Is BMI documented in preoperative surgical notes (or elsewhere in medical record)? | 98.1 | 93.8 |
| Are height and weight documented in preoperative surgical notes (or elsewhere in medical record) such that BMI can be calculated? | 98.6 | 93.8 |
| Is current tobacco use documented in preoperative surgical notes (or elsewhere in medical record)? | 93.3 | 81.0 |
| Is past tobacco use documented in preoperative surgical notes (or elsewhere in medical record)? | 87.6 | 80.5 |
| Is smoking status in pack years documented in preoperative surgical notes (or elsewhere in medical record)? | 28.6 | 24.8 |
| Is use of chronic (any given period of 90 days) preoperative narcotics documented in preoperative surgical notes? | 33.8 | 26.7 |
| If yes, is dose greater than 24 mg hydromorphone per 24 hours or equivalent for a duration of four or more weeks documented in preoperative surgical notes? | 11.3 | 8.9 |
| Is Charnley classification documented in preoperative surgical notes? | 21.9 | 20.5 |

| Question | Patients with "Yes" Response (%) | Patients with "Yes, in EHR Standardized Field" Response (%) |
|---|---|---|
| Is a total painful joint count documented in preoperative surgical notes? | 28.6 | 19.0 |
| Is homunculus with painful joints documented in preoperative surgical notes? | 0.5 | 0 |
| Is quantified spinal pain using validated instrument (e.g., visual analog scale, Oswestry or other formal assessment) documented in preoperative surgical notes? | 1.0 | 1.0 |
| Is subjective or semi-quantified spinal pain (chronic low back pain, spinal stenosis, sciatic pain) documented in preoperative surgical notes? | 31.7 | 20.2 |
| Is any other subjective estimate of painful joints (not asked above) documented in preoperative surgical notes? | 42.6 | 30.8 |
| Is quantified hip/knee range of motion (i.e., in degrees) documented in preoperative surgical notes? | 84.8 | 51.0 |
| Is subjective hip/knee range of motion (e.g., "limited") documented in preoperative surgical notes? | 54.1 | 34.4 |
| Is the use of a specific gait aide(s) documented in preoperative surgical notes? | 68.1 | 45.2 |
| Is walking distance capacity documented in preoperative surgical notes? | 50.0 | 26.7 |
| Is the presence of retained hardware requiring removal during the THA or TKA procedure documented in preoperative surgical notes? | 1.4 | 0.5 |
| Are dynamometric results for hip abduction strength (abductor muscle deficiency) documented in preoperative surgical notes? | 17.9 | 16.8 |
| Is a semi-quantified strength assessment (i.e., on 1–5 scale) documented in preoperative surgical notes? | 56.3 | 46.9 |
| Is the presence of a Trendelenberg gait documented in preoperative surgical notes? | 44.8 | 42.7 |
| Is a subjective strength assessment (e.g., "weak abductors") documented in preoperative surgical notes? | 35.1 | 29.9 |
| Is a history of developmental dysplasia of hip (DDH) with high displacement or dislocation documented in preoperative surgical notes? | 19.1 | 19.1 |
| Is a history of other childhood developmental abnormalities documented in preoperative surgical notes? | 20.0 | 18.9 |
| Are quantified angular, translational, or rotational deformities of the proximal femur (in degrees) documented in preoperative surgical notes? | 4.21 | 2.1 |

| Question | Patients with "Yes" Response (%) | Patients with "Yes, in EHR Standardized Field" Response (%) |
|---|---|---|
| Is subjective angular femoral deformity (e.g., coxa varus/valgus without degree) documented in preoperative surgical notes? | 21.1 | 14.7 |
| Is other subjective deformity of femoral neck or shaft documented in preoperative surgical notes? | 26.3 | 11.6 |
| Is a quantified anatomic angle (femoro-tibial angle in degrees with vaurs/valgus) documented in preoperative surgical notes? | 43.5 | 21.7 |
| Is varus/valgus deformity greater than or equal to 15 degrees documented in preoperative surgical; notes? | 26.7 | 24.1 |
| Is subjective angular thigh-foot deformity (e.g., varus/valgus without degree) documented in preoperative surgical notes? | 48.7 | 29.1 |
| Are dynamometric results for extensor mechanism strength documented in preoperative surgical notes? | 6.0 | 6.0 |
| Is a semi-quantified strength assessment (i.e., on 1–5 scale) documented in preoperative surgical notes? | 29.9 | 27.4 |
| Is a subjective strength assessment (e.g., "weak abductors" or "extensor lag") documented in preoperative surgical notes? | 45.3 | 45.3[b] |

[a] Total number of THA and TKA was 95 and 115 records, respectively.
[b] "Yes, in EHR standardized field" was not a response option for this question. Percentage represents all "Yes" responses.

# Appendix C: Data Collection Specifications

Per the 2015 CJR Final Rule, data would be consistently collected at the hospital level, and would be used for development of a THA/TKA PRO-PM. Complete data specifications for the development data collection as part of CJR are listed here.

Please note that clinical expert input subsequent to THA/TKA PRO-PM development and testing with PRO data collected from CJR participant hospitals included strong recommendation to extend the postoperative data assessment window established for CJR PRO data collection to better align with clinical workflow and typical one-year follow-up scheduling and to allow for better postoperative PRO data capture.

## C.1 Preoperative Assessments (To Be Collected between 90 and Zero Days Prior to THA/TKA Procedure)

- PROMIS-Global (all items) or VR-12 (all items)
- For THA Patients, HOOS or HOOS, JR
- For TKA patients, KOOS or KOOS, JR
- Medicare Provider Number
- Medicare Health Insurance Claim (HIC) Number
- Date of Birth
- Date of Collection
- Mode of Collection
- Person Completing the Survey
- Race
- Ethnicity
- Single-Item Health Literacy Screening (SILS2) Questionnaire
- BMI or Weight (kg)/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)

## C.2 Postoperative Assessments (To Be Collected between 270 and 365 Days Following THA/TKA Procedure in CJR; we propose data be collected between 300 and 425 Days Following THA/TKA Procedure for future measure use)

- PROMIS-Global (all items) or VR-12 (all items)
- For THA Patients, HOOS or HOOS, JR
- For TKA Patients, KOOS or KOOS, JR
- Medicare Provider Number
- HIC Number
- Date of Birth
- Date of Collection
- Mode of Collection
- Person Completing the Survey
- Date of Admission to Anchor Hospitalization
- Date of Eligible Procedure

**Table C1. Data Elements That Were Finalized in the CJR Final Rule or Requested in the CJR Model for Prospective Data Collection and Testing Project**

| Data Element | Format | Range | Collection | Source | Required (Core) Data Elements |
|---|---|---|---|---|---|
| VR-12 (all items) | See Table C2 | See Table C2 | Pre- and postoperative | Patient reported | X |
| PROMIS-Global (all items) | See Table C3 | See Table C3 | Pre- and postoperative | Patient reported | X |
| HOOS (all items, or six HOOS, JR items) | See Table C4 | See Table C4 | Pre- and postoperative | Patient reported | X |
| KOOS (all items, or seven KOOS, JR items) | See Table C5 | See Table C5 | Pre- and postoperative | Patient reported | X |
| Medicare Provider Number | Six-digit Medicare provider number, also known as CCN | N/A | Pre- and postoperative | Provider reported | N/A |
| Medicare Health Insurance Claim (HIC) Number | Ten or 11-digit account number (e.g., 123456789A) | N/A | Pre- and postoperative | Provider reported | X |
| Date of Birth | Date (MM/DD/YYYY) | N/A | Pre- and postoperative | Provider reported | X |
| Race | 0 = White<br>1 = Black or African American<br>2 = Asian<br>3 = American Indian/Alaska Native<br>4 = Native Hawaiian/Other Pacific Islander | 0–4 | Preoperative | Patient or provider reported | X |
| Ethnicity | 0 = non-Hispanic or not Latino<br>1 = Hispanic or Latino | 0,1 | Preoperative | Patient or provider reported | X |

| Data Element | Format | Range | Collection | Source | Required (Core) Data Elements |
|---|---|---|---|---|---|
| Date of Collection | Date (MM/DD/YYYY) | N/A | Pre- and postoperative | Provider reported | N/A |
| Mode of Collection | 0 = Paper<br>1 = Telephone (active interactive voice response)<br>2 = Electronic (web-based, EHR, etc.) | 0–2 | Pre- and postoperative | Provider reported | N/A |
| Person Completing the Survey | 0 = Self<br>1 = Surrogate | 0,1 | Pre- and postoperative | Patient or provider reported | N/A |
| SILS2 questionnaire ("How comfortable are you filling out medical forms by yourself?") | 0 = Not at all<br>1 = A little bit<br>2 = Somewhat<br>3 = Quite a bit<br>4 = Extremely | 0–4 | Preoperative | Patient reported | X |
| Body Mass Index (BMI) | Weight (kg)/Height (cm) | 10–70 | Preoperative | Medical record/EHR | X[a] |
| Height | Centimeters (cm) | Positive number with one decimal digit | Preoperative | Medical record/EHR | X[b] |
| Weight | Kilograms (kg) | Positive number with one decimal digit | Preoperative | Medical record/EHR | X[b] |
| Use of Chronic (≥ 90 days) Narcotics | 0 = No<br>1 = Yes | 0,1 | Preoperative | Medical record/EHR (provider reported) | X |
| Total Painful Joint Count: Patient-Reported Pain in Non-Operative Lower Extremity Joint ("What amount of pain have you experienced in the last week in your other knee/hip?") | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Preoperative | Patient reported | X |

| Data Element | Format | Range | Collection | Source | Required (Core) Data Elements |
|---|---|---|---|---|---|
| Quantified Spinal Pain: Patient-Reported Back Pain, Oswestry Index Question ("My BACK PAIN at the moment is" | 0 = None<br>1 = Very mild<br>2 = Moderate<br>3 = Fairly severe<br>4 = Very severe<br>5 = Worst imaginable | 0–5 | Preoperative | Patient reported | X |
| Date of Admission to Anchor Hospitalization | Date (MM/DD/YYYY) | N/A | Postoperative | Medical record/EHR | X |
| Date of Eligible Procedure | Date (MM/DD/YYYY) | N/A | Postoperative | Medical record/EHR | X |

Note: Core Data Element refers to those elements required for successful submission per the CJR Final Rule; non-Core Elements are requested to enhance measure development data quality

[a] Conditionally yes; collection of Height and Weight together will substitute the requirement to collect BMI

[b] Conditionally yes; collection of BMI will substitute the requirement to collect Height and Weight

**Table C2. Data Elements from the VR-12 Health Survey**

| Item | Format | Range | Collection |
|---|---|---|---|
| General health | 1 = Excellent<br>2 = Very good<br>3 = Good<br>4 = Fair<br>5 = Poor | 1–5 | Pre- and postoperative |
| Does your health limit you in moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf? | 1 = Yes, limited a lot<br>2 = Yes, limited a little<br>3 = No, not limited at all | 1–3 | Pre- and postoperative |
| Does your health limit you in climbing several flights of stairs? | 1 = Yes, limited a lot<br>2 = Yes, limited a little<br>3 = No, not limited at all | 1–3 | Pre- and Postoperative |
| During the past four weeks, have you accomplished less in work or other daily activities than you would like because of your physical health? | 1 = No, none of the time<br>2 = Yes, a little of the time<br>3 = Yes, some of the time<br>4 = Yes, most of the time<br>5 = Yes, all of the time | 1–5 | Pre- and postoperative |
| During the past four weeks, were you limited in the kind of work or other daily activities because of your physical health? | 1 = No, none of the time<br>2 = Yes, a little of the time<br>3 = Yes, some of the time<br>4 = Yes, most of the time<br>5 = Yes, all of the time | 1–5 | Pre- and postoperative |
| During the past four weeks, have you accomplished less in work or other daily activities than you would like as a result of any emotional problems (such as feeling depressed or anxious)? | 1 = No, none of the time<br>2 = Yes, a little of the time<br>3 = Yes, some of the time<br>4 = Yes, most of the time<br>5 = Yes, all of the time | 1–5 | Pre- and postoperative |
| During the past four weeks, did you not do work or other activities as carefully as usual as a result of any emotional problems (such as feeling depressed or anxious)? | 1 = No, none of the time<br>2 = Yes, a little of the time<br>3 = Yes, some of the time<br>4 = Yes, most of the time<br>5 = Yes, all of the time | 1–5 | Pre- and postoperative |
| During the past four weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? | 1 = Not at all<br>2 = A little bit<br>3 = Moderately<br>4 = Quite a bit<br>5 = Extremely | 1–5 | Pre- and postoperative |
| How much of the time during the past four weeks have you felt calm and peaceful? | 1 = All of the time<br>2 = Most of the time<br>3 = A good bit of the time<br>4 = Some of the time<br>5 = A little of the time<br>6 = None of the time | 1–6 | Pre- and postoperative |

| Item | Format | Range | Collection |
|---|---|---|---|
| How much of the time during the past four weeks have you had a lot of energy? | 1 = All of the time<br>2 = Most of the time<br>3 = A good bit of the time<br>4 = Some of the time<br>5 = A little of the time<br>6 = None of the time | 1–6 | Pre- and postoperative |
| How much of the time during the past four weeks have you felt downhearted and blue? | 1 = All of the time<br>2 = Most of the time<br>3 = A good bit of the time<br>4 = Some of the time<br>5 = A little of the time<br>6 = None of the time | 1–6 | Pre- and postoperative |
| During the past four weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)? | 1 = All of the time<br>2 = Most of the time<br>3 = Some of the time<br>4 = A little of the time<br>5 = None of the time | 1–5 | Pre- and postoperative |
| General physical health now compared to one year ago | 1 = Much better<br>2 = Slightly better<br>3 = About the same<br>4 = Slightly worse<br>5 = Much worse | 1–5 | Pre- and postoperative |
| Emotional problems (such as feeling anxious, depressed, or irritable) now compared to one year ago | 1 = Much better<br>2 = Slightly better<br>3 = About the same<br>4 = Slightly worse<br>5 = Much worse | 1–5 | Pre- and postoperative |

**Table C3. Data Elements from PROMIS-Global (Format Reflects PROMIS-Global Version 1.2)**

| Item | Format | Range | Collection |
|---|---|---|---|
| General health | 1 = Poor<br>2 = Fair<br>3 = Good<br>4 = Very good<br>5 = Excellent | 1–5 | Pre- and postoperative |
| General quality of life | 1 = Poor<br>2 = Fair<br>3 = Good<br>4 = Very good<br>5 = Excellent | 1–5 | Pre- and postoperative |
| General physical health | 1 = Poor<br>2 = Fair<br>3 = Good<br>4 = Very good<br>5 = Excellent | 1–5 | Pre- and postoperative |

| Item | Format | Range | Collection |
|---|---|---|---|
| General mental health | 1 = Poor<br>2 = Fair<br>3 = Good<br>4 = Very good<br>5 = Excellent | 1–5 | Pre- and postoperative |
| General satisfaction with social activities and relationships | 1 = Poor<br>2 = Fair<br>3 = Good<br>4 = Very good<br>5 = Excellent | 1–5 | Pre- and postoperative |
| General ability to carry out social activities and roles | 1 = Poor<br>2 = Fair<br>3 = Good<br>4 = Very good<br>5 = Excellent | 1–5 | Pre- and postoperative |
| Ability to carry out everyday physical activities | 1 = Not at all<br>2 = A little<br>3 = Moderately<br>4 = Mostly<br>5 = Completely | 1–5 | Pre- and postoperative |
| Emotional problems in past seven days | 1 = Always<br>2 = Often<br>3 = Sometimes<br>4 = Rarely<br>5 = Never | 1–5 | Pre- and postoperative |
| Average fatigue in past seven days | 1 = Very severe<br>2 = Severe<br>3 = Moderate<br>4 = Mild<br>5 = None | 1–5 | Pre- and postoperative |
| Average pain in past seven days | Ten-point scale (no pain to worst imaginable pain) | 0–10 | Pre- and postoperative |

**Table C4. Data Elements from HOOS Survey**

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Symptoms: Feel grinding or hear clicking or any other type of noise from hip during the last week | 0 = Never<br>1 = Rarely<br>2 = Sometimes<br>3 = Often<br>4 = Always | 0–4 | Pre- and Postoperative | N/A | N/A |
| Symptoms: Difficulties spreading legs wide apart during the last week | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Symptoms: Difficulties striding out when walking during the last week | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Stiffness: Severity of hip joint stiffness after first wakening in the morning during the last week | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Stiffness: Severity of hip stiffness after sitting, lying, or resting later in the day during the last week | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Pain: How often is your hip painful? | 0 = Never<br>1 = Monthly<br>2 = Weekly<br>3 = Daily<br>4 = Always | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week straightening hip fully | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Pain: Amount of hip pain last week bending hip fully | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week following walking on a flat surface | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week going up or down stairs | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Pain: Amount of hip pain last week at night while in bed | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week sitting or lying | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week standing upright | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week walking on a hard surface (asphalt, concrete, etc.) | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of hip pain last week walking on an uneven surface | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Function (Daily Living): Degree of difficulty last week due to your hip when descending stairs | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when ascending stairs | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when rising from sitting | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Function (Daily Living): Degree of difficulty last week due to your hip when standing | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when bending to the floor/picking up an object | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Function (Daily Living): Degree of difficulty last week due to your hip when walking on a flat surface | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when getting in/out of car | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when going shopping | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Function (Daily Living): Degree of difficulty last week due to your hip when putting on socks/stockings | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when rising from bed | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when taking off socks/stockings | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when lying in bed (turning over, maintaining hip position) | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Function (Daily Living): Degree of difficulty last week due to your hip when getting in/out of bath | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip when sitting | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Function (Daily Living): Degree of difficulty last week due to your hip when getting on/off toilet | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Function (Daily Living): Degree of difficulty last week due to your hip during heavy domestic duties (moving heavy boxes, scrubbing floors, etc.) | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty last week due to your hip during light domestic duties (cooking, dusting, etc.) | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty last week due to your hip during squatting | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty last week due to your hip during running | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty last week due to your hip during twisting/pivoting on loaded leg | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty last week due to your hip during walking on an uneven surface | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: How often are you aware of your hip problem? | 0 = Never<br>1 = Monthly<br>2 = Weekly<br>3 = Daily<br>4 = Constantly | 0–4 | Pre- and postoperative | N/A | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Quality of Life: Have you modified your lifestyle to avoid activities potentially damaging to your hip? | 0 = Not at all<br>1 = Mildly<br>2 = Moderately<br>3 = Severely<br>4 = Totally | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: How much are you troubled with lack of confidence in your hip? | 0 = Not at all<br>1 = Mildly<br>2 = Moderately<br>3 = Severely<br>4 = Extremely | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: General difficulty with hip | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |

**Table C5. Data Elements from the KOOS Survey**

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Symptoms: Swelling in knee during the last week | 0 = Never<br>1 = Rarely<br>2 = Sometimes<br>3 = Often<br>4 = Always | 0–4 | Pre- and postoperative | N/A | N/A |
| Symptoms: Feel grinding or hear clicking or any other type of noise when knee moves during the last week | 0 = Never<br>1 = Rarely<br>2 = Sometimes<br>3 = Often<br>4 = Always | 0–4 | Pre- and postoperative | N/A | N/A |
| Symptoms: Knee catches or gets hung up when moving during the last week | 0 = Never<br>1 = Rarely<br>2 = Sometimes<br>3 = Often<br>4 = Always | 0–4 | Pre- and postoperative | N/A | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Symptoms: Ability to straighten knee fully during the last week | 0 = Always<br>1 = Often<br>2 = Sometimes<br>3 = Rarely<br>4 = Never | 0–4 | Pre- and postoperative | N/A | N/A |
| Symptoms: Ability to bend knee fully during the last week | 0 = Always<br>1 = Often<br>2 = Sometimes<br>3 = Rarely<br>4 = Never | 0–4 | Pre- and postoperative | N/A | N/A |
| Stiffness: Severity of knee joint stiffness after first wakening in the morning during the last week | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Stiffness: Severity of knee stiffness after sitting, lying, or resting later in the day during the last week | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Frequency of knee pain | 0 = Never<br>1 = Monthly<br>2 = Weekly<br>3 = Daily<br>4 = Always | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of knee pain last week when twisting/pivoting on knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Pain: Amount of knee pain last week when straightening knee fully | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Pain: Amount of knee pain last week when bending knee fully | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Pain: Amount of knee pain last week when walking on flat surface | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of knee pain last week when going up or down stairs | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Pain: Amount of knee pain last week at night while in bed | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of knee pain last week when sitting or lying | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Pain: Amount of knee pain last week when standing upright | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Function (Daily Living): Degree of difficulty descending stairs in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty ascending stairs in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty rising from sitting in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Function (Daily Living): Degree of difficulty standing in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty bending to floor/picking up an object in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | X |
| Function (Daily Living): Degree of difficulty walking on flat surface in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty getting in/out of car in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty going shopping in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty putting on socks/stockings in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty rising from bed in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty taking off socks/stockings in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Function (Daily Living): Degree of difficulty lying in bed (turning over, maintaining knee position) in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty getting in/out of bath in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty sitting in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty getting on/off toilet in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty with heavy domestic duties (moving heavy boxes, scrubbing floors, etc.) in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Daily Living): Degree of difficulty with light domestic duties (cooking, dusting, etc.) in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | X | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty squatting in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty running in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |

| Data Element | Format | Range | Collection | Data Element from Subscale as Finalized in CJR Final Rule | Data Element in JR Survey as Finalized in CJR Final Rule |
|---|---|---|---|---|---|
| Function (Sports and Recreational Activities): Degree of difficulty jumping in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty twisting/pivoting on your injured knee in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Function (Sports and Recreational Activities): Degree of difficulty kneeling in last week due to knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: How often are you aware of your knee problem? | 0 = Never<br>1 = Monthly<br>2 = Weekly<br>3 = Daily<br>4 = Constantly | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: Have you modified your lifestyle to avoid potentially damaging activities to your knee? | 0 = Not at all<br>1 = Mildly<br>2 = Moderately<br>3 = Severely<br>4 = Totally | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: How much are you troubled with lack of confidence in your knee? | 0 = Not at all<br>1 = Mildly<br>2 = Moderately<br>3 = Severely<br>4 = Extremely | 0–4 | Pre- and postoperative | N/A | N/A |
| Quality of Life: General difficulty with knee | 0 = None<br>1 = Mild<br>2 = Moderate<br>3 = Severe<br>4 = Extreme | 0–4 | Pre- and postoperative | N/A | N/A |

# Appendix D: TEP Risk Variable Survey Results

In 2014, we surveyed the TEP regarding the relative priority of various risk variables. We asked the TEP to consider both clinical importance and feasibility in their overall prioritization. The rating options included an overall priority score of High, Medium, or Low. Nine of 15 TEP members completed the survey.

**Table D1. Summary Results from TEP Survey on Risk Variable Priority**

| Risk Factor | Score Summary (Number Voted for High, Medium, Low) |
|---|---|
| Age | High (8); Medium (1) |
| Baseline expectations about joint replacement | Medium (1); Low (8) |
| Gender | High (8); Low (1) |
| Living circumstances | High (1); Medium (3); Low (5) |
| Motivation score | High (1); Medium (1); Low (6) |
| Social functioning | High (6); Medium (2) |
| Social support | High (3); Medium (2); Low (4) |
| Functional Independence Measure (FIM) | High (1); Medium (1); Low (7) |
| Harris Hip Score | Medium (2); Low (7) |
| Health Status Questionnaire (HSQ) | Medium (1); Low (8) |
| Preoperative PRO score | High (7); Low (2) |
| Walking distance | Medium (4); Low (5) |
| Number of comorbidities | High (2); Medium (1); Low (2) |
| American Society of Anesthesiologists (ASA) physical status classification system | High (3); Medium (1); Low (1) |
| Bodily pain | High (2); Medium (1); Low (2) |
| BMI | High (3); Medium (2) |
| Cancer | High (2); Medium (2); Low (1) |

| Risk Factor | Score Summary (Number Voted for High, Medium, Low) |
|---|---|
| Cardiac disease | High (4); Low (1) |
| Chronic pain management | High (2); Medium (3) |
| Diabetes | High (3); Low (2) |
| Duration of symptoms | Medium (1); Low (4) |
| History of Deep Venous Thrombosis/Pulmonary Embolism (DVT/PE) | High (4); Medium (1) |
| Immunocompromised/HIV | High (3); Medium (2) |
| Inflammatory arthritis | High (4); Low (1) |
| Nutritional status | High (4); Low (1) |
| Peripheral vascular disease | High (2); Medium (2); Low (1) |
| Other joint or musculoskeletal problems | High (1); Medium (3); Low (1) |
| Poor circulation | High (3); Medium (1); Low (1) |
| Pain when walking | High (1); Medium (1); Low (3) |
| Pain visual analogue scale | High (1); Medium (2); Low (2) |
| Smoking status | High (3); Medium (1); Low (1) |
| Depression | High (8) |
| Mini Mental State Examination (MMSE) score | Medium (3); Low (5) |
| Presence of anxiety or depression | High (7); Medium (1) |
| Helplessness | High (1); Medium (2); Low (5) |
| Abduction deficiency | High (1); Medium (3); Low (3) |
| Angular deformity | High (4); Medium (1); Low (2) |
| Congenital deformity | High (3); Medium (3); Low (1) |
| Extensor mechanism deficiency | High (4); Low (3) |
| Flexion contracture | High (1); Medium (5); Low (1) |

| Risk Factor | Score Summary (Number Voted for High, Medium, Low) |
|---|---|
| Gait aides | Medium (4); Low (3) |
| Infrapatellar index less than 75% | Medium (1); Low (6) |
| Index of severity for knee disease (ISK) | Medium (1); Low (5) |
| Post traumatic arthritis | Medium (3); Low (4) |
| Previous hip injury | Medium (4); Low (3) |
| Previous infection | High (5); Low (1) |
| Previous open surgery | High (4); Medium (1); Low (2) |
| Previous surgery on lower limb | Medium (2); Low (5) |
| Radiographic severity | High (1); Medium (2); Low (4) |
| Range of motion | High (1); Medium (3); Low (3) |
| Surgical approach | Low (7) |
| Education | High (2); Medium (2); Low (5) |
| Workman's compensation | High (5); Low (4) |
| Marital status | Medium (3); Low (6) |
| Residence (urban/rural) | Medium (3); Low (6) |
| Employment | High (1); Medium (2); Low (6) |
| Race | High (2); Medium (3); Low (4) |
| Socioeconomic status/Income | High (2); Medium (5); Low (2) |

# Appendix E: Candidate Risk Variables Included in Risk Modeling

The variables in Table E1 are the candidate variables from which we selected the proposed risk model variables. The Condition Category (CC) groupings below reflect combinations based upon clinical input and empirical data, as well as how the CCs have been grouped in existing CMS measures, specifically the elective primary THA/TKA readmission and complication measures.

**Table E1. List of Risk Variables Used for Risk Model Selection**

| Risk Variables | Data Source |
|---|---|
| Age | PRO data collection |
| Sex | PRO data collection |
| BMI | PRO data collection |
| Health Literacy (Comfort Filling Out Medical Forms by Yourself) | PRO data collection |
| Narcotic Use for >90 days | PRO data collection |
| Pain in Non-Operative Lower Extremity Joint in past week | PRO data collection |
| Back Pain at preoperative assessment | PRO data collection |
| PROM Survey Respondent (Self or Surrogate) | PRO data collection |
| PROMIS-Global Mental Health subscale score at preoperative assessment | PRO data collection |
| History of Congenital Hip Dysplasia | Derived from ICD-10 codes |
| History of Chronic Narcotic Use or Opioid Abuse | Derived from ICD-10 codes |
| History of Deep Vein Thrombosis/Pulmonary Embolism | Derived from ICD-10 codes |
| History of Hip/Knee Deformities | Derived from ICD-10 codes |
| History of Post-traumatic Arthropathy | Derived from ICD-10 codes |
| History of Smoking | Derived from ICD-10 codes |
| Severe infection; other infectious diseases (CC 1, 3-7) | Condition Categories (CCs) and CC groupings |
| Septicemia, sepsis, systemic inflammatory response syndrome/shock (CC 2) | Condition Categories (CCs) and CC groupings |
| Metastatic cancer or acute leukemia (CC 8) | Condition Categories (CCs) and CC groupings |
| Other major cancers (CC 9-12) | Condition Categories (CCs) and CC groupings |
| Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15) | Condition Categories (CCs) and CC groupings |
| Benign neoplasms of skin, breast, eye (CC 16) | Condition Categories (CCs) and CC groupings |

| Risk Variables | Data Source |
|---|---|
| Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123) | Condition Categories (CCs) and CC groupings |
| Protein-calorie malnutrition (CC 21) | Condition Categories (CCs) and CC groupings |
| Morbid obesity (CC 22) | Condition Categories (CCs) and CC groupings |
| Other significant endocrine and metabolic disorders; disorders of fluid/electrolyte/acid-base balance (CC 23-24) | Condition Categories (CCs) and CC groupings |
| Disorders of thyroid, cholesterol, lipids (CC 25-26) | Condition Categories (CCs) and CC groupings |
| Liver disease (CC 27-31) | Condition Categories (CCs) and CC groupings |
| Gallbladder and biliary tract disorders (CC 32) | Condition Categories (CCs) and CC groupings |
| Chronic pancreatitis (CC 34) | Condition Categories (CCs) and CC groupings |
| Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 36) | Condition Categories (CCs) and CC groupings |
| Other gastrointestinal disorders (CC 38) | Condition Categories (CCs) and CC groupings |
| Bone/joint/muscle infections/necrosis (CC 39) | Condition Categories (CCs) and CC groupings |
| Rheumatoid arthritis and inflammatory connective tissue disease (CC 40) | Condition Categories (CCs) and CC groupings |
| Disorders of the vertebrae and spinal discs (CC 41) | Condition Categories (CCs) and CC groupings |
| Osteoarthritis of hip or knee (CC 42) | Condition Categories (CCs) and CC groupings |
| Osteoporosis and other bone/cartilage disorders (CC 43) | Condition Categories (CCs) and CC groupings |
| Other musculoskeletal and connective tissue disorders (CC 45) | Condition Categories (CCs) and CC groupings |
| Severe hematological disorders (CC 46) | Condition Categories (CCs) and CC groupings |
| Disorders of immunity (CC 47) | Condition Categories (CCs) and CC groupings |
| Coagulation defects and other specified hematological disorders (CC 48) | Condition Categories (CCs) and CC groupings |

| Risk Variables | Data Source |
|---|---|
| Iron deficiency or other/unspecified anemias and blood disease (CC 49) | Condition Categories (CCs) and CC groupings |
| Delirium and encephalopathy (CC 50) | Condition Categories (CCs) and CC groupings |
| Dementia or other specified brain disorders (CC 51-53) | Condition Categories (CCs) and CC groupings |
| Drug/alcohol abuse/dependence/psychosis (CC 54-56) | Condition Categories (CCs) and CC groupings |
| Major psychiatric disorders (CC 57-59) | Condition Categories (CCs) and CC groupings |
| Depression (CC 61) | Condition Categories (CCs) and CC groupings |
| Anxiety disorders (CC 62) | Condition Categories (CCs) and CC groupings |
| Other psychiatric disorders (CC 63) | Condition Categories (CCs) and CC groupings |
| Mental retardation or developmental disability (CC 64-68) | Condition Categories (CCs) and CC groupings |
| Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190) | Condition Categories (CCs) and CC groupings |
| Neuropathy; muscular dystrophy (CC 75-76) | Condition Categories (CCs) and CC groupings |
| Multiple sclerosis; mononeuropathy, other neurological conditions/injuries (CC 77, 81) | Condition Categories (CCs) and CC groupings |
| Parkinson's and Huntington's diseases (CC 78) | Condition Categories (CCs) and CC groupings |
| Seizure disorders and convulsions (CC 79) | Condition Categories (CCs) and CC groupings |
| Coma, brain compression/anoxic damage (CC 80) | Condition Categories (CCs) and CC groupings |
| Respirator dependence/respiratory failure (CC 82-83) | Condition Categories (CCs) and CC groupings |
| Cardio-respiratory failure and shock (CC 84 plus ICD-10-CM codes R09.01 and R09.02, for discharges on or after October 1, 2015; CC 84 plus ICD-9-CM diagnosis codes 799.01 and 799.02, for discharges prior to October 1, 2015) | Condition Categories (CCs) and CC groupings |
| Congestive heart failure (CC 85) | Condition Categories (CCs) and CC groupings |

| Risk Variables | Data Source |
|---|---|
| Acute coronary syndrome (CC 86-87) | Condition Categories (CCs) and CC groupings |
| Coronary atherosclerosis or angina (CC 88-89) | Condition Categories (CCs) and CC groupings |
| Heart infection/inflammation, and valvular and rheumatic heart disease (CC 90 -91) | Condition Categories (CCs) and CC groupings |
| Congenital cardiac/circulatory defects (CC 92-93) | Condition Categories (CCs) and CC groupings |
| Hypertension and hypertensive disease (CC 94-95) | Condition Categories (CCs) and CC groupings |
| Specified arrhythmias and other heart rhythm disorders (CC 96-97) | Condition Categories (CCs) and CC groupings |
| Other and unspecified heart disease (CC 98) | Condition Categories (CCs) and CC groupings |
| Stroke (CC 99-100) | Condition Categories (CCs) and CC groupings |
| Cerebrovascular disease (CC 101-102, 105) | Condition Categories (CCs) and CC groupings |
| Vascular or circulatory disease (CC 106-109) | Condition Categories (CCs) and CC groupings |
| Chronic obstructive pulmonary disease (COPD) (CC 111) | Condition Categories (CCs) and CC groupings |
| Fibrosis of lung or other chronic lung disorders (CC 112) | Condition Categories (CCs) and CC groupings |
| Asthma (CC 113) | Condition Categories (CCs) and CC groupings |
| Pneumonia; pleural effusion/pneumothorax (CC 114-117) | Condition Categories (CCs) and CC groupings |
| Other respiratory disorders (CC 118) | Condition Categories (CCs) and CC groupings |
| Legally blind (CC 119) | Condition Categories (CCs) and CC groupings |
| Other retinal disorders (CC 125) | Condition Categories (CCs) and CC groupings |
| Other eye disorders (CC 128) | Condition Categories (CCs) and CC groupings |
| Significant ear, nose, and throat disorders (CC 129) | Condition Categories (CCs) and CC groupings |

| Risk Variables | Data Source |
|---|---|
| Other ear, nose, throat, and mouth disorders (CC 131) | Condition Categories (CCs) and CC groupings |
| Transplants (CC 132, 186) | Condition Categories (CCs) and CC groupings |
| Dialysis status (CC 134) | Condition Categories (CCs) and CC groupings |
| Renal failure (CC 135-140) | Condition Categories (CCs) and CC groupings |
| Nephritis (CC 141) | Condition Categories (CCs) and CC groupings |
| Urinary incontinence (CC 143) | Condition Categories (CCs) and CC groupings |
| Urinary tract infection (CC 144) | Condition Categories (CCs) and CC groupings |
| Other urinary tract disorders (CC 145) | Condition Categories (CCs) and CC groupings |
| Other female genital disorders (CC 148) | Condition Categories (CCs) and CC groupings |
| Male genital disorders (CC 149) | Condition Categories (CCs) and CC groupings |
| Decubitus ulcer or chronic skin ulcer (CC 157-161) | Condition Categories (CCs) and CC groupings |
| Cellulitis, local skin infection (CC 164) | Condition Categories (CCs) and CC groupings |
| Other dermatological disorders (CC 165) | Condition Categories (CCs) and CC groupings |
| Trauma (CC 166-168, 170-173) | Condition Categories (CCs) and CC groupings |
| Vertebral fractures without spinal cord injury (CC 169) | Condition Categories (CCs) and CC groupings |
| Other injuries (CC 174) | Condition Categories (CCs) and CC groupings |
| Poisonings and allergic and inflammatory reactions (CC 175) | Condition Categories (CCs) and CC groupings |
| Major complications of medical care and trauma (CC 176-177) | Condition Categories (CCs) and CC groupings |
| Major symptoms, abnormalities (CC 178) | Condition Categories (CCs) and CC groupings |

| Risk Variables | Data Source |
|---|---|
| Minor symptoms, signs, findings (CC 179) | Condition Categories (CCs) and CC groupings |

# Appendix F: Clinically Derived vs. Empirically Derived Risk Model Measure Results

Model performance statistics for the clinically derived models are highly comparable to those for the empirically derived models, and scatterplot analyses demonstrated high correlation of hospital-specific RSIRs regardless of the model approach (Figure F1 and Figure F2).

**Figure F1. Scatterplot of Clinically Derived RSIR by Empirically Derived RSIR by Hospital for THA Patients on HOOS, JR**

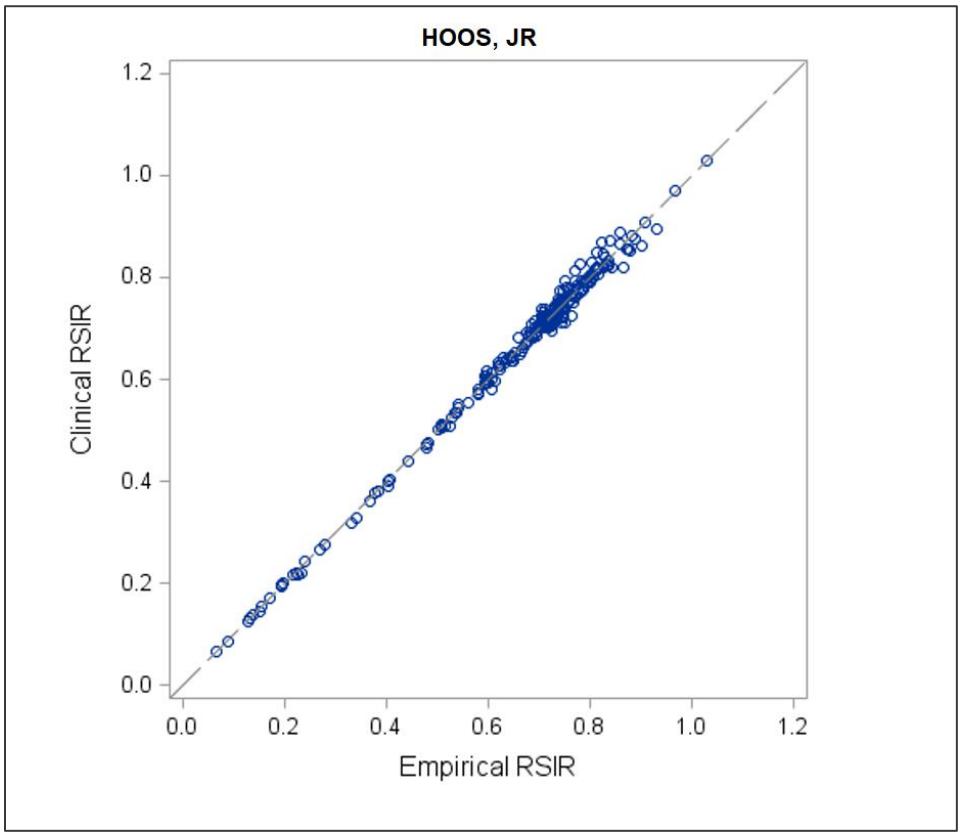


**Figure F2. Scatterplot of Clinically Derived RSIR by Empirically Derived RSIR by Hospital for TKA Patients on KOOS, JR**