Summary of Technical Expert Panel (TEP) Meetings Clinician-Level and Clinician Group-Level Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA and TKA) Patient-Reported Outcome-Based Performance Measure (PRO-PM)

October 2021

Prepared by:
Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

This material was prepared by CORE under contracts to the Centers for Medicare & Medicaid Services (CMS), an agency of the U.S. Department of Health and Human Services. The contents presented do not necessarily reflect CMS policy.

Table of Contents

Background

The Centers for Medicare & Medicaid Services (CMS) has contracted with Yale New Haven Health Services Corporation - Center for Outcomes Research and Evaluation (CORE) to develop quality measures of hospital and clinician performance. Under this contract, CORE is developing a Clinician-Level and Clinician Group-Level Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA and TKA) Patient-Reported Outcome-Based Performance Measure (PRO-PM). The contract name is Development, Reevaluation, and Implementation of Outcome/Efficiency Measures for Hospital and Eligible Clinicians, Option Period 2. The contract number is HHSM-75FCMC18D0042, Task Order Number HHSM-75FCMC19F0001.

CORE is obtaining expert and stakeholder input on the proposed measure. The CORE measure development team is comprised of experts in quality outcomes measurement and measure development. As is standard with all measure development processes, CORE convened a technical expert panel (TEP) of clinicians, patient advocates, and other stakeholders. Collectively, the TEP members brought expertise in performance measurement, quality improvement, the patient experience, and orthopedics, specifically THA and TKA procedures.

This report summarizes the feedback and recommendations received from the TEP during meetings one, two, and three which focused on the measure concept, the proposed measure development approach, preliminary measure specifications, the risk model, approach to non-response bias, social risk factor analyses, measure reliability, and measure specifications updates.

Measure Development Team

Rachelle Zribi, BA leads the measure development team for this Clinician-Level and Clinician Group-Level THA/TKA PRO-PM. Ms. Zribi is a Research Project Coordinator II for the Quality Measurement Team at CORE and has supported several novel measure development teams including those developing PRO-PMs. The remainder of the measure development team provide a range of expertise in outcome measure development, health services research, clinical medicine, statistics, and measurement methodology. See Appendix A for the full list of members for the CORE Measure Development Team.

The TEP

In alignment with the CMS Measures Management System (MMS), CORE held a 30-day public call for nominations and convened a TEP for the development and reevaluation of orthopedic measures, including the development of the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM. CORE solicited potential TEP members via emails to individuals and organizations recommended by the measure development team and stakeholder groups, email blasts sent to CMS physician and hospital email listservs, and through a posting on CMS's website. The TEP is composed of 20 members, listed in Table 1.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

The role of the TEP is to provide feedback and recommendations on key methodological and clinical decisions. The appointment term for the TEP is from March 2021 to March 2022.

Specific Responsibilities of the TEP Members

- Complete and submit all nomination materials, including the TEP Nomination Form, statement of interest, and curriculum vitae
- Review background materials provided by CORE prior to each TEP meeting
- Attend and actively participate in TEP conference calls
- Provide input on key clinical, methodological, and other decisions
- Provide feedback on key policy or other non-technical issues
- Review the TEP summary report prior to public release
- Be available to discuss recommendations and perspectives following TEP meetings and public release of the TEP Summary Report to CMS

Table 1. TEP Member Name, Affiliation, and Location

Name	Title, Organization	Location
David C. Ayers, MD	Professor and Chair of Orthopaedics, University of Massachusetts Medical School	Worcester, MA
Thomas C. Barber, MD	Deputy Physician in Chief, Memorial Sloan Kettering Hospital	New York, NY
Phyllis Bass	Patient Representative	Cypress, TX
Vinod Dasa, MD	Professor of Clinical Orthopedics and Director of Research, Louisiana State University Health Sciences Center	New Orleans, LA
Rachel DuPré Brodie	Senior Director of Measurement & Accountability, Purchaser Business Group on Health (PBGH)	San Francisco, CA
Cheryl Fahlman, PhD, MBA, BSP	President, CAF Consulting Solutions	Gaithersburg, MD
William G. Hamilton, MD	Chair of Orthopedic Surgery at Inova Mount Vernon Hospital and FOCAL Chair of the American Association of Hip and Knee Surgeons	Alexandria, VA

Name	Title, Organization	Location
Cynthia S. Jacelon, PhD, RN-BC, CRRN, FGSA, FAAN	Professor and Executive Associate Dean, University of Massachusetts Amherst School of Nursing	Greenfield, MA
Craig T. Miller, PT	Director of Home Care Therapy and Senior PT, Rivetus Rehabilitation; American Physical Therapy Association	Macomb, MI
Michael H. Perskin, MD	Clinical Professor of Medicine, New York University School of Medicine and American Geriatrics Society	New York, NY
Patient A	Patient Representative	Chicago, IL
Patient B	Patient Representative	Los Angeles, CA
Nan Rothrock, PhD	Associate Professor of Medical Social Sciences, Feinberg School of Medicine at Northwestern University	Chicago, IL
Jonathan L. Schaffer, MD, MBA, FACS, FHIMSS, FABOS	Staff and Program Director, The Cleveland Clinic	Cleveland, OH
Adam Schwartz, MD, MBA	Associate Professor of Orthopaedic Surgery, Mayo Clinic	Phoenix, AZ
Robert Sterling, MD	Associate Professor of Orthopaedic Surgery and Vice Chair for Quality Safety, and Service, Johns Hopkins University; Association of Hip and Knee Surgeons	Baltimore, MD
Margaret A. VanAmringe, MHS	Vice President for Public Policy and Government Relations, The Joint Commission	Washington, DC
Christine Von Raesfeld	Patient Representative	Santa Clara, CA
Patricia Walker, PhD	Patient Representative	South Holland, IL

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

Name	Title, Organization	Location
Kevin Woodward, PA-C, MMS	Physician Assistant of Orthopaedic Surgery, John Hopkins University; American Academy of Physician Assistants; Maryland Academy of Physician Assistants	Baltimore, MD
Adolph J. Yates, MD	Professor and Vice Chairman, Department of Orthopedic Surgery, University of Pittsburgh; American Association of Hip and Knee Surgeons	Pittsburgh, PA

TEP Meetings

CORE held its first TEP meeting in August 2020, a meeting at which both the Clinician-Level and Clinician-Group Level THA/TKA PRO-PM and Center for Medicare & Medicaid Innovation (CMMI) Episode Payment Model (EPM) THA/TKA Complication Measure were presented. The Clinician-Level and Clinician Group-Level THA/TKA PRO-PM team held its second TEP meeting on February 10, 2021, the third TEP meeting on March 23, 2021, and the fourth meeting on July 13, 2021 (see Appendix B for the TEP meeting schedule). This summary report contains a summary of the four TEP meetings. The presentation of the EPM THA/TKA Complication measure is presented in a separate summary report.

TEP meetings follow a structured format consisting of the presentation of key issues identified during measure development, as well as CORE's proposed approaches to addressing the issues, followed by an open discussion of these issues by the TEP members.

First TEP Meeting Overview

Prior to the first TEP meeting, TEP members received detailed meeting materials outlining the measure background and proposed approach to measure re-specification for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM and the EPM THA/TKA Complication Measure. One TEP member shared several measure specification questions and provided input prior to the meeting. For further details, please see Appendix C.

During the first TEP meeting, CORE solicited feedback from the TEP on the measure concept and proposed approach to measure re-specification. CORE educated the TEP on the background and approach to developing the EPM THA/TKA Complication Measure and the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM Measure. Information on how the EPM THA/TKA Complication Measure and the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM align with the existing Hospital-level THA/TKA Complication Measure and the Hospital-level THA/TKA PRO-PM, respectively, was also provided. The TEP was invited to provide input on the measure concepts and approaches to each measure re-specification.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

Following the meeting, TEP members unable to join the TEP teleconference were given the meeting recording and detailed meeting minutes.

The following bullets represent a high-level summary of what was presented and discussed relevant for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM during the first TEP meeting. For further details, please see Appendix D. (Please note that the high-level summary of the presentation and discussion of the EPM THA/TKA Complication Measure is presented in a separate TEP Summary Report.)

Measure Background and Approach to Re-specification

- CORE Presentation of a Measure Overview to the TEP
    - CMS contracted CORE to re-specify the existing hospital-level THA/TKA PRO-PM as a clinician or clinician group measure for the Quality Payment Program (QPP).
    - Ms. Zribi reviewed the project background, the existing hospital-level THA/TKA PRO-PM specifications, the approach to measure re-specification, and future topics the team will raise with the TEP.
    - For measure re-specification, CORE will use data from the development of the hospital-level THA/TKA PRO-PM (submitted through CMMI's Comprehensive Care for Joint Replacement [CJR] Model voluntary PRO data collection), and will solicit input from the TEP, the Orthopedic Clinical Working Group, the Patient Working Group, and via a public comment period.
    - TEP Feedback
    - Several TEP members noted challenges to implementation and response rates. One TEP member noted the challenge of implementing the Clinician-Level and Clinician Group-Level and Hospital-level THA/TKA PRO-PMs without national data collection and submission pathways. Several TEP members noted challenges of obtaining high response rates at follow up. One member specifically identified that small or rural hospitals with resource limitations may be specifically challenged.
    - Two TEP members advocated for incentivizing providers. One TEP member suggested that CMS should provide larger financial incentives to support the capacity to build the data infrastructure and develop clinical workflows for patient-reported outcome measure (PROM) data collection.
        - CORE agreed that patient-level data collection, such as PROs, are dependent on hospital and clinician resources. CORE noted that the TEP can help CMS learn from the institutions capturing PROM data as well.
    - Two TEP members noted the importance of considering measuring PROMs in other care settings, such as ambulatory surgical centers (ASCs) and hospital outpatient departments (HOPDs), as procedures are commonly performed outside of the inpatient setting. One TEP member noted that many clinicians or clinician groups will perform procedures in more than one type of facility and the measure would collect a greater volume of data including procedures in other settings. One TEP member noted the need to measure differences in outcomes

based on inpatient and outpatient status. One TEP member noted a need for the orthopedic outcome measures, including the PRO-PM, to ensure harmonization of the measure cohorts across measurement of ASC, HOPD, and inpatient settings.
    - CORE acknowledged the performance of procedures in multiple settings and noted that considerations of including multiple settings will require measurement considerations, specifically differences in patient severity per setting. More understanding must be gained before a decision is made.
- TEP members commented on the post-operative PRO data collection. One TEP member questioned whether response rates would be higher if the timeframe for post-operative data collection was 6-9 months after the procedure instead of 9-12 months. Another TEP member suggested expanding the timeframe two months before and after the 12-month post-operative data collection deadline for patients who cannot come in exactly at one year for follow-up.
    - CORE noted that the 9-12-month post-operative timeframe was chosen because clinicians wanted to allow enough time for adequate recovery. CORE has heard from some physicians requesting that this data collection window expand beyond 12 months to broaden data capture.
- Summary
- TEP members generally supported the measure developers' approach to re-specification while providing critical points for developers to consider regarding implementation and response rates, expansion of the post-operative PRO data collection timeframe, and measuring outcomes in outpatient and ASC settings. Several TEP members raised concern about the burden of measure implementation and challenges in achieving high response rates.

Further comments from a TEP member about concerns with socioeconomic risk factor analysis, measure implementation, and additional measure specifications are contained in a detailed summary of the pre-TEP meeting email provided in Appendix C.

Second TEP Meeting Overview

Prior to the second TEP meeting, TEP members received a copy of the meeting slides that contained information regarding CORE's approach to clinician and clinician group attribution as well as the risk model approach and results. One TEP member sent an email prior to the meeting, following receipt of the meeting materials, with several comments regarding the risk model: a concern regarding the c-statistic results; a suggestion to use pre-operative Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS JR)/Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS JR) scores in risk adjustment; a question about how body mass index (BMI) is captured in the risk model; and a comment regarding risk adjustment for social risk factors. Two TEP members provided feedback via email after the TEP meeting. One TEP member asked about variation in outcome rates based on hospital location (rural, urban, suburban) or public versus private institutions, the relationship

between changes in PROs and use or lack or use of ancillary services (such as physical therapy), and clarification regarding prior joint replacements as an exclusion criterion. The other TEP member recommended consideration of high-risk patients by including a history of organ transplantation, active cancer diagnosis, use of immunosuppressant medications, race and disparities as factors in the risk model, and suggested novel statistical methods such as machine learning to create the risk model. For further details on emails sent prior to and following the second TEP meeting, please see Appendix E.

During the second TEP meeting, CORE solicited feedback from the TEP on the clinician and clinician group attribution approach and results as well as risk model approach and results. CORE described the approach taken for clinician and clinician group attribution, including rationale as to why certain groups of clinician specialties were removed from consideration in the attribution approach. Afterwards, CORE shared the approach taken for the risk model and summary of the risk model results. CORE additionally explored alternative options and gave rationale for the approach ultimately taken by the measure team. Results were shared and the TEP was invited to provide feedback.

Following the meeting, TEP members unable to join the TEP teleconference were given the meeting recording and detailed meeting minutes.

The following bullets represent a high-level summary of what was presented and discussed during the second Clinician-Level and Clinician Group-Level THA/TKA PRO-PM TEP meeting. For further details, please see Appendix F.

Clinician and Clinician Group Attribution Approach and Results

- CORE presentation of the Attribution Approach and Results to the TEP
    - CORE identified that the attribution approach is based on the attribution algorithm developed for the clinician- and clinician group-level THA/TKA Complication measure. The algorithm attributes the outcome for each patient in the cohort to a single clinician using billing data for the procedure. For clinician group attribution, the group is determined by aggregating providers who use the same Tax Identification Number (TIN).
    - The CORE team proposed a refinement to the existing attribution algorithm to only consider these clinician specialties: Orthopedic Surgery, Sports Medicine, Hand Surgery, Orthopedic Manipulative Medicine, and General Surgery to increase face validity and remove anomalous clinician specialties unrelated to orthopedics. CORE presented the frequency and percentage of patients attributed to each clinician specialty in the testing dataset.
    - TEP Feedback
    - A TEP member asked for clarification on whether outpatient procedures are excluded from this measure.

        - Ms. Zribi explained the measure is currently specified for inpatient procedures, but that CORE received input suggesting expansion of the cohort to include procedures performed in outpatient and ASC settings.
    - One TEP member requested clarification of inpatient procedures billed as outpatient procedures.
        - Ms. Zribi noted that current cohort criteria only consider procedures billed as inpatient procedures.
    - Several TEP members commented on the Two Midnight Rule and billing approaches now that CMS removed TKAs and THAs from the Inpatient Only (IPO) list. One TEP member noted that a patient could be hospitalized for less than two midnights and still be considered an inpatient. Another TEP member noted that many hospitals bill as "extended recovery" if an inpatient was expected to stay less than two midnights.
        - CORE agreed with these concerns about appropriate identification of inpatient from outpatient procedures and anticipates expansion of the measure cohort to include procedures performed in the HOPD and ASC settings. CORE will bring this topic to the TEP at the next meeting.
        - Four TEP members agreed with capturing procedures across settings given that an increasing number of THA/TKA procedures are being performed at ASC and HOPD settings. One TEP member specifically noted the concern that excluding ASC and HOPD procedures from the current measure could misrepresent a provider's eligible cohort size. One TEP member noted a benefit that inclusion of procedures in ASC and HOPD settings could increase the number of providers that would be eligible for the measure.
        - One TEP member suggested separating inpatient and outpatient data because the outcomes could be different.
    - A TEP member asked why specialties with a low number of attributed procedures (Hand Surgery and Osteopathic Manipulative Medicine) were still included in the attributed algorithm.
        - CORE replied that providers from these specialties were included in testing since they had patients with complete PRO data.
        - Dr. Kevin Bozic, CORE's clinical consultant, added that this increases face validity. The measure will not separate these specialties and analyze them individually; instead, they are in the total denominator of cases used to create the measure.
    - One TEP member asked how the specialty description was assigned to a surgeon, and another TEP member asked if CORE would examine the types of providers in the attribution algorithm over time.
        - CORE explained that the specialty description is assigned by Medicare through the coding process. CORE noted that the measure team will continue to evaluate the attribution algorithm to include as many

clinically appropriate patients and providers as possible for face validity purposes.

- Three TEP members commented on specific specialty descriptions. One TEP member noted the low number of Sports Medicine clinicians. Another TEP member asked whether the hand surgeons in the testing data performed hemiarthroplasties for fractures. Another TEP member asked if any procedures in the dataset were attributed to adult reconstructive surgeons.
  - CORE noted that the CJR dataset used for measure testing is limited and may not be representative of frequencies at the national level.
  - CORE further noted only elective, primary procedures are included in the measure cohort and revisions and fractures are excluded from the denominator.
  - After the TEP meeting, CORE shared that Medicare Claims documentation indicates a specialty assignment group of "Plastic and reconstructive surgery". No reconstructive surgeons were attributed procedures in our dataset using the attribution algorithm. These providers may have submitted claims for elective, primary hip and knee procedures, but they were not identified as the primary clinician using the attribution algorithm.
- One TEP member noted that because hospitals and medical systems decide whether to participate in CJR, this clinician- and clinician group-level measure is an unfair evaluation of surgeons. They noted that if hospitals were held accountable in this measure, it would drive universal reporting of the measure for all entities.
- One TEP member asked if the testing dataset included the latest CJR data.
  - CORE confirmed that the testing dataset used included the most recent CJR data for which cleaning and matching had been completed.

Risk Model Approach and Results

- CORE Presented the Risk Model Approach to the TEP
  - CORE presented the risk model approach, a clinically- and empirically-derived risk model initially developed by the hospital-level THA/TKA PRO-PM development team, with risk variables identified in the literature and selected by orthopedic experts for their relationship with the measure outcome. CORE reviewed the 19 risk variables included in the model.
  - CORE reviewed the risk model performance results. The c-statistic indicated moderate model discrimination across the development and validation models and was slightly lower than that of the hospital-level THA/TKA PRO-PM. CORE described approaches taken to explore differences in c-statistic results between the clinician-level and hospital-level risk model performance results, and noted specifically that health literacy, a risk variable in the model, showed a strong, statistically significant and linear relationship with clinical improvement following THA/TKA in the first year of testing data but not in the second year of

data. CORE intends to use the clinically- and empirically-derived risk models but given that the CJR data represents a unique sampling of patients undergoing THA/TKA procedures, will re-test the risk model before mandatory use of the measure in the future.

- TEP Feedback
    - Multiple TEP members supported CORE's recommendations and expressed interest in seeing updated results using newer CJR data.
    - One TEP member noted that patient selection by providers may interfere with risk adjustment. They additionally addressed the challenges of PRO data collection.
    - Several TEP members commented on the risk model results. One TEP member suggested that differences in CJR performance year data could be due to an increase in coding for comorbidities over time. Another TEP member suggested that improved care via the CJR model could result in a lower c-statistic over time. A third TEP member suggested crediting the clinician if they enhance health literacy rates by educating patients. They asked how health literacy was captured.
        - CORE uses a widely validated health literacy questionnaire with minimal provider and patient burden.
    - Two TEP members asked for clarification regarding the interpretation of the c-statistic values. One TEP member expressed concern regarding the c-statistic and asked if the CORE team had a target for the c-statistic value.
        - CORE explained that it is challenging to determine a threshold for a c-statistic given many factors influencing the outcome. Of note, other quality outcome measures have similar c-statistic values as this measure.
        - Another TEP member found that the c-statistic was acceptable.
    - One TEP member commended CORE on their testing methodology. Another TEP member highlighted the importance of risk adjustment and expressed support of aligning the risk model with the hospital-level measure.
    - One TEP member noted that some social risk factors are omitted from the model.
        - CORE noted that social risk and response bias will be discussed at a later TEP meeting.
    - One TEP member noted the importance of understanding response rates and associations with "cherry picking".
    - Several TEP members suggested consideration of additional risk variables:
        - BMI as a continuous variable
        - smoking history
        - preoperative physical composite score (as measured by the Veteran's RAND 12-Item Health Survey [VR-12] and Patient-Reported Outcomes Measurement Information System [PROMIS] Global-10 PROMs)
        - orthopedic comorbidities such as concurrent hip and knee pain

        - the Charlson Comorbidity Index to account for the synergistic effects of multiple comorbidities
        - Parkinson's Disease, as it may be a primary reason for dislocation
        - patients' long-term use of medications for treatment of comorbidities
        - patient expectation, as it is highly associated with good clinical outcomes
            - CORE acknowledged suggested risk variables.
            - CORE noted that variables indicating frailty are included in the risk model and will investigate Parkinson's Disease as a potential risk factor.
            - CORE noted that medication data can be difficult to capture.
    - One TEP member asked if CORE examined differences between THA and TKA and the proportion of those surgery types. They noted that some clinicians or clinician groups may perform more of one procedure than the other, which could affect the fairness of the measure. They also recommended patient stratification by race and ethnicity for the risk model.
- Summary
    - TEP members generally supported the measure developers' approach to clinician and clinician group attribution and supported ongoing evaluation of the specialties included. TEP members generally supported the measure developers' approach to the risk model and suggestion consideration of additional variables in the risk model. TEP members supported ongoing evaluation of the risk model in the future.

Third TEP Meeting Overview

Prior to the third TEP meeting, TEP members received meeting materials that contained information regarding CORE's approach to social risk factors, non-response bias, measure reliability testing results, and measure specification updates. These topics were discussed during the third TEP meeting and TEP members were invited to provide feedback.

Following the meeting, TEP members unable to join the TEP teleconference were given the meeting recording and detailed meeting minutes.

The following bullets represent a high-level summary of what was presented and discussed during the third Clinician-Level and Clinician Group-Level THA/TKA PRO-PM meeting. For further details, please see Appendix G.

Social Risk Factor Approach and Results

- CORE Presented the Approach to Social Risk Factors and Results to the TEP
    - CORE reviewed the approach to social risk factor adjustment for non-White race, dual eligibility, and Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index.

- CORE reviewed the social risk factor analyses results, noting that there was little to no impact in adding these additional risk factors to the risk model as evidenced by c-statistics.
- CORE recommends adjusting for health literacy in the risk model and including the other social risk factors in the statistical weighting used to address potential non-response bias. CORE will monitor the relationship between social risk, response rate, and measure results in the future.

- TEP Feedback
  - Multiple TEP members agreed with continuous monitoring of the impact of social risk on the measure results.
  - One TEP member noted that only 7.7% of THA/TKA patients were non-White and asked how the testing data cohort compares to the Medicare population.
    - CORE noted that this percentage does reflect the Medicare patient population undergoing primary elective THA/TKA and that the small proportion likely reflects unequal access to healthcare.
  - A TEP member asked how race is reported.
    - CORE explained that race is identified using the Medicare enrollment database and may not be self-reported. CORE reviewed the options for race in claims data.
  - A TEP member suggested using racial stratification instead of social risk factors in risk adjustment, such as the method used in the 21st Century Cures Act.

Non-Response Bias Approach and Results

- CORE Presented the Statistical Approach to Non-Response Bias to the TEP
- CORE reviewed the importance of assessing non-response bias in PRO-PMs, noting that bias may be introduced if there are systematic differences between patients that respond to PROM surveys versus those who do not, including differences in social risk or clinical outcomes.
- CORE presented PRO submission rates for clinicians and clinician groups.
- CORE reviewed the statistical approach for accounting for non-response bias which was utilized for the hospital-level THA/TKA PRO-PM. CORE presented risk-standardized improvement rates (RSIRs) with and without weighting for potential non-response bias.
- While there was little impact from weighting for non-response bias on the results using the dataset, CORE recommends including non-response weighting in the calculation of final measure results and continuously evaluating non-response bias moving forward.
- TEP Feedback
- A TEP member asked what the variability was in response rates between clinicians.
- CORE reviewed the PRO submission rates. CORE noted that the submission rates are likely low because the CJR Model gradually incentivized response rates over multiple years.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

- One TEP member suggested using the propensity score to calculate observed and expected response rates for clinicians and clinician groups. They recommended considering a clinician-level adjustment that is based on their specific patient population.
- Another TEP member suggesting using surgical volume to calculate a clinician- or clinician-group specific expected PRO response rate.
- CORE noted that since the CJR model incentivizes an absolute number or proportion of successful PRO submissions, there is variation in how surgical volume will be associated with PRO response rates.
- Two TEP members expressed concern regarding the opportunity for hospitals to offer surveys only to patients who are likely to show substantial clinical benefit (i.e., "cherry picking").
- CORE recommended maximum flexibility in PRO data collection to mitigate unintended consequences. CORE noted that the CJR model provides a variety of ways for hospitals to collect PRO data and agrees that CMS should consider an implementation structure that addresses this concern.
- One TEP member suggested coupling the measure with rewards and penalties for the proportion of successful PRO data collected. Additionally, they recommended sharing with clinicians that hospitals with high PRO response rates are typically more successful. This strategy may be effective in encouraging hospitals to increase their reporting rates overall.
- One TEP member clarified that CORE assumes social determinants of health may contribute to non-random missing PRO survey data.
- CORE confirmed this assumption given that the testing dataset was relatively small, and when the measure is implemented in a national scale, social risk factors will likely impact non-response. CORE seeks to support clinicians and hospitals in limiting non-response bias and noted the importance of integrating PRO data collection into the clinical workflow.

Measure Results

- CORE Presented Measure Results to the TEP
  - CORE presented the signal-to-noise reliability testing results, which showed excellent reliability for clinician and clinician groups.
- TEP Feedback
  - One TEP member clarified with CORE how the signal-to-noise results were calculated.

Measure Specification Updates

- CORE Presented Proposal to Expand Measure Cohort to the TEP
  - CORE proposed an expansion of the measure cohort to include eligible THA/TKA procedures that occur in HOPD and ASC settings. This recommendation comes in response to recent CMS rulings that expands payment for THA/TKA procedures

in these settings, and feedback from the Rural Measures Application Partnership and from clinical experts.

- TEP Feedback
    - TEP members overwhelmingly supported the recommendation to expand the measure cohort. One TEP member noted that not expanding the cohort would be detrimental, as clinicians could evade the measure by switching a patient to another setting.
    - Multiple TEP members emphasized the importance of standardizing response rate incentives across settings.
    - One TEP member asked if the expected improvement rates would be equal across settings. The TEP member emphasized the importance of having comprehensive, transparent, and publicly reported quality scores as well as having equitable measure expectations across settings. Another TEP member suggested CORE look at each entity as a unit of accountability.
    - Another TEP member expressed concern regarding separate public reporting across settings, noting that patients may prefer to have procedures performed in the setting with the highest improvement scores even if that setting might be clinically inappropriate for the patient.
        - CORE clarified that the current measure is being developed for the Quality Payment Program (QPP), which is setting agnostic.
    - One TEP member noted that this would not be significantly burdensome as the CJR model would begin collecting data from HOPDs according to the new CJR Final Rule. They asked whether the measure would identify eligible patients using International Classification of Diseases, Tenth Revision (ICD-10) codes or Current Procedural Terminology (CPT) codes.
        - CORE noted that at the time of this meeting, the CJR Final Rule had not been published. Eligible patients from HOPD and ASC settings would be captured in the clinician-level measure in the future regardless of how the procedures are coded.
    - The TEP member noted an increase in same-day discharges for Medicare patients and emphasized the importance of capturing data from HOPDs and ASCs because of an increase in cardiopulmonary complications in Medicare patients who were discharged the same day of their surgery.
    - Another TEP member supported capturing patients who are discharged the same day. They asked whether CMS could distinguish same-day discharge patients.
        - CORE relies on billing codes which indicate the setting of the procedure and will investigate whether same-day discharges can be identified in billing codes.
    - Another TEP member noted that if a patient is billed as an outpatient under the Healthcare Common Procedure Coding System (HCPCS) codes, the physician cannot bill for an observation night because the overnight stay is bundled into the HCPCS code for the procedure. They recommended that CORE look into the date of the procedure and discharge to identify whether the patient had an

overnight observation stay. The TEP member also noted that some states allow ASCs to act as overnight facilities.
    - A TEP member asked whether the American Society of Anesthesiologists (ASA) Score data is captured.
        - CORE had explored including ASA scores but heard from stakeholders that this would be unnecessarily burdensome for anesthesiologists.
    - One TEP member noted that the California Joint Replacement Registry aimed to add ASC data to the registry since this would increase the number of facilities in the registry, increase the volume of patients per surgeon, and also volume of patients per group since many of the surgeons perform surgeries in more than one facility. Additionally, inclusion of ASC and HOPD data would mean that more clinicians and groups would be eligible for performance measurement.
- CORE Presented Proposal to Extend Postoperative Data Collection to the TEP
    - CORE proposed an extension of the post-operative data collection window from 9-12 months post-procedure to 10-14 months post-procedure, aligning with stakeholder feedback that post-operative follow-up appointments are typically scheduled for one year or later. This extension would likely give providers more opportunity to collect complete post-operative PRO data.
- TEP Feedback
    - TEP members overwhelmingly supported this recommendation.
    - Two TEP members suggested expanding the post-operative data collection window to 9-24 months, noting the outcome likely will not change between 12-24 months and that a longer window would capture patients who cannot return for a one-year follow-up.
    - Another TEP member noted this lengthy timeframe could hinder actionable use of data and supported the 10–14-month window.
    - One TEP member asked how this window would operate for patients who received staged procedures.
        - CORE is actively discussing how to handle staged procedures in this measure. Currently, staged procedures that occur within the measurement period are excluded from this measure due to overlap of recovery time. CORE invites feedback regarding the defined timeframe for staged procedures.
- Summary:
    - TEP members generally supported the measure developers' approach to social risk factor analyses, non-response bias approach and results, measure reliability, and measure updates. TEP members supported continuous monitoring of the impact of social risk on the measure results and ongoing evaluation of response rates. TEP members were strongly supportive of the measure updates to expand the cohort and extend the postoperative PRO data collection timeframe.

Fourth TEP Meeting Overview

Prior to the fourth TEP meeting, TEP members received meeting materials that contained information regarding CORE's approach to measure score validity and results, updated measure specifications, and questions inquiring about the measure's face validity. These topics were discussed during the fourth TEP meeting and TEP members were invited to provide feedback.

Following the meeting, TEP members unable to join the TEP teleconference were given the meeting recording. One TEP member provided feedback via email after the TEP meeting with details about physical therapy outcomes measures and the American Physical Therapy Association (APTA) Qualified Clinical Data Registry (QCDR). For further details on emails sent prior to and following the fourth TEP meeting, please see Appendix H.

The following bullets represent a high-level summary of what was presented and discussed during the fourth Clinician-Level and Clinician Group-Level THA/TKA PRO-PM meeting. For further details, please see Appendix I.

Measure Score Validity Approach and Results

- CORE Presented the Approach to Measure Score Validity and Results to the TEP
    - CORE reviewed the approach to measure score validity testing which compared the measure results of the clinician- and clinician group-level THA/TKA PRO-PM RSIRs with the clinician- and clinician group-level THA/TKA complication measure risk-standardized complication rates (RSCRs). CORE also examined the distribution of RSIRs compared to the RSCRs, as well as to the RSCR performance categories.
    - CORE presented the correlation results for clinicians and clinician groups with more than 25 procedures showing a low correlation between RSIRs and RSCRs.
    - CORE presented the clinician- and clinician group-level results of RSIRs (without the 25-volume restriction) categorized by national performance categories, noting few outliers among clinicians and clinician groups and a wide range of RSIRs which made it difficult to identify a relationship between improvement rates and RSCR performance categories.
    - CORE noted they were not surprised by the results given a priori expectations: correlation results are influenced by the testing dataset, including collection incentives in the CJR model at the hospital level, differences in timeframes and severities of measure outcomes, and differences in the groups of patients included in the cohorts, which can reduce any expected correlation. CORE noted that the difference in outcome rates may also impact these analyses as the complication measure assesses a rare occurrence (around 3% for clinicians and clinician groups), while RSIRs are much higher (around 64%).
    - CORE acknowledged the limitations of the testing dataset and will advise CMS to retest measure score validity in a larger dataset in the future as feasible.
- TEP Feedback

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

- Multiple TEP members noted they did not expect to see correlations between these measures, and were also not surprised by the results.
- One TEP member expressed concern over combining low- and high-volume clinicians in the results, as clinicians with fewer procedures could have higher complication rates. They suggested separately comparing low-volume and high-volume clinicians.
    - CORE noted they looked at all clinicians in the dataset (including those without the 25-volume restriction) and provided the correlation details as correlation without volume thresholds: -0.034 (p=0.2316) at clinician level and, -0.008 (p=0.8491) at clinician-group level.
    - CORE acknowledged the importance of looking at the relationship between volume and quality outcomes in THA/TKA procedures. CORE plans to continue to explore issues of variable response rates and overall provider volume in the future.
- One TEP member asked if CORE explored other outcome measures that may show a potentially better correlation with the QPP THA/TKA PRO-PM, possibly measures of utilization or Medicare Spending Per Beneficiary (MSPB).
    - CORE explored using the Consumer Assessment of Healthcare Providers and Systems (CAHPS) for MIPS, but these patient experience measures were specific to primary care providers, and it was unlikely clinicians in this measure would report on that measure. CORE also reviewed the AHRQ's CAHPS Surgical Care Survey and chose not to pursue it given the survey focused on patient experience rather than outcomes.
- One TEP member asked about the distribution of the 200 hospitals in CJR and if the hospitals were similar or diverse.
    - CORE noted there are approximately 200 hospitals participating in the voluntary CJR PRO data collection, which is a small proportion of the approximately 4,000 hospitals that perform THAs/TKAs and reconfirmed there are few statistical outliers on the complication measure.
    - CORE confirmed they looked at the overall distribution of hospital types submitting PRO data in CJR. Overall, the distributions of hospital types (based on teaching status and bed size) were similar across the performance years.
    - Another TEP member noted the hospital characteristics in the CJR data were based on metropolitan statistical areas (MSAs) and as such contain mostly urban and suburban hospitals, which may result in some bias.
- The same TEP member noted these measures and timeframes are very different with potential for selection bias. They suggested using the THA/TKA readmission measure (National Quality Forum [NQF] #1551) to correlate readmissions as an alternative to the complications. Also, the range of the complication measure outcomes is narrow, so there's a small window for complications versus improvement in PROs. Given the CJR model gradually incentivized response rates

incrementally, hospital readiness to collect these data may have contributed to selective PRO reporting.

- One TEP member expressed concerns regarding comparing the complication measure with the PRO-PM, noting the significant time difference of PRO data collection at one year. PROs are highly correlated with the expectations of the patient and matching expectations with the provider that may not be reflected in the complication results. They also noted the 1-day stays nor the outpatient THA/TKA are not captured in either measure results.
- TEP members suggested alternative measures for future validity testing. Two TEP members suggested consideration of length of stay and another TEP member suggested consideration of discharge to skilled nursing facilities for future validity testing.
- One TEP member asked if other disciplines such as physical therapy report THA/TKA measures. The TEP member suggested outpatient utilization as a potential measure and exploring physical therapy measures used by APTA.
    - CORE noted they were unaware of any measures and asked for group input.
    - CORE thanked the TEP member for their suggestion and will explore measures supported by the APTA.
- One TEP member asked about the RSCR categories and if there was a way to parse out the types of complications pertinent to THA/TKA from medically related one that would improve over time.
    - CORE noted there is a threshold of severity in the complication measure outcomes. A patient must be readmitted and have a debridement or go to the operating room to be captured as a significant surgical bleed or a significant infection. CORE noted they could potentially look at mechanical complications and prosthetic joint infections.
- The same TEP member suggested separating mechanical complications and infections, as defined in THA/TKA complication measure.
    - CORE thanked the TEP member for their feedback and reiterated the TEP member feedback on the discordance in timing, provider volume, and other suggestions to better correlate with the PRO-PM were helpful suggestions.

Summary of Measure Specifications

- CORE Presented the Measure Specifications to the TEP
    - CORE reviewed the summary of the final measure recommendations. Data sources include the PROMs, Medicare administrative claims data, and additional risk variables.
    - CORE noted the cohort includes Medicare fee-for-service FFS patients aged 65 and older that undergo elective primary THA/TKA and the dataset is limited to inpatient procedures based on the available CJR data.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

- CORE noted the outcome measure is attributed to the clinician or a clinician group, and this approach is based on the algorithm that is used by the existing clinician- and clinician group-level THA/TKA complication measure with a refinement to limit to clinicians that are responsible for the procedure.
- CORE noted two PROMs are used in the measure, the HOOS,JR. and KOOS,JR. assess pain and function both preoperatively (up to 90 days) and postoperatively (10-14 months) the THA/TKA procedures. TEP members supported the shifted postoperative timeframe from the initial 9-12 months.
- CORE noted the patient-level outcome is defined as the substantial clinical benefit (SCB), which is a threshold of 22-point improvement for hip patients using the HOOS, JR and 20-point improvement for knee patients using the KOOS, JR.
- CORE Presented the Approaches to Risk Adjustment to the TEP
- CORE reviewed the approaches to risk adjustment. The clinically derived and empirically tested risk model included 19 factors such as health literacy, back pain, pain in non-operative lower extremity joint, and the baseline PROMIS Global-10 Mental Health subscale score.
- CORE used the approach of stabilized inverse probability weighting to account for potential non-response bias.
- CORE noted the ongoing reevaluation and surveillance items for future investigation include retesting the risk model prior to mandatory measure use and continued investigation on the relationship between social risk, response rates to surveys, and overall measure results over time. Also, CORE will assess the impact of including procedures performed in the outpatient and ASC settings.

- TEP Feedback
  - One TEP member suggested limiting clinician reporting burden by only requiring reporting of the mental health subscale score instead of the entire PROM. They also suggested incentivizing providers to report PROM data through the American Joint Replacement Registry (AJRR), noting some hospitals already participate in AJRR to qualify for incentives such as the Joint Commission (TJC) Center of Excellence status. Participation in a joint registry database could help inform potential risk adjustment variables to refine and improve the measures.
    - CORE will share these suggestions about reporting burden and reporting through AJRR for the clinician-level measure with CORE's Strategic Planning Implementation team and CMS.

Face Validity of Measure

- CORE Presented the Face Validity of Measure Questions to the TEP
  - CORE noted the TEP member responses for two questions would be included in the NQF forms and measure methodology report.
  - CORE asked TEP members to rate their level of agreement on a six-point scale with the following questions.

1. The clinician-and clinician group-level hip/knee PRO-PM, as specified, will provide a valid assessment of improvement and functional status and pain following elective primary hip or knee replacements.
2. The clinician- and clinician group-level hip/knee PRO-PM, as specified, can be used to distinguish between better and worse quality care among clinicians and clinician groups.

- Summary
    - TEP members supported the measure developers' approach to measure score validity analyses and results and the measure specifications. TEP members supported continuous monitoring of measure score validity testing. TEP members provided feedback about reducing provider reporting burden and offered ideas for additional approaches to measure score validity testing in the future.
    - Among the 17 out of 21 TEP members who responded, face validity survey results were as follows:
        - The clinician-and clinician group-level hip/knee PRO-PM, as specified, will provide a valid assessment of improvement and functional status and pain following elective primary hip or knee replacements.
            - Strongly agree: 7
            - Moderately agree: 6
            - Somewhat agree: 4
            - Somewhat disagree: 0
            - Moderately disagree: 0
            - Strongly disagree: 0
        - The clinician and clinician-group level THA/TKA PRO-PM as specified can be used to distinguish between better and worse quality care among clinicians and clinician groups.
            - Strongly agree: 3
            - Moderately agree: 6
            - Somewhat agree: 6
            - Somewhat disagree: 2
            - Moderately disagree: 0
            - Strongly disagree: 0

Next Steps

Ongoing Measure Development

CORE will continue to encourage further feedback and questions from TEP members via email. Additionally, CORE will continue to engage stakeholders in a Clinical Working Group and will hold a public comment to solicit feedback on measure specifications.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

Conclusion

TEP feedback of CORE's approach to measure development helped inform the development of measure specifications and ongoing reevaluation items. CORE will continue to engage with the TEP as the measure moves through measure endorsement and implementation planning.

Appendix A. CORE Measure Development Team

Center for Outcomes Research and Evaluation (CORE) Team Members

Name	Role
Rachelle Zribi, BA	Project Lead
Kerry McDowell, MS, MPhil	Project Manager
Emma Turchick, MPH	Research Support
Sheng Zhou, MD, ScM	Lead Analyst
Kyaw (Joe) Sint, PhD, MPH	Supporting Analyst
Shani Legore, BA	Person and Family Engagement Communication Specialist
Kathleen Balestracci, PhD, MSW	Division Lead
Zhenqiu Lin, PhD	Director, Data Management and Analytics
Lisa Suter, MD, PhD	Contract Director, Quality Measurement Program

Appendix B. TEP Call Schedule

TEP Meeting #1

Wednesday, August 26, 2020 – 5:00-7:00PM EST (Zoom Teleconference)

TEP Meeting #2

Wednesday, February 10, 2021 – 5:00-6:30PM ET (Zoom Teleconference)

TEP Meeting #3

Tuesday, March 23, 2021 – 5:00-7:00PM ET (Zoom Teleconference)

TEP Meeting #4

Tuesday, July 13, 2021 – 5:30-7:00 PM ET (Zoom Teleconference)

Appendix C. Summary of Questions Received Prior to TEP Meeting #1

TEP members received detailed meeting materials prior to the first TEP meeting. Provided below are questions submitted via email from a TEP member, as well as the responses provided by CORE, prior to the first TEP meeting on August 26, 2020.

1. Data continues to demonstrate that socioeconomic status (SES) risk factors are important. There is a potential for SES driven discrepancies as to which patients undergo procedures in outpatient versus inpatient status. Has this been analyzed?

   Social risk is a very important issue for quality measures, especially for measures that evaluate patients undergoing elective procedures such as THA and TKA. There are many potential approaches for addressing social risk in value-based payment programs, including risk adjustment of the quality metrics and stratification of either the quality metrics and/or payments. The Assistant Secretary for Planning and Evaluation (ASPE) has recently released guidance on this topic and CMS has prioritized reducing disparities in its measurement programs. For example, the Hospital Readmission Reduction Program (HRRP) now applies payment incentives within groups of hospitals categorized by their proportion of dually Medicare and Medicaid eligible patients. Dual eligibility is a potent marker of social risk and this approach ensures that hospitals with more complicated patients, such as dual eligible patients, are not financially penalized because their patients are more socially and economically disadvantaged. We will investigate the impact of social risk in detail when we talk about risk adjustment and look forward to the TEP's input on how best to ensure that quality measures do not result in negative consequences such as reduced access to care or worsening disparities. While CORE does not make decisions on how CMS implements these quality measures, all of the TEP's input is shared with CMS and will inform CMS' future implementation planning.
2. The PROM-PM data was collected at a hospital level with incentivization through the CJR. Is the data as reported by the hospital to be used for the surgeons?

   The development and testing data for the surgeon-level PRO-PM will use the data collected through CJR. These data were readily available for testing and no national surgeon-level dataset containing PRO data is currently available. Future implementation plans have not been finalized by CMS.
3. Will this be a voluntary Merit-based Incentive Payment System (MIPS) measure? If it is not voluntary, what are the minimum collection rates? How does the measure protect against selective reporting?

   These are all very important questions. CMS is funding the surgeon-level PRO-PM work through money allocated for the QPP, suggesting their intention is to eventually implement it in QPP. However, no specific implementation plans have been shared with CORE for either the hospital- or surgeon-level PRO-PMs. The TEP's input on this will be critical to CMS' planning. Other measures that involved novel data collection, such as the hybrid hospital-wide readmission measure that

combines clinical vital sign and laboratory data with administrative claims data, are being implemented gradually, starting with voluntary submission and measurement.

4. Why is the PROM-PM not being applied to the hospitals (or is it)? The data captured to provide the testing of this measure was through hospital incentivization, not surgeon.

   As noted above, there are no specific plans for either measure yet and no plans to hold hospitals in CJR accountable for their measure results. The proposed CJR model extension has extended the PRO data collection but has not added any plans to implement the PRO-PM in the model.

5. When was the final version of the PROM-PM presented to the TEP before undergoing National Quality Forum (NQF) endorsement? Is it to be applied at the hospital level? Has that rule proposal been submitted?

   As noted above, there are no specific plans for either measure yet – the final measure specifications for the Hospital-level THA/TKA PRO-PM measure was shared with the TEP on January 29, 2020. That measure passed the Scientific Methods Panel and NQF Committee and is now in a NQF Public Comment period.

6. There is an economy of scale in terms of capturing the PROMs. The collection of these through incentivization to the hospitals through the CJR relies on the application of collection efforts that are beyond a single surgeon or group. The surgeons being required to capture these will face variable resources depending on the size of the group and or hospital employment. Any consideration to this factor?

   This input is important for CMS to hear as they consider implementation planning for PRO-PMs.

7. Will either of these measures be used for surgeon specific NPIs (National Provider Identification) versus group TINs (Taxpayer Identification Number)? The latter spread risk variably depending on size. If it is going to rely on TINs, how does the measure work in very large multispecialty groups numbering in the thousands?

   In general, the surgeon-level measures can be reported at the level of the individual clinician (a unique TIN-NPI combination) or at the level of the clinician group (a unique TIN). This means that very large multispecialty groups could have results reported for the THA/TKA complication measure if they perform enough THA/TKA procedures. In such cases, the group receives a measure result, even if it only represents the care provided by a subset of their clinicians.

Appendix D. Detailed Summary of TEP Meeting #1

Orthopedic Technical Expert Panel (TEP) Meeting #1 Minutes

Wednesday, August 26, 2020 5:00-7:00 PM ET

Participants:

- Technical Expert Panel (TEP): Phyllis Bass, Vinod Dasa, Rachel Dupré Brodie, Cheryl Fahlman, William Hamilton, Cynthia S. Jacelon, Benita Lattimore, Craig Miller, Michael H. Perskin, Christine Von Raesfeld, Adam Schwartz, Robert Sterling, Margaret A. Vanamringe, Patricia Walker, Kevin Woodward, Adolph J. Yates
- Yale New Haven Health Services Corporation- Centers for Outcomes Research and Evaluation (CORE): Andrea Barbo, Kathleen Balestracci, Susannah Bernheim, Jacqueline Grady, Andreina Jimenez, Miriam Katz, Shani Legore, Yixin Li, Fior Rodriguez, Lisa Suter, Kyaw Sint, Lori Wallace, Sheng Zhou, Rachelle Zribi
- Expert Clinical Consultant: Kevin Bozic

Executive Summary

- The purpose of the first TEP meeting was to educate the TEP on the background and approach to developing the EPM THA/TKA Complication Measure and Clinician-Level and Clinician-Group Level THA/TKA PRO-PM. The TEP was invited to provide input on the measure concepts and approaches to re-specifications.
- The TEP shared several considerations for both measures.
    - EPM THA/TKA Complication Measure:
        - TEP members noted concern that the measure does not include claims for procedures performed in ambulatory surgery centers (ASCs). TEP members discussed the EPM THA/TKA Complication Measure outcome and provided recommendations for complications that should be considered in the measure outcome.
    - Clinician-Level and Clinician-Group Level THA/TKA PRO-PM Measure:
        - Measure Implementation: TEP members noted the importance of a national data collection and submission mechanism for the measure to be successful. TEP members noted the importance of incentivizing adoption of collecting PROs and recommended a phased approach to allow practices of all sizes/locations, including small, rural, or low resources practices, to build the capacity to collect and submit PROMs.
        - Data Collection Timeframe: TEP members recommended consideration of allowing for a longer post-operative follow up timeframe.
        - Clinical Settings: TEP Members noted that the measure should consider procedures performed in ASCs and HOPDs.

TEP Action Items:

- TEP members were invited to email cmsorthopedicmeasures@yale.edu with any additional comments and suggestions. Members were asked to review and comment on the meeting summary.

CORE Action Items:

- Immediate next steps: The development team will continue measure development and testing activities, with consideration of specific issues raised by the TEP.
- The team will convene the next TEP meeting by webinar, mostly likely in the Fall or Winter of 2020.

Detailed Discussion Summary

Welcome

- Ms. Andreina Jimenez welcomed the group on behalf of CORE. She reminded the group that the purpose of bringing together the TEP is for the development and re-evaluation of two orthopedic measures. She noted that the minutes and summary report will be distributed following the meeting.
- Ms. Jimenez reviewed the meeting agenda and reminded the group that the content of TEP discussions must remain confidential until made public by the Center for Medicare & Medicaid Services (CMS) and that all personal opinions and experiences, including personal health information, shared during TEP meetings are to remain confidential. Ms. Jimenez stated that TEP members represent themselves and not the organizations with which they are affiliated. She noted that the work is funded by CMS and that CMS, the Center for Medicare & Medicaid Innovation (CMMI), or Quality Payment Program (QPP) members may sit in on these calls.
- Ms. Jimenez provided a brief description of CORE and its measure development work.

Introductions

- Ms. Jimenez introduced the EPM THA/TKA Complication Measure team members.
- Ms. Fior Rodriguez introduced the Clinician-Level and Clinician-Group Level THA/TKA PRO-PM team members.
- Dr. Lisa Suter provided an introduction and thanked members of the TEP for joining the call. The purpose of this meeting is to provide foundational information about the two measures that CORE will re-specify. CORE's goal is to create a partnership with the TEP over time with the aim of engaging this TEP with many different measures. Some questions TEP members have regarding these measures may be out of scope for today, for example social risk and implementation questions, but these could be pertinent in future conversations. There may not be full consensus on the topics discussed, but CORE is eager to ensure all voices are heard and all perspectives are respected. CMS reviews the summary report and posts it publicly following these meetings for maximum transparency.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

- TEP members briefly introduced themselves and described their key interests or experiences related to these measures. Members disclosed any potential conflict of interest (COI).
- Dr. Kevin Bozic, a member of the clinical workgroup and consultant working with CORE for 12 years in performance measure initiatives development, introduced himself.

Review and Approval of the TEP Charter

- Ms. Rodriguez facilitated the review and approval of the TEP charter. Members agreed there were no concerns and the charter was unanimously ratified and approved.

Measure Background: Current Orthopedic Measures

- Ms. Rodriguez presented the current orthopedic measures. She noted where the orthopedic measures fit into the reporting and payment programs.
- Ms. Rodriguez reiterated the focus of the discussion would be the Clinician-Level and Clinician Group-Level Total Hip Arthroplasty/Total Knee Arthroplasty (THA/TKA) Patient-Reported Outcome-based Performance Measure (PRO-PM) and the 90-Day Risk-Standardized Complication Measure Following Elective THA and/or TKA for a Potential Combined Inpatient and Outpatient EPM.
- Ms. Rodriguez noted the EPM 90-day THA/TKA Complication Measure is focused on hospital performance and tied to hospital payment. The measure specifications are based on the existing Hospital-level THA/TKA Complication Measure. The Clinician-Level and Clinician Group-Level THA/TKA PRO-PM is focused on clinician and clinician groups using patient-reported outcome data. The QPP program is tied to clinician payment and the measure specifications are based on the existing Hospital-level THA/TKA PRO-PM.

Measure Overview: EPM 90-day THA/TKA Complication

- Dr. Lori Wallace, the project lead, welcomed the group and indicated she would be providing a high-level overview of the EPM 90-day THA/TKA Complication Measure. She noted that this will be referred to as the EPM measure, and that the presentation would review the existing measure, provide the timeline for measure development, and list potential topics and questions for future discussion.
- Dr. Wallace noted that the purpose of the measure is a re-specification of the existing inpatient hip/knee replacement measure for a combined inpatient and outpatient CMMI Episode Payment Model (EPM). CMMI supports innovative payment models. The rationale for expanding the measure is an increase in TKA procedures in the outpatient setting, indicating that this setting should be assessed in order to accurately capture the quality of care. The existing Hospital-level THA/TKA Complications Measure will be referred to interchangeably as the inpatient or hospital measure.
- The purpose of the original measure was to identify the medical and surgical complications that could be attributable to the care provided during and after an elective total hip or total knee arthroplasty procedure. The outcome is a dichotomous yes or no assessment of whether a complication occurred during an index admission for

the elective procedure or if a complication occurred during a readmission within the specified time for that complication.

- Dr. Wallace described the components of the measures, focusing on those areas that require refinement. The cohort setting, the patients included in the measure, is being expanded in the new measure to include patients with procedures performed in outpatient departments. The cohort definition, the patients eligible to be included in the measure, will remain the same in both measures. These patients are Medicare beneficiaries aged 65 and older who have had a qualifying elective primary THA/TKA during index admissions. This excludes fracture, bony metastases or partial or revision THA/TKA procedures. Patients must be enrolled in Medicare FFS Part A during the index admission and enrolled in Parts A and B for 12 months prior to the admission date.
- Dr. Wallace described the expansion of the new measure to include complications that occur during the index procedures in the hospital or outpatient departments as well as complications that occur during emergency department visits and observation stays. The current codes to identify complications are ICD-10-Clinical Modification (CM), Procedure Coding System (PCS) and Present on Admission (POA) codes. The new measure will expand the type of procedure codes to include the CPT and HCPCS codes used in outpatient billing. POA is not used in outpatient claims.
- Dr. Wallace described the outcome definition for those conditions that are included in the complication outcome. These are the same for both the current and new measures. Clinically significant outcomes are those attributed to the THA or TKA procedure and identifiable using claims data. These complication outcomes were clinically vetted during the development and reevaluation of the original hospital measure, which has been in use by CMS since 2013. In order to identify if a complication occurred during the procedure, there are two overarching questions which are as follows:
    - Did the condition or event occur?
    - Did it occur within the specified timeframe?
- The algorithm for both measures indicates it is considered a complication if any of the following occur:
    - An acute myocardial infarction (AMI), pneumonia or other acute respiratory complication, sepsis or shock occur during the index admission or subsequent inpatient admission within 7 days from the start of the index admission.
    - A pulmonary embolism, surgical site bleeding or other surgical site complication, or death occurs during the index admission or subsequent inpatient admission within 30 days from the start of the index admission.
    - A mechanical complication, periprosthetic joint infection or wound infection occurs during the index admission or subsequent admission within 90 days from the start of the index admission.
- The EPM measure is also considering observation and emergency department visits as potential settings.
- Dr. Wallace walked through the timeline for this measure and noted that many activities will happen concurrently. The schedule begins with this first TEP meeting. In the fall, winter, and into the spring of 2021, measure testing will be conducted, which involves

running statistical analyses to test measure validity. CORE will continue with stakeholder engagement in the fall of 2020. CORE will hold meetings with the TEP and the Clinical Working Group at various points throughout measure development, which would extend through the winter of 2021. CORE aims to finalize this measure and obtain public comment feedback by the spring or winter of 2022.

- Dr. Wallace presented future topics and anticipated questions for the group. CORE will seek to confirm that outpatient THA/TKA procedures were adequately captured. CORE will request TEP members to review and provide feedback on complications captured during emergency room visits and observation stays. A consideration for the group is how to accurately identify complications in the outpatient setting in the absence of POA coding. CORE asks that the TEP members help define the measure outcome algorithm for both the inpatient and outpatient setting.
- TEP members had the following questions and comments regarding the measure:
    - A TEP member asked why the measure did not include cases from freestanding ambulatory surgical centers (ASCs). With many surgeries now being performed in that setting, a fair number of cases would be missed.
        - Dr. Wallace noted that the CMS quality measures are aligned with payment programs and models. ASCs have their own payment programs. CORE can ask CMMI if they have plans to include ASCs in future models.
    - A TEP member asked why POA codes were removed in the outpatient setting.
        - Dr. Wallace indicated she was unsure of the history of POA codes being included in the outpatient realm or captured in outpatient claims.
        - Dr. Suter noted that CPT codes do not contain modifiers that allow providers to indicate present issues the same way inpatient codes do. This is likely due to the extended period of an inpatient encounter where any onset issue is more critical. She recommended reviewing the document circulated earlier by email in response to questions by a TEP member. When the first complication measure for hospitals was developed, POA codes existed but were not being used. CORE worked with clinicians to create an algorithm to identify potential complications of care. For example, when a pneumonia code is present during the index admission where the procedure was performed with no history of pneumonia in the previous 12 months, this was attributed to complication of care as opposed to a risk-adjustment. For this measure, it is not a prolonged encounter and the procedure is performed on relatively healthy people. Coding for issues such as heart attack, diabetes or pneumonia in an outpatient setting for an outpatient elective procedure reflects the health history of a patient. TEP members should keep this approach in mind to accurately capture the patient's clinical history. Although it is frustrating not to have granularity in the codes, it is something to work around.

- A TEP member inquired about the existence of a master document to see the granularity of the definitions of the complications, noting these definitions vary even amongst hospitals.
    - Dr. Wallace responded that CORE could share the original methodology reports and specification and update reports, which provide some information.
    - Dr. Suter added that CORE can provide more detail. CORE has well-documented ICD-10 codes for complications in the inpatient setting. She noted that the role of the TEP is to ensure that the translation of this measure to the outpatient setting is valid. The goal is to create measures that ultimately incentivize improvement. CORE does not want to create a measure that incentivizes poor behavior or negatively affects clinical practice. There will be areas of tradeoff. These are high stakes measures that are likely to be implemented in a future payment model. The team will discuss the tradeoffs and CORE will gather member feedback for what feels most appropriate to minimize harm.
- A TEP member asked about the volume of inpatient versus outpatient procedures and if the intent for reporting on measures is to separate those that occur in these two settings.
    - Dr. Wallace replied that the group will address this in more detail in future meetings and offline. The discussion may circle back to this later in the call if time permits.
- A TEP member asked whether surgical site infections count as complications under this outcome definition.
    - Dr. Wallace confirmed that these are included under the periprosthetic joint infection and wound infection definition and would require both a diagnosis and prognosis code.

Measure Overview: Clinician-Level and Clinician-Group Level THA/TKA PRO-PM

- Ms. Rachelle Zribi, the project lead for this measure, presented the overview of this measure.
- CMS contracted CORE to re-specify the existing Hospital-level THA/TKA PRO-PM as a clinician or clinician group measure for the Quality Payment Program (QPP). CORE will adapt the current hospital-level measure to be applicable to clinician and clinician groups. This is CORE's third re-specification project aimed at re-adapting hospital-level measures to be applicable to clinicians and clinician groups.
- The QPP was created in 2015 and transformed the Medicare clinician payment system from FFS to Pay for Performance. Participants receive an overall score that includes quality measures such as process measures, outcome measures, and experience measures.
- Ms. Zribi reviewed terminology related to this work. Patient Reported Outcomes (PROs) describe patient-reported concepts such as pain or function. Patient Reported Outcome Measures (PROMs) are the instruments that capture PROs. A Patient Reported

Outcome-based Performance Measure (PRO-PM) is the performance measure that uses PRO data to define the measure outcome.

- Ms. Zribi described the history of the existing Hospital-level THA/TKA PRO-PM. The Hospital-level THA/TKA PRO-PM began measure development in 2013. The measure was specified and tested with input from patients, providers, and clinical experts. It passed endorsement by the National Quality Forum (NQF) in 2020 and is currently undergoing public comment.
- Ms. Zribi described the existing Hospital-level THA/TKA PRO-PM specifications and noted that they will be discussed in greater detail in the future. The current Hospital-level THA/TKA PRO-PM uses two PROMs. For hip patients, the PROM is the HOOS, JR, which is a survey that consists of 6 questions on pain and function. For knee patients, there is the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR), a survey with 7 questions on pain and function. These surveys are psychometrically valid and were selected with length and overall burden in mind.
- The data sources for the PRO-PM are the PROM data, administrative claims data, and additional risk variable data. The cohort is Medicare FFS beneficiaries ages 65 and older who undergo elective, non-urgent surgeries. Pre-operative PRO data is collected 90 to 0 days before the procedure and post-operative PRO data is collected 270 to 365 days after the procedure. Both the pre-operative and post-operative PROM collection windows allow some flexibility in data collection. The post-operative timeframe aligns with existing follow up appointments and allows enough time for patient recovery.
- Ms. Zribi described the risk-adjustment process, which accounts for varying patient case mix across entities such as hospitals. The hospital-level measure team developed a clinically-derived risk model with 19 variables including health literacy, back pain, pain in non-operative lower extremity joint, and the baseline PROMIS Global Mental Health subscale score. PRO-PMs rely on novel data collection, so PRO-PMs need to consider response bias. The hospital measure conducted analyses to address potential non-response bias.
- The patient-level outcomes were defined using both patient input and empirical evidence. The outcome definition for THA is whether the patient meets or exceeds the substantial clinical benefit (SCB) threshold, defined as an increase of 22 points on the HOOS, JR, from their pre-operative to post-operative PROM assessment. Similarly, the outcome definition for TKA is whether the patient meets or exceeds an SCB threshold of 20 points on the KOOS, JR from their pre-operative to post-operative PROM assessment. The hospital-level outcome is the risk-standardized proportion of patients undergoing elective primary THA/TKA who meet or exceed the SCB thresholds.
- The goal of the Clinician-Level and Clinician Group-Level PRO-PM is to capture the full spectrum of care and incentivize quality. Patients have expressed a desire to have measure results that reflect physician level performance. CORE will develop and test the measure using data from the Hospital-level PRO-PM development, specifically the CMMI Comprehensive Care for Joint Replacement (CJR) voluntary data collection. CORE will solicit input from the TEP, Clinical Working Group, and Patient Working Group. In the future, there will be a public comment period.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

- Ms. Zribi noted potential future discussion topics. For the clinician and clinical group attribution methodology, CORE proposes using the approach that was developed for the QPP THA/TKA Complication Measure. For risk adjustment, CORE proposes using the clinically-derived risk model that was developed for the Hospital-level THA/TKA PRO-PM. CORE will analyze non-response and incomplete PRO data using the approach developed by the Hospital-level THA/TKA PRO-PM. CORE will investigate future inclusion of the outpatient procedures in the measure cohort when data are available.
- Ms. Zribi presented the measure development timeline for this measure, which is similar to the timeline for the EPM measure. This first TEP meeting occurred in August and CORE will continue measure development and testing through the winter of 2020. CORE aims to meet with stakeholder groups throughout the development process, which will continue through the summer of 2021.
- TEP members had the following questions and comments in response:
    - A TEP member asked how the PROM data required for this measure would be collected at the national scale, particularly the follow-up PROM data one year later when many patients may not return to their physician. The member noted that without a mechanism to capture this information, many clinicians may not collect the data.
        - Ms. Zribi thanked the TEP member for their question and responded that there is no current implementation plan for this measure. CORE is interested in gathering the TEP feedback on how best to incentivize clinicians to capture PROM data and what an ideal mechanism would be to allow for high response rates and have a low burden.
    - The TEP member noted that capturing PROMs is aspirational and even with dedicated efforts, they have seen poor response rates. The TEP member noted that electronic methods have increased their research institute's patient responses for the HOOS, JR and KOOS, JR, and highlighted that rural and small hospitals may not have any resources to implement this. The TEP member noted that national implementation using a database, such as the American Joint Replacement Registry (AJRR), may be an option. The TEP member reiterated that implementation is an important consideration before designing the measure.
        - Another TEP member agreed with the importance of considering measure implementation and shared that joint surgeons are motivated to collect data on their patients, but it is challenging to achieve high responses. The TEP member noted that it would be equally important to incentivize patient response.
        - Dr. Suter shared that CMS has embarked on a mission to move to digital quality measures, with the goal of 100% digital quality measurement by 2030. Historically, electronic health records (EHRs) were developed first for billing, then for clinical care, and quality measurement was a lower priority. There is parallel work within CMS to innovate digital measurement; though it is not specifically geared towards PROs or orthopedic measures, it will inform this measure. CMS aims to collect

information from a range of sources and integrate this at the point of care for integration with clinical decision making and benchmarking. The hospital measure was developed at the beginning of working towards EHR measures. In addition, the CJR model did not restrict the mode of data collection nor how clinicians communicated the PROM scores to patients and integrated data collection efforts into clinical workflow. This group can help CMS learn from the institutions capturing PROM data well and encourage institutions to use these best practices.

- Dr. Suter agreed that patient-level data collection, such as PROs, are dependent on hospital and clinician resources. However, it is difficult for CMS to incentivize this work until there are quality measures. Although technology allows PRO data collection to be done, not everyone in the country can invest resources in those technologies. The TEP can highlight these issues and share ways that CMS can implement a measure to move the field forward. For example, CMS has implemented hybrid measures for a voluntary reporting period combining claims data and electronic clinical data for a single quality measure. Therefore, there is a precedent for implementing novel measures and learning from them without penalty.
- Dr. Suter further noted that non-response is an issue for all PRO-PMs and there will never be an expectation for 100% response rates. We believe there are reasonable targets to reach. HCAHPS currently has 10-15% response rates and their response rates have declined over time for many reasons. This measure will assess the impact of non-response on the measure and follow the current Hospital-level measure approach to accounting for non-response. In the future, some of these challenges may be addressed by removal of clinicians from public reporting if they do not meet a certain response rate threshold or stratification by peer groups serving the same patient groups. Dr. Suter noted that it is possible that this measure may not move forward until it is electronically specified, and CMS establishes all electronic standards in the future, and TEP feedback can help CMS progress towards that.
    - A TEP member acknowledged from the perspective of a CJR-participating hospital that although it is challenging to capture PROs, they have been participating in the voluntary data collection effort for 5 years and, along with other CJR participant hospitals, have successfully captured hundreds of thousands of outcomes that allowed for development of the hospital-level PRO-PM. The TEP member noted that their hospital utilized the mandatory bundle to incentivize the data collection. The TEP member noted that an economy of scale must exist for that type of data collection at the hospital or hospital system level for the measure to succeed. There is the expectation of CMS to move to more universal alternative payment models and it is possible that CJR would expand to a national scale. At that point, hospitals would need to capture PROs, so there will be dual-incentivization for surgeons to capture information electronically.

However, if not all hospitals are participating in an alternative payment model with that built in incentivization, it will be difficult for PROs to be captured nationally. The TEP member also noted that, regarding a surgeon-specific measure, many surgeons are performing THA/TKAs in ASCs. Therefore, even with alternative payment models imposed on hospitals, the ASCs are a different environment in terms of economies of scale and the setting where ultimately these measures will need to be captured.

- A TEP member shared that it would be beneficial if CMS provided larger financial incentives to support capacity building, data infrastructure, and workflows to report on these measures. The member noted that many CJR-eligible hospitals decided to not invest in the voluntary data submission because the incentive was not large enough. The TEP member noted that their organization promoted a phased approach to submit PROs over 3 to 4 years to first incentivize the baseline measurement and then, as hospitals became experienced, funding was incentivized for the follow up PRO measurement. The TEP member suggested CMS support IT infrastructure to allow for the capture of these data or the AJRR existing infrastructure, especially for smaller hospitals or those without a high level of experience. The TEP member also commented that clinician buy-in would increase if the measures are useful in the clinical setting. The TEP member also noted that CMS incentivization of practice transformation and measurement-based care may help providers collect PROMs.

- A TEP member raised consideration about the SCB thresholds and the post-operative PROM collection window. The TEP member asked if the research indicates clinically significant thresholds could be applied at an earlier timeframe; for example, if a 15-point increase between 6-9 months could be acceptable.
    - Dr. Suter clarified if the question was that shorter post-op timeframe would garner higher response rates.
    - The TEP member agreed and questioned if low follow up response rates were due to the fact that some patients are not seen during the 9-12-month timeframe. The TEP member questioned if the measure had a shorter follow up timeframe, whether the response rate would increase and still be a valid measure. The TEP member questioned if evidence existed to show an increase in scores from 6-9 months equates to what would have been seen for the 9-12-month scores.
    - Dr. Suter commented that these questions are important for aligning quality measures with clinical practice due to the lack of uniform clinical practice around the country. The 9-12-month post-operative PROM data collection timeframe was chosen because clinicians wanted the post-operative data collection window to extend far enough out from the

procedure to reflect adequate recovery. Dr. Suter explained that during the development of the hospital-level measure, there were discussions that hip replacement patients recover faster than knee replacement patients, so a longer timeframe might be required for TKAs. CORE has heard feedback from some physicians requesting an extension on the data collection beyond 12 months to broaden the post-operative PROM data capture. Dr. Suter noted that extending the timeframe allows more flexibility for that one-year follow-up, but some patients have high social risk factors, or may change physical location, making the extension difficult.

- A TEP member inquired, based on recollection from a previous TEP meeting, whether 365 days was used as a cut-off because of the inability of CMS logistics to capture data beyond that time.
    - Dr. Suter replied that the timeframe was selected based on preference rather than a logistical barrier. Dr. Suter noted that it is an innovative measure and experience from CJR demonstrated that more flexibility in the post-operative PRO data collection window may be needed.
    - Dr. Bozic noted that previous evidence was presented to the TEP indicating there was a difference between when scores peaked and leveled off for hip and knee replacement patients. The evidence showed that the scores leveled off between 6-9 months but the TEP decided to use the timeframe of 12 months to capture all potential change. Dr. Bozic noted that because patients do not come back for follow ups at exactly one year, he suggested centering the window around a year but allow flexibility on both ends of the window.
    - Another TEP member agreed and recommended expanding the timeframe to allow for two months on either side.

- A TEP member asked about CORE's expectation, given that HOPD is now a focus, if the outcomes for outpatient and inpatient would be measured similarly. For example, if a patient is sent home at 27 hours versus 18 hours, is there a clinical difference?
    - Dr. Suter noted that CMS does not reimburse for procedures that are not full inpatient stays. The clinical practice evolution and improvements in clinical care and perioperative protocols have made it reasonable to perform procedures in the outpatient setting. The change in CMS reimbursement is based on these changes in the level of clinical work. Quality programs are more restrictive so there are no measures that cross the different settings in those payment programs. CMMI aims to be more comprehensive and flexible by creating a payment model that crosses these settings. Similar to CJR, it is possible there may be forgiveness or exclusion from other payment models. Many clinical

events or procedures do occur in multiple settings. This is one of the first elective procedures that considers cross setting measurement and will offer insight for CMS to develop a multiple setting payment program. That point relates back to the earlier question as to whether CMS would separate inpatient and outpatient measurement. This is an open question. Fundamentally, QPP is the one area with some flexibility because it assigns who is held responsible for the measure result, such as the clinician or clinician group, rather than detail the setting of the event or encounter. Hospital measurement programs are very clear in what is considered inpatient versus outpatient and they do not have measures that cross those settings. The TEP's responsibility is to flag the unintended consequences and areas of concern when moving across settings. For elective hip and knee procedures, is this going to reduce access? Will healthy patients be seen outpatient and very sick patients be admitted for an inpatient stay? If that is the case, the TEP may need to separate the measurements because it would be inappropriate to combine these different populations. More understanding must be gained before that decision is made.

        - TEP members indicated that this point was helpful. The TEP could find that the inpatient measurement is no different than the outpatient measurement and the measure can transcend location, or these two measurements are different, and the group will need to rework the modeling for the measures.

- A TEP member pointed out that if CMMI proceeds with the current rule proposal for Performance Year 6 for CJR, the data will combine the HOPD and inpatient THA and TKA. There will be data coming in with PROs that include outpatient and inpatient procedures up to 20-25% of cases through outpatient HCPCS billing. The denominator definition must be modified in order to find those HCPCS codes, which have historically been Diagnosis-related Group (DRG) 469 or 470. If implemented, there is a need to consider the cost and develop a quality measure. The complications measure would follow, then ideally the PRO, and finally the harmonization across ASC, HOPD, and inpatient so there are similar measures in these settings. This mix of outpatient and inpatient ensures that all Medicare patients getting hip and knee replacements are captured.
- A TEP member inquired for those in the bundle whether there was any major shift seen in ratio of DRG 469 and 470 in inpatient cases and if there were lower amounts of complex patients performed in the bundled setting.
    - A TEP member noted based on their experience, it depends on how aggressively a hospital is using DRG 469. Hospital coding blogging site experts advise hospitals should run around 9% for DRG 469, but the actual CMS experience is around 6%. Their institution is conservative regarding what is considered DRG 469 and are at around 3%. Many of those 3% are actually hip fractures within CJR and thus sit in a separate

category. Overall, no matter how aggressively a hospital uses DRG 469, it is a small percentage. It is relevant to look for any significant change in comorbidities or risk of mortality. DRG 469 seems arbitrary upon admission and is generated for an inpatient complication during the stay but does not necessarily have a risk appearance for a patient going into operation.

- Dr. Wallace revisited the question regarding the volume of cases for TKAs. She provided the volume for inpatient hip and knee procedures was 296,314 and TKA outpatient procedures was 96,006, which includes observation stays. These numbers do not account for cohort exclusions.
    - A TEP member speculated that with trend data it would show movement to the outpatient setting.
    - Dr. Wallace agrees that we can anticipate that the outpatient numbers will increase.
    - A TEP member noted something billed as an outpatient case could actually have been admitted under observation for a one-night stay. Hospitals have been risk averse to Quality Improvement Organizations (QIO) and rack audits for patients going home, so these are not truly outpatient. This is about 25% of those cases. Parts of the country using CJR drop to about 15%. People respond in an economically appropriate fashion by keeping healthy patients in the CJR and not going outside for the procedure, which would generate a DRG.
- Dr. Suter noted that COVID-19 may accelerate hip and knee procedures being performed in the outpatient setting. Even though the COVID-19 infection rate is low in Connecticut, it is still understandably difficult to get patients to come to the hospital setting for care at this time.
- A TEP member noted that the way a claim is coded does not always represent how a patient was treated. The TEP must consider separating the claims as it affects the appearance of the results. Outpatient numbers may appear better, but that may not be the case if there is a coding issue on the claim.

- A TEP member noted the inclusion of complications such as acute myocardial infarction (AMI), pneumonia, sepsis, and mechanical complications for the EPM measure. Their patients experience other complications in addition to these. The member challenged why other medical complications including urinary tract infection (UTI), urinary retention, chest pain, and nausea, which are more common, are not included.
    - Dr. Suter noted it was important for this group to consider how to re-specify the current Hospital-level EPM measure. It is currently limited to the complications listed. The measure does not include UTI, constipation or others based on the historical perspective of ensuring claims data are capturing a complication that hits a severity threshold. This does not

diminish other aspects of a patient's post-operative experience as unimportant or potentially life threatening, but due to challenges of the data source, they have been chosen not to be included. Any changes the TEP makes to the list of complications would have implications for all other versions of this measure. This EPM complication measure already exists at the hospital level and at the surgeon and surgical team level. It is possible to move away from the current complications if the group finds that necessary, but the group must consider the data set and the implications across CMS's measures. It was suggested that it may be best to pause making any changes to that list and come back to this in the next iteration of the measure.

- A TEP member noted that because this is the patient-reported outcome measure other functional deficits such as stroke, acute kidney injury or renal failure that lead to prolonged disability are significant. The counter argument is if the loss of function is picked up in another measure. While the other complications such as UTI and constipation lead to a poor patient experience and maybe some disability, these are more difficult.
- A TEP member indicated from previous TEP panel discussions that the current complications were chosen because of their incidence and significance. The member suggested going forward that there may be other things equally important to capture. It was suggested to ask a patient group to rate the importance of these complications as another way to gain patient input.
- A TEP member agreed that incontinence is an important quality of life issue.
- A TEP member commented that complications such as sepsis or AMI may have been more impactful previously when patients stayed longer in the inpatient setting. We may see that things are not as important as practice evolves.
- Another TEP member noted that the data shows more cardiopulmonary complications in outpatient than inpatient settings.
- A TEP member recommended that blood transfusion should also be considered as an addition to the list of complications.

Next Steps

- The CORE team will circulate the meeting minutes and summary report. It is requested that TEP members review these items before they are posted for public comment.
- CORE anticipates holding one or two more of these TEP meetings between September 2020 and March 2021. Those meetings will be measure specific, with one for EPM and one for QPP. Surveys or emails may also be forthcoming in between meetings to request TEP member input.
- On behalf of CORE, Ms. Jimenez thanked the group for their feedback and asked that any additional questions to be emailed to the team.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

Appendix E. Summary of Email Correspondence Before and After TEP Meeting #2

Email Correspondence Before TEP Meeting #2

TEP Member Email

Hello,

I have reviewed the PowerPoint in advance and have two questions/suggestions that are complex, and I might not ask them correctly or fully in the conversation this afternoon.

The c-statistics are surprisingly low. Was there an attempt (or is it methodologically possible) to use the pre-operative HOOS/KOOS JR scores as a risk adjustment using highest quartile scoring or scores above the standard deviation? False expectations of what can be achieved might mute the delta; I know this was previously discussed. As an alternative possible approach, some kind of mental health/higher preop patient score ratio?

I raise the question because of the literature that shows a high proportion of dissatisfied total knee patients having very mild radiographic changes (if any in some cases). Second question: how was BMI captured and applied? As a continuous variable or by class of obesity (versus morbid, versus super-obesity)? Success, although achievable, is reported as muted in the very high ranges.

I have attached what I feel is a good paper regarding SES that adds credence to the lack of added risk adjustment using SES; it came out this past week.

CORE's Response

Thank you for your questions. We provided responses to your questions below.

Using the pre-operative PROM score as a risk adjustment variable

- We have not included pre-operative PROM scores in the risk model because conceptually we want to model improvement (which is defined as meeting a pre-defined threshold of improvement between baseline and follow up), and inclusion of/controlling for pre-operative PROM score in the risk model shifts the model focus to post-operative health status. We decided to measure improvement (instead of post-operative health) to mitigate an unintended consequence of incentivizing performance of procedures on patients that may achieve a high post-operative PROM score over patients with more severity.

Mental health score

- Thank you for this suggestion. The risk model considers a patient's baseline mental health subscale score using the PROMIS-Global or VR-12.

BMI

- Patient BMI is submitted by hospitals in the voluntary reporting of PRO data and risk variable data for CJR. Hospitals can report a patient's BMI, or patient's height and weight which we can use to calculate the BMI. The measure applies BMI as a continuous variable.

SES

- Thank you for this feedback and for sharing this article. We plan to present and discuss with you all our social risk factor approach and analyses during TEP Meeting 3, which will occur in late March.

Thank you again for your comments and questions. We look forward to tonight's discussion with you all.

Email Correspondence After TEP Meeting #2

TEP Member Email

Before my additional comments, I'd like to commend the team, as I think the information shared is valuable and also given in appropriately sized chunks!

Additional points/questions from TEP meeting #2:

1) Any variation noted, if captured on hospital locale such as Rural VS Urban VS Suburban?
2) Similar to #1, any variation noted with respect to private VS public institutions?
3) Is there any correlation w/ changes in outcome w/ use or lack of use of ancillary services, such as PT, when looking at PROs?
4) I believe prior joint replacement may have been an exclusion criterion, if so, is that only within a certain time frame such as most recent 12 months?
5) Just a comment...since I was asking about the impact of improved capturing of risk factors from a coding perspective...probably going to be more depression, thank you COVID-19!

Thank you for the opportunity to participate and ask these questions!

TEP Member Email

Thank you very much for the interesting discussion tonight – I agree with everyone that this was a very thoughtfully organized, and well-done presentation.

I agree that the c-statistic is not as high as I might have expected, although I understand the reasoning behind why this might be.

There is likely quite a bit of unmeasured risk accounting for this – in my own practice, we care for a number of "high-risk" patients that would not be identified by the variables listed on page

22 – e.g. history of transplant, active cancer diagnosis (even if the surgical diagnosis is OA), use of immunosuppressant medication, and others.

I wonder if this is a scenario where machine learning would perhaps outperform more traditional statistical methods?

Not sure if that is an option, just a thought.

Finally, it was briefly mentioned during the discussion, but I think that race is an important variable not to be overlooked.

Disparities in healthcare can be exacerbated when the input data for a given model are not reflective of the population to which the model is intended to be applied.

At a bare minimum, I would think that it would be helpful to include race as a variable on page 22.

I am grateful to be a part of this important work.

CORE's Response to TEP Member's Email Communications and Outstanding Action Items from TEP Meeting 2

Clinician Attribution

- How are provider specialty descriptions assigned?
    - According to Medicare Claims documentation, the CMS Provider Specialty Code is assigned by the Medicare Administrative Contractor (MAC) based on the corresponding provider identification number (performing National Provider Identifier [NPI]).
- Hand Surgeons attributed to procedures in our dataset
    - To be attributed a procedure in our dataset, the procedure must meet cohort inclusion/exclusion criteria. Therefore, the 5 hand surgeons identified in the attribution algorithm performed elective, primary procedures (with no fractures or hemiarthroplasty). We believe we should keep these providers in the measure as they were paid for the elective, primary hip and knee procedures and are still considered orthopedic surgeons.
- Adult reconstructive surgeons
    - According to Medicare Claims documentation, there is a specialty assignment group of "Plastic and reconstructive surgery". No reconstructive surgeons were attributed procedures in our dataset using the attribution algorithm. These providers may have submitted claims for elective, primary hip and knee procedures, but they were not identified as the primary clinician using the attribution algorithm.

Risk Model Feedback

- Additional variables of interest:

    - Parkinson's, body mass index (BMI) continuous, Smoking, Charlson Comorbidity Index, Physical Health Component of PROMIS-Global/VR-12, long term use of steroids, long term toxicity, history of transplant, active cancer diagnosis, and use of immunosuppressant medications.
        - Thank you for raising these important variables. Please note that we do use BMI as a continuous variable in the risk model. Many other risk factors were also considered but deferred during this iteration, either due to low volume (Parkinson's), data limitations (steroid use), or overlap with the measure outcome (Physical Health Component of PROMIS/VR-12). We will continue to evaluate the risk model over time, and we will take these variables into consideration during ongoing reevaluation of the risk model.
- How does the c-statistic help distinguish performance of providers?
    - The c-statistic describes how well the risk model predicts the measure outcome. Please see the one-pager attached for more information on c-statistic interpretation. Of note, it is important for the risk model to incorporate patient-level risk factors because if they are not taken into consideration, we could not fairly compare providers that have differences in their case mix (for example, if a provider has many patients with existing comorbidities). The measure takes into consideration patient-level risk factors as well as the provider's influence on the patient's outcome in the overall measure calculation.
- Use of machine learning to outperform more traditional statistical methods
    - Thank you for this suggestion. We have explored various analytical approaches for this measure and will continue to evaluate specific machine learning approaches that may improvement model performance. The current approach allows for validation in different datasets and is interpretable to the public. A model developed using machine learning techniques may be more challenging to validate, to interpret and explain to stakeholders, and may not align with the risk variables identified as important by orthopedic experts. Of relevance to this, a recent paper just published illustrates that machine learning may not improve mortality risk prediction after acute myocardial infarction (AMI) (Khera 2021, JAMA Cardiology March 10), to quote the accompanying editorial, "generalized linear model is powerful, and only rarely is there a price – a substantial loss of performance – for choosing it."
- Consideration of race and disparities in health care
    - We appreciate your comments and perspective on disparities in healthcare. We will discuss social risk factor analyses during TEP Meeting 3. We recognize that this is a complex and complicated conversation.
    - We include below information we hope will be helpful for you to review in advance of our meeting tomorrow. We share information on the conceptual relationship between social risk factors and outcomes used as well as a summary of our findings in our dataset. We acknowledge that our testing dataset is limited (and this may be reflective of access issues across the country) and we plan to

continue investigating the impact of social risk on the measure results in the future.

- The conceptual relationship, or potential causal pathways by which social risk factors influence improvement following hip and knee replacement procedures, like the factors themselves, are varied and complex. Similar to other outcome measures, we present four potential pathways that are important to consider:
    1. **Patients with social risk factors may have worse health at the time of admission for their surgery**. Patients who have lower income/education or unstable housing may have a worse general health status and may present for their procedure with a greater severity of underlying illness. These social risk factors may contribute to worse health status at admission due to competing priorities, lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment.
    2. **Patients with social risk factors often receive care at lower quality hospitals**. Patients of lower income, lower education, or unstable housing have inequitable access to high quality facilities, in part, because such facilities are less likely to be found in geographic areas with large populations of poor patients. Thus, patients with low income are more likely to be seen in lower quality hospitals, which can explain decreased likelihood of achieving the improvement outcome following hospitalization.
    3. **Patients with social risk factors may receive differential care within a hospital**. The third major pathway by which social risk factors may contribute to likelihood of not achieving the improvement outcome is that patients may not receive equivalent care within a facility. For example, patients of non-white race or non-English-speaking patients may receive differential or inadequate care and/or education during their stay (such as failure to provide adequate pain control due to biases about pain perception among patients of color or failure to provide educational materials in a patient's preferred language), leading to poorer health outcomes.
    4. **Patients with social risk factors may experience worse health outcomes beyond the control of the health care system.** Some SRFs, such as income or wealth, may affect the likelihood of improvement following hip or knee replacement without directly affecting health status at admission or the quality of care received during the stay. For instance, while a hospital or provider may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing financial priorities which don't allow for adequate recuperation or access to needed treatments, or a lack of access to care outside of the hospital.

- Social risk factors often act on multiple pathways, and as such, individual pathways can be complex to distinguish analytically. Some social risk factors, despite having a strong conceptual relationship with worse outcomes, may not have statistically meaningful effects on the risk model. Some social risk factors also have different implications on the decision to risk adjust or not.
- Based on this model and because the following factors are currently consistently available in our dataset, the following social risk variables were considered for risk-adjustment:
    - Dual-eligible status
        - Following guidance from the Department of Health and Human Services Assistant Secretary for Policy and Evaluation (ASPE) and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable (ASPE 2016, ASPE 2020). We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous variable. However, the eligibility threshold for over 65-year-old Medicare patients is valuable, as it considers both income and assets and is consistently applied across states for the older population.
    - Agency for Healthcare Research and Quality (AHRQ) Socioeconomic status (SES) index
        - The AHRQ SES index score is a well-validated variable that describes the average SES of people living in small defined geographic areas (Bonito et al., 2008). Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. AHRQ-validated SES index score summarizes information from the following 7 variables: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12^{th} grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room.
- Please note:
    - We do not consider race a marker of socioeconomic status; we include it 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 (Irgit and Nelson, 2011; Kerman et al, 2018).
    - While health literacy also reflects social risk, the Hospital-level THA/TKA PRO-PM upon which this measure is based found that patients and

clinical experts on the TEP strongly supported including health literacy in the risk model for a PRO-based measure, due to its very nature of asking patients to complete survey instruments as part of measurement. For this reason, we included it in the final risk model; we therefore do not include health literacy in the specific social risk factor testing.

- Social Risk Factor assessment: We examined the associations of dual eligibility, AHRQ SES index lowest quartile (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 associations between non-white race and substantial clinical benefit (SCB) improvement, AHRQ SES index lowest quartile and SCB improvement, or dual eligibility and SCB improvement (see slide 38). In addition, accounting for these risk factors in the model had little effect on clinician- and clinician group-level RSIRs (see slide 10). 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 and together included in the risk model, correlation coefficients indicated near-perfect correlation in our data (see Slide 11).
- CMS' decision regarding whether or not to adjust for social risk factors is based both on the empirical results (impact on model and measure scores) and the conceptual model and the use of the measure (in a payment program or for public reporting). The Clinician-Level and Clinician Group-Level THA/TKA PRO-PM is not yet included in public reporting or a payment program.
- In making the decision about whether or not to risk adjust for these factors, CMS also considers the potential unintended consequence of adjusting, and the fairness to patients and providers that care for patients with social risk factors of the unadjusted measure score. If the relationship is driven by poorer quality, adjusting will mask the disparity in care. In contrast, an unadjusted measure will illuminate quality differences and create an incentive to mitigate them. Not adjusting, however may disadvantage providers who care for low SES patients, and unintentionally create an incentive for clinicians to care for fewer patients with social risk factors, potentially reducing access to care. CMS considers this to be a small risk currently, given the correlations between the measure scores calculated with and without social risk factors in the model. CMS also considers alternate approaches to risk adjustment for SRFs, such as stratifying payment based upon performance among peer groups of hospitals caring for similar patients with SRFs.
- In addition to our empirical results, we plan to bring your input to CMS to consider. Ongoing research aims to identify valid patient-level social risk factors and highlight disparities related to social risk. As additional variables become available, they will be considered for testing and inclusion within the measure. There are also alternative ways to account for social risk as part of measure program implementation.

<u>References</u>:

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

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

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

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2019.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/pdf-report/second-impact-report-to-congress*. Accessed July 2, 2020.*

Khera R, Haimovich J, Hurley NC, et al. Use of Machine Learning Models to Predict Death After Acute Myocardial Infarction. JAMA Cardiol. Published online March 10, 2021. doi:10.1001/jamacardio.2021.0122

Irgit, K. and Nelson, C.L, Defining Racial and Ethnic Disparities in THA and TKA. Clinical Orthopaedics and Related Research®, 2011. 469(7): p. 1817-1823.

Kerman, H.M., 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): p. 1326-1334.

Measure Results

- Did we look at variation in hospital locale rural vs urban vs suburban? Private vs public institutions?
    - Thank you for these suggestions. We will keep these suggestions in mind during ongoing reevaluation. Of note, when the hospital-level measure was presented to the Rural Measures Application Partnership (MAP), we heard concern regarding adequate case volume for assessing rural hospitals. We will continue to monitor variation in volume and measure results using national-level data in the future. The current Comprehensive Care for Joint Replacement (CJR) dataset is limited.
- Changes in outcomes with use or lack of use of ancillary services, such as physical therapy (PT)
    - Thank you for these suggestions. Of note, the improvement outcome aims to in part evaluate the quality of collaboration among providers and all the care delivered to the patient during and after their procedure. Use of ancillary services is important in considering patient outcomes. We hope that this measure makes hospitals and clinicians accountable to ensure high quality

communication and collaboration among providers and ensuring coordinated care for patients throughout their recovery process.

Cohort Inclusion/Exclusion Criteria

- Exclusion of staged procedures
    - Thank you for your question. The measure exclusion criterion is: Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed. We are actively reviewing this criterion and plan to bring this criterion to the TEP in the near future for ongoing discussion.

Appendix F. Detailed Summary of TEP Meeting #2

Orthopedic Technical Expert Panel (TEP) Meeting #2 Minutes

Wednesday, February 10, 2021 5:00-6:30 PM ET

Participants:

- YNHHS/CORE: Lisa Suter, Rachelle Zribi, Miriam Katz, Fior Rodriguez, Katie Balestracci, Kyaw Sint, Shani Legore, ZQ Lin, Hannah Stiles
- TEP Participants: David Ayers, Thomas Barber, Barbara Lewis, Rachel DuPré Brodie, Cheryl Fahlman, William Hamilton, Cynthia Jacelon, Benita Lattimore, Craig Miller, Michael Perskin, Adam Schwartz, Robert Sterling, Margaret VanAmringe, Christine Von Raesfeld, Patricia Walker, Kevin Woodward, Adolph Yates, Phyllis Bass, Vinod Dasa

CORE Action Items:

- The CORE team will investigate whether adult reconstructive surgeons were attributed procedures in the dataset.

Detailed Discussion Summary

Welcome

- Ms. Fior Rodriguez explained the goal of the Technical Expert Panel (TEP) panel meeting was to review measure cohort results and gain feedback on the approach to clinician attribution and risk model for the Clinician-Level and Clinician-Group Level Total Hip Arthroplasty/Total Knee Arthroplasty (THA/TKA) Patient-reported outcome-based performance Measure (PRO-PM). She noted that the minutes and summary report will be distributed to TEP members following the meeting.
- Dr. Katie Balestracci welcomed the TEP members and provided an update on Hospital-level THA/TKA PRO-PM which was endorsed by the National Quality Forum (NQF) and supported by the Measure Applications Partnership (MAP).

Timeline and Updates

- Ms. Rachelle Zribi reviewed the measure development timeline.
- Ms. Zribi discussed feedback received from the Patient Working Group and Clinical Working Group. She noted their recommendation for ongoing evaluation of the risk model.

Measure Cohort Specifications

- Ms. Zribi reviewed the Measure Testing Dataset, which includes data voluntarily collected from hospitals through the Center for Medicare and Medicaid (CMMI) Comprehensive Care for Joint Replacement (CJR) Model. For measure testing, patient-reported outcome measure (PROM) and risk variable data for eligible procedures between July 1, 2016 – June 30, 2018 were matched to Medicare administrative claims data.
- Ms. Zribi shared the measure cohort inclusion and exclusion criteria.

The materials within this document do not represent final measure specifications for the Clinician-Level and Clinician Group-Level THA/TKA PRO-PM.

Clinician and Clinician Group Attribution Approach and Results

- Ms. Zribi shared the clinician attribution approach which is based on the attribution algorithm developed for the clinician- and clinician group-level THA/TKA Complication measure. This method attributes the outcome for each patient in the cohort to a single Eligible Clinician (EC), who is identified using Part B Physician/Supplier claims (billing claims) for the THA/TKA procedure. She reviewed the algorithm and noted that only six patients were unable to be attributed to clinicians using the attribution algorithm.
    - A TEP member asked if patients were captured via Part B claims using ICD-10 codes or CPT codes for the clinician approach.
        - Ms. Zribi answered that the elective primary procedure is identified using ICD-10 codes.
        - Dr. Lisa Suter noted that inpatient procedures and procedures at hospital facilities will continue to be defined using ICD-10 codes. As the measure expands to include ambulatory surgical centers (ASCs) and hospital outpatient departments (HOPDs), we will define procedures using CPT codes to define eligible procedures.
    - A TEP member noted that there could be simple revision procedures that would be performed in an ASC or an outpatient environment. This could be billed as a partial revision to prevent it from being billed as CPT code 27447-Under Repair, Revision, and/or Reconstruction Procedures on the Femur (Thigh Region) and Knee Joint.
    - A TEP member asked for clarification on whether outpatient procedures are excluded from this measure.
        - Ms. Zribi explained this measure is currently specified for inpatient procedures. Ms. Zribi noted CORE has previously heard interest in expanding the cohort to include procedures performed in outpatient and ASC settings.
    - A TEP member asked about inclusion of inpatient procedures that are billed as outpatient procedures.
        - Ms. Zribi explained the current cohort criteria only considers procedures that occurred in an inpatient setting and are billed as inpatient procedures. Ms. Zribi noted that the CORE team plans to discuss expansion of the cohort to HOPD and ASC settings in an upcoming meeting. Ms. Zribi noted that the CJR data used for measure testing only captures PRO data on hip and knee procedures that occur in the inpatient setting.
    - A TEP member asked if there was a requirement for the minimum number of nights stayed.
        - Ms. Zribi noted the CMS billing rule (Two Midnight Rule), which requires a patient be hospitalized for at least two midnights to be considered an inpatient and be billed for a Part A claim.

- A TEP member stated that TKA was removed from the inpatient only (IPO) list, so patients undergoing THA/TKA could stay less than two midnights and still be considered an inpatient in the billing data.
- Another TEP member noted that patients are classified as an inpatient if the expectation is to remain in the inpatient setting for two midnights, which requires specific documentation. Most hospitals are taking the safe route and billing as an "extended recovery". THAs and TKAs are now off the IPO list and under the Two Midnight rule. They noted additional codes for musculoskeletal procedures have been taken off the IPO list.
    - Dr. Suter agreed with these concerns and reiterated that CORE anticipates discussing these topics in an upcoming meeting. CORE anticipates that this measure will capture eligible procedures performed in the inpatient, outpatient, and ASC settings.
- Four TEP members expressed interest in capturing procedures across these settings given more procedures are being performed in these settings. One member specifically noted concern about a provider's eligible cohort size if the measure only included inpatient procedures.

- Ms. Zribi noted the CORE team refined their clinician attribution algorithm to limit the eligible clinician specialties to Orthopedic Surgery, Sports Medicine, Hand Surgery, Osteopathic Manipulative Medicine, and General Surgery. This approach removes clinician specialties that were anomalies, and in doing so, increases face validity. Ms. Zribi reviewed the frequency and percent of patients attributed to these clinician specialties in the testing dataset.
    - A TEP member suggested separating inpatient and outpatient data because the outcomes could be different.
    - A TEP member asked why specialties with a low number of attributed procedures (Hand Surgery and Osteopathic Manipulative Medicine) were still included in the attribution algorithm.
        - Ms. Zribi noted that these providers were included because they had patients with complete, PRO data in the testing dataset.
        - Dr. Kevin Bozic added that including these specialties increases face validity. He explained that the measure will not separate these specialties and analyze them individually; instead, they are in the total denominator of cases used to create the measure.
    - A TEP member asked whether examination of the types of clinicians in the attribution algorithm would continue for validity purposes.
        - Dr. Suter confirmed that the team will continue to evaluate the attribution algorithm over time. Dr. Suter noted that CORE intends to include as many clinically appropriate patients and providers as possible in the measure for face validity purposes. CMS makes these determinations based on testing data.
    - A TEP member asked how a specialty description is assigned to a surgeon.

        - Dr. Suter responded that the specialty description is assigned by Medicare through the coding process.
    - A TEP member noted surprise that a low number of Sports Medicine clinicians were attributed procedures in the dataset.
        - Dr. Suter reminded the group that the testing dataset includes inpatient procedures at CJR participating hospitals that elected to voluntarily submit PRO data. Dr. Suter noted that the dataset is limited and may not be representative of frequencies at the national level.
    - A TEP member asked if CORE investigated whether hand surgeons attributed to procedures in the dataset performed hemiarthroplasties for fractures, or if the elective flag was incorrectly applied.
        - Dr. Suter noted only elective, primary procedures are included in the measure cohort and revisions and fractures are excluded from the denominator.
    - A TEP member asked for the mean number and range of cases per surgeon as well as the total number of cases with complete PRO data.
        - Ms. Zribi noted that the next TEP meeting will cover PRO response rates.
    - A TEP member clarified that hospitals and medical systems supported the collection and reporting of PRO data in CJR. The TEP member noted it is not a fair evaluation at the surgeon- or group-level if hospitals are not required to report PROs. The TEP member noted that if hospitals were held accountable to the measure, it would drive universal reporting of the measure for all entities.
    - A TEP member asked if any procedures in the dataset were attributed to adult reconstructive surgeons.
        - **Action Item:** The CORE team will investigate whether adult reconstructive surgeons were attributed procedures in the dataset.
    - A TEP member asked if the testing dataset was based on the latest CJR data.
        - Ms. Zribi explained that the CJR data used for testing includes the most recent data that has been cleaned and matched for CJR since measure testing began. PRO data collected for CJR Performance Year 5 was submitted by hospitals last year, and CORE has been working on cleaning and matching these data. The team will have updated results based on new CJR data in the future.
        - Dr. Suter noted that there is a one-year post-operative data collection window after surgery, which elongates the timeline for data matching.

Risk Model Approach and Results

- Ms. Zribi reviewed the risk model approach, noting risk adjustment is used to account for case mix difference across providers. CORE is using a clinically-derived and empirically tested risk model developed by the Hospital-level THA/TKA PRO-PM measure development team, with risk variables identified in the literature and selected by orthopedic experts for their relationship with the measure outcome.

- Ms. Zribi noted that during development of the Hospital-level THA/TKA PRO-PM, the CORE team also tested a purely empirically derived risk model in which there was no prior selection of risk variables but instead purely a statistical approach to variable selection. CORE favored the combined clinically- and empirically derived model due to greater face validity, to avoid year-to-year variation, and because model performance was comparable to the empirically-derived risk model performance.
- Ms. Zribi reviewed the 17 variables included in the risk model. Some variables, such as health literacy and back pain, are reported by the patient. Other risk variables are identified using Medicare administrative claims data from the procedure claim and 12 months before the procedure.
- Ms. Zribi reviewed the risk model performance results. The c-statistic indicated moderate model discrimination across the development and validation models and was lower than the c-statistic for the Hospital-level THA/TKA PRO-PM. CORE found good calibration of the model between the development and validation datasets and noted that predictive ability statistics indicated moderate discrimination in distinguishing high- from low-risk patients.
- Ms. Zribi described CORE's interpretation of the c-statistic results given differences in findings between risk model assessment of this measure and the Hospital-level THA/TKA PRO-PM on which this measure is based. Findings indicate differences in model performance using the first year of THA/TKA procedure data (a higher c-statistic of 0.68) and the second year of THA/TKA procedure data (a lower c-statistic of 0.59). There were three notable differences between the first and second years of CJR data:
    - There were slightly lower patient improvement rates using the first year of CJR data (64.37%) compared to the second year (70.02%).
    - There were fewer patients with low health literacy in the first year compared to the second year.
    - In the first year of CJR data, health literacy showed a strong, statistically significant, linear relationship with the improvement outcome (such that higher health literacy was associated with improvement) while in the second year of CJR data, the association was weaker and without a clear linear trend.
- Ms. Zribi explained that the CORE team tested other risk model approaches for potential better model performance, including the purely empirically derived risk model and inclusion of social risk factors other than health literacy, but none of these approaches provided a better model fit.
- Ms. Zribi noted that the CJR data represents a unique sampling of patients undergoing THA/TKA procedures and the trend of less influence of social risk in the second year of CJR data was unexpected. She noted that these results may reflect shifts in patient selectivity across performance years. For this measure, CORE recommends using the combined clinically- and empirically derived and empirically tested risk model to ensure measure alignment with the Hospital-level THA/TKA PRO-PM, and plans to re-test the risk model prior to mandatory use of the measure in the future.
    - A TEP member agreed with CORE's recommendations and expressed interest in updating results to include procedures in subsequent CJR performance years.

They noted that capturing PROs was an evolving process and although individuals and groups can practice patient selection, hospitals cannot. The TEP member noted that patient selection by providers may interfere with risk adjustment. They shared that different hospitals may have submitted data across different years, and that PRO implementation across sites required an economy of scale to capture patients. The current results may show a smoothing out over time.

- A TEP member asked if the risk variable list was comprehensive and noted that Parkinson's Disease is the top reason for a dislocation. The TEP member also suggested surgeons having more experience could have resulted in better outcomes in the second performance year. They also noted surgeons may be learning to communicate better with individuals with lower health literacy.
    - Ms. Zribi noted that age and other variables indicating patient frailty are included.
    - Dr. Suter explained that the team would investigate Parkinson's Disease and noted that CMS updates measures to incorporate changes in coding and clinical practice.
- A TEP member asked how the health literacy variable was captured and noted the importance of giving a provider credit if they enhance health literacy by communicating and educating patients.
    - Ms. Zribi described the health literacy variable, which is based on a widely used, validated questionnaire.
    - Dr. Suter noted that health literacy is a complicated concept, but that evidence suggests that the health literacy questionnaire captures health literacy well with little burden to completing the questionnaire.
- A TEP member noted that single-question items such as health literacy and back pain were chosen to limit the total number of questions and reporting burden. They expressed that data collection burden can decrease rates of reporting from patients and providers.
- A TEP member expressed support for the risk model. They expressed interest in investigating whether the model stays stable with future CJR data.
    - Dr. Suter noted that CORE will share updated results with additional CJR data when available. She reiterated that the risk variables are supported by the literature and orthopedic experts.
- A TEP member asked if the CORE team had a target for the c-statistic. They noted concern regarding the lower c-statistic.
    - Dr. Suter noted that it is challenging to determine a threshold for a c-statistic given many factors influencing the outcome. Dr. Suter shared that CORE has seen c-statistics similar to this measure previously. For example, readmission measures typically have c-statistic values in the 0.62 and 0.64 range, while other measures such as mortality measures typically have higher c-statistic values.

- A TEP member commended the CORE team on their testing methodology and supported continued examination of the stability of the risk model over time as well as a better understanding of patient selection.
- A TEP member requested clarification regarding how differences in patient populations versus differences in provider quality may come into play in the measure. They noted that social risk factors are not captured in the risk model and may be contributing to the measure outcome.
    - Dr. Suter thanked the TEP member for their comments and noted that social risk and response bias will be discussed at an upcoming TEP meeting.
- A TEP member highlighted the importance of risk adjustment and expressed support of aligning the risk model with the hospital-level measure. The TEP member asked for clarification on interpretation of c-statistic values and how the risk model distinguishes providers with different patient populations.
- A TEP member noted support for the risk model.
- A TEP member expressed appreciation for CORE's analyses and noted that the current risk model performance results meet the threshold of acceptable.
- A TEP member agreed with CORE's recommendations. They noted importance of understanding response rates and associations with "cherry picking". Regarding the risk model, they noted that the Function and Outcome Research for Comparative Effectiveness in Total Joint Replacement (FORCE TJR) team saw improved risk prediction when they considered body mass index (BMI) as a continuous variable, smoking history, preoperative physical composite score (as measured by the VR-12 and PROMIS Global PROMs), orthopedic comorbidities in addition to those noted in the risk model such as concurrent hip and knee pain, and the Charlson Comorbidity Index to account for the synergistic effects of multiple comorbidities.
- A TEP member asked about differences between data in the first two years of the CJR model and asked whether there was increased coding of comorbidities overtime. They noted that as people are exposed to factors that may improve outcomes, people may better code those risk factors.
    - Ms. Zribi said that the CORE team has not looked at this issue in terms of coding changes over time but can consider this in future reevaluation.
- A TEP member asked if CORE considered patients' long-term use or toxicity of treatment for other comorbidities (i.e. steroids from treating other comorbidities).
    - Dr. Suter thanked the TEP member for their feedback and explained that information about medications may be challenging to capture.
- A TEP member noted that the risk models are based on clinical factors. They noted that the factor most highly associated with good outcomes is patient expectation, not clinician outcomes. The TEP member noted that this measure does not capture patient expectation. They also noted that CJR aimed to improve care of populations at hospitals and if the program was successful, care

would improve and less complications would result. They expressed that it is possible that improved care via the CJR model could result in a decrease in the c-statistic over time.

- A TEP member appreciated CORE's comparison of the purely empirically derived and combined clinically- and empirically-derived risk models. They asked if the team examined differences between THA and TKA and the proportion of those surgery types. The TEP member noted that some surgeons or groups may treat more or less of one type of patient and might see this as important to fairness. They also recommended patient stratification by race and ethnicity for the risk model.
    - The CORE team thanked the TEP member for their feedback.

Next Steps

- Ms. Zribi thanked the group for their feedback and asked that any additional questions be emailed to the team.

Appendix G. Detailed Summary of TEP Meeting #3

Orthopedic Technical Expert Panel (TEP) Meeting #3 Minutes

Tuesday, March 23, 2021 5:00-7:00PM ET

Participants:

- YNHHS/CORE: Lisa Suter, Rachelle Zribi, Miriam Katz, Fior Rodriguez, Katie Balestracci, Kyaw Sint, Shani Legore, ZQ Lin
- CMS: Lisa Marie Gomez
- TEP Participants: Thomas Barber, Phyllis Bass, Rachel DuPré Brodie, Barbara Lewis, Margaret VanAmringe, Benita Lattimore, Nan Rothrock, Bill Hamilton, Robert Sterling, Christine Von Raesfeld, Adam Schwartz, Kevin Woodward, Michael Perskin, Craig Miller, Adolph Yates

CORE Action Items:

- The CORE team will follow up with more detailed information as to how the population in the testing dataset compares to the Medicare population and overall US population.
- CORE will share more information regarding how accurately same-day procedures can be identified in billing codes.

Detailed Discussion Summary

Welcome

- Ms. Fior Rodriguez welcomed Technical Expert Panel (TEP) members, reviewed the goals for the meeting, and reminded members that all discussions must remain confidential until made public by the Centers for Medicare and Medicaid Services (CMS).

Social Risk Factor Approach and Results

- Ms. Rachelle Zribi reviewed CORE's goals of assessing the impact of social risk factors on the measure results to reduce known disparities in the field and ensuring that the measure does not have any unintended consequences. She reviewed the list of social risk factors that were examined for potential inclusion in the final risk model: non-white race, dual eligibility, and bottom quartile of the Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index. This approach is aligned with that taken in the development of the Hospital-level THA/TKA Patient-Reported Outcome Based-Performance Measure (PRO-PM) on which this measure is specified. Ms. Zribi noted that health literacy was included in the risk model due to its importance and relevance for a measure which asks patients to complete patient-reported outcome measures (PROMs), and due to its strong correlation with social risk based on literature.
- Ms. Zribi reviewed the social risk factor analyses results. She noted there was little to no impact in adding these risk factors to the model, as evidenced by comparing risk model c-statistics and noting the near perfect correlations in calculated measure results of risk models with and without social risk factors included. Therefore, the CORE team

recommends following the approach taken with the Hospital-level THA/TKA PRO-PM to adjust for health literacy in the risk model, and to leave non-white race, dual eligibility and bottom quartile of the AHRQ SES Index out of the risk model; but to include these social risk factors in the propensity model to adjust for potential response bias. Additionally, CORE recommends monitoring the relationship between social risk, response rate, and measure results throughout measure testing for potential reevaluation of the risk model. Ms. Zribi also reviewed alternative opportunities for consideration in accounting for social risk in payment programs.
    - A TEP member asked if these social risk variables are not adding any value, why the CORE team recommends keeping these variables in the risk model. The TEP member also asked if there was any burden in collecting these social risk factors.
        - Ms. Zribi noted that these variables will not be included in the risk model but will be included in non-response bias approach.
        - Dr. Lisa Suter noted that these social risk factors are available in claims and enrollment data, and therefore do not place additional burden on providers.
- A TEP member noted with interest a recent article that identified the impact of the expansion of Medicaid under the Affordable Care Act on the volume of THA and TKA procedures.
- A TEP member requested clarification that the aforementioned social risk factors, while important for consideration, do not add any further discrimination within the risk model.
    - Ms. Zribi confirmed this interpretation.
- A TEP member applauded CORE for investigating the effects of including social risk variables in the risk model and recommended continuously monitoring the impact of these social determinants of health on the measure results in the future.
- A TEP member noted that only 7.7% of THA/TKA patients were non-white and asked how the population in the testing data compares to the Medicare population in the United States. They noted concern that the sample population is not representative of the Medicare population.
    - Dr. Suter noted that according to recent data, non-white patients are a small proportion of all Medicare patients undergoing primary elective THA/TKA. CORE believes the population in the testing dataset is reasonably representative of the national population, and that the low percentage of non-white patients likely reflects unequal access to healthcare.
    - Another TEP member agreed with these concerns and asked how race is reported and which other social risk variables CORE considered adding to the risk model based on current literature.
        - Dr. Suter noted that race is identified via the Medicare enrollment database and that there are limited options for identifying race in this database. The options for race in the Medicare Enrollment database are white, Black, other, Asian, Hispanic, and North American native. This list is not as exhaustive as that identified by the Office of Minority Health. Dr.

Suter noted that race information is collected during Medicare enrollment and may not be self-reported.
    - **Action Item:** The CORE team will follow up with more detailed information as to how the population in the testing dataset compares to the Medicare population and overall US population.
  - Dr. Kyaw Sint shared a link to the options for race identification in claims data: https://resdac.org/cms-data/variables/beneficiary-race-code-base.
- A TEP member noted surprise that social risk did not impact the measure results.
- A TEP member noted that the color of a patient's skin or their income should not affect their change in pain or function after a THA/TKA procedure. They noted that social risk factors should not be used in risk adjustment and instead suggested stratification for payment programs, as supported by the 21st Century Cures Act. This legislation requires that CMS assess penalties based on a hospital's performance relative to that of other hospitals with a similar proportion of patients that are dually eligible for Medicare and full benefit Medicaid.

Non-Response Bias Approach and Results

- Ms. Zribi reviewed the importance of assessing non-response bias on the measure results given that survey completion is voluntary. Ms. Zribi noted that patient-reported outcome (PRO) data is unlikely to be missing at random and potential bias may be introduced if there are systematic differences between patients that respond versus those who do not, including differences in social risk or clinical outcomes.
- Ms. Zribi presented PRO submission rates for clinicians and clinician groups.
- Ms. Zribi reviewed the team's statistical approach to accounting for non-response bias, stabilized inverse probability weighting. This approach was vetted and used by the Hospital-level PRO-PM measure development team. Incorporating stabilized weights in the calculation of the measure results helps to account for 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.
- Ms. Zribi presented clinician and clinician group risk-standardized improvement rates (RSIRs) with and without weighting for potential non-response. She noted that there was little impact of adjusting for non-response on the results in the dataset, but that CORE recommends including non-response weighting in the calculation of final measure results and continuing to evaluate non-response in the future, given the expectation that non-response may continue to be a concern.
  - A TEP member asked for clarification of the RSIR and asked what the variability was in response rates between clinicians.
    - Ms. Zribi clarified that the hospital RSIR reflects the risk-standardized percentage of patients that met the improvement threshold. For clinicians, RSIRs with and without weighting for non-response ranged from about 18% to 88% with a mean around 64%. For clinician groups, RSIRs with and without weighting for non-response ranged from about 21% to 86% with a mean of around 64%. The overall PRO submission

rates for clinicians and clinician groups are shown on slide 16 and CORE found PRO submission rates of 42% for clinicians and 36% for clinician groups.
        - Dr. Suter reminded the TEP that these data are from the Center for Medicare and Medicaid Innovation (CMMI) Comprehensive Joint Replacement (CJR) program, in which data submission was gradually incentivized over multiple years. Dr. Suter explained that this incentivized approach which started at 50% complete data or 50 cases, as well as the difficulty of collecting PRO data, contribute to the low response rates. Dr. Suter asked the TEP how CMS can be more transparent about response rates and what a reasonable response rate might be.
    - A TEP member asked if it was possible to use the propensity score to calculate observed and expected response rates for clinicians and clinician groups. They recommended consideration of including a clinician-level adjustment that is based on their patient population in the overall measure results.
        - CORE thanked the member for their suggestion and will share it with CMS.
    - A TEP member noted that in CJR, larger hospitals with more resources and larger volumes often are more successful in capturing and reporting PRO data. The TEP member asked if surgical volume can be used to calculate an expected PRO data submission rate.
        - Dr. Suter thanked the member for their suggestion and will share it with CMS. Dr. Suter noted that the CJR model incentivizes an absolute number of successful PRO submissions or a proportion of surgeries performed, and hospitals can choose either threshold, resulting in variation in how surgical volume will be associated with PRO response rates in CJR data.
    - A TEP member asked who selects which patients are given PRO surveys.
        - Ms. Zribi responded that the PRO data submission in the CJR model is voluntary, so hospitals can select if and how to participate. If the current measure went to national reporting, CORE would recommend that clinicians and clinician groups attempt to capture PRO data for all patients so that the data is representative.
        - The TEP member noted this leaves room for clinicians to "cherry pick" patients who are more likely to show substantial clinical benefit. They asked for clarification on what the CORE team is doing to prevent this.
            - Dr. Suter noted that CORE will recommend maximum flexibility in PRO data collection to mitigate unintended consequences. Dr. Suter noted that in CJR there are a variety of ways in which hospitals can collect PRO data, either in a systematic way or a single clinician may be activated to collect the PRO data. Dr. Suter thanks the TEP member for this input and noted that CORE agrees that CMS should consider an implementation structure that addresses this concern.

        - Another TEP member agreed with concerns about cherry picking and noted that it is often easier to collect the baseline PROM. They asked if CORE investigated whether there are differences between patients who have completed PROMs versus those with incomplete or missing PROMs.
            - Dr. Katie Balestracci noted that the CORE team explored this question using data collected in the first year of CJR collection and did not see many significant differences in patient characteristics between patients with complete PROMs and those without. She noted that the statistical approach to potential non-response bias in the measure is intended to address this issue, and that the CORE team will continue monitoring non-response bias going forward.
    - A TEP member noted that the measure could be coupled with rewards and penalties for the proportion of patients for whom providers successfully collect and submit PRO data. Additionally, they recommended sharing with providers that hospitals with high response rates are typically more successful. This strategy may be effective in encouraging hospitals to increase their reporting rates overall.
    - A TEP member clarified their understanding that CORE assumes social determinants of health may contribute to non-random missing PRO survey data.
        - Dr. Suter and Dr. Balestracci confirmed this interpretation and noted that social risk factors are included in the non-response bias approach. Dr. Balestracci noted that the testing dataset was relatively small, and CORE assumes that social risk factors will be influential to PRO response during national measure implementation.
        - Dr. Balestracci noted that CORE seeks to support clinicians and hospitals in limiting non-response bias and noted the importance of integrating PRO data collection in clinical workflows. CORE believes that using PRO data can enhance patient engagement in understanding their outcomes and support clinician decision making.

Measure Results

- Ms. Zribi explained the importance of reliability testing and reviewed the signal-to-noise reliability results with the TEP. The results showed excellent reliability for clinician and clinician groups with more than 25 patients with complete PRO data.
    - A TEP member asked how the signal-to-noise results were calculated.
        - Dr. ZQ Lin clarified that reliability calculations are derived based on the hierarchical model used to estimate measure results; as part of the model, CORE estimates between-provider variance. The reliability calculation is a ratio of between-provider variance divided by between-provider variance plus within-provider variance. If between-provider variance accounts for most of the variance, then CORE will be able to discriminate one provider from another.

Measure Specification Updates

- Ms. Zribi presented CORE's measure specifications updates, including expanding the measure cohort to include eligible THA/TKA procedures that occur in hospital outpatient department (HOPD) and ambulatory surgical center (ASC) settings. This is in response to recent CMS rulings that expands payment for THA/TKA procedures in these settings, and feedback from the Rural Measures Application Partnership and various stakeholders to expand the measure cohort.
    - A TEP member agreed with the recommendation to expand the cohort, noting this change would be beneficial to the measure, and not expanding the cohort would be detrimental. They noticed that many patients opt to undergo THA/TKA procedures in ASCs.
    - A TEP member agreed with the recommendation to expand the cohort and suggested CORE ensures that incentives for response rates across settings should be standardized.
    - Four other TEP members stated their support for expanding the measure cohort. One TEP member suggested CORE investigates response rates across the different settings in the future.
    - A TEP member strongly agreed with the recommendation to expand the cohort. They noted that this expansion would not be significantly burdensome. They noted that according to the new CJR Final Rule, the CJR model would begin collecting data on HOPD THA/THA procedures and asked whether the measure would identify eligible patients using ICD-10 codes or CPT codes.
        - Dr. Suter noted that the CJR Final Rule has not yet been published. She noted that eligible patients from outpatient and ASC settings would be captured regardless of how those procedures are coded.
        - The TEP member noted an increase in same-day discharges for Medicare eligible patients and emphasized the importance of capturing data from HOPDs and ASCs because of increased cardiopulmonary complications in Medicare patients who have been discharged on the same day of their surgery. They also noted that some patients who intend to stay overnight are discharged on the same day, and vice versa.
        - Another TEP member noted their support for capturing patients who are discharged the same day in the measure cohort. They asked whether CMS would be able to distinguish patients who were discharged on the same day as their procedure from patients who stayed overnight.
            - Dr. Suter explained that CORE relies on billing codes which indicate the setting of the procedure, inpatient, outpatient, or ASC.
                - **Action Item:** CORE will share more information regarding how accurately same-day procedures can be identified in billing codes.
            - Another TEP member noted that if a patient is billed as an outpatient under the HCPCS codes, the physician cannot bill for an

observation night because the overnight stay is bundled into the HCPCS code for the procedure. Unless there is a medical complication that generates a "g" code for the observation night, these overnight stays are not captured. They recommended that CORE look into the date of the procedure and discharge to identify whether the patient had an overnight observation stay. The TEP member also noted that some states allow ASCs to act as overnight facilities.

- A TEP member asked if the American Society of Anesthesiologists (ASA) Score data is captured.
  - Dr. Suter noted that the ASA Score is not captured. She explained that the ASA Score rates an individual's surgical or anesthetic risk. Dr. Suter noted that the Hospital-level PRO-PM team looked into using ASA Score data as a potential risk variable, but orthopedic stakeholders flagged that this would be too burdensome on anesthesiologists.
  - The TEP member noted that various research articles highlight the importance of using the ASA Score for dislocation risk.
- A TEP member expressed support for the recommendation to expand the measure cohort and asked whether there would be an expectation for equal improvement rates for patients that undergo procedures in inpatient, HOPD, and ASC settings.
  - Ms. Zribi stated that for the current measure, procedures across different settings but performed by the same clinician or clinician group would be grouped together. Therefore, there would be one measure applied to all surgical settings. Ms. Zribi asked the TEP member if they think an overall clinician or clinician group score encompassing all surgical settings would be important for patients/providers, or if separate scores for each individual setting would be more helpful.
  - The TEP member noted that both may be helpful and expressed the importance of having clear quality scores on publicly reported sites given that patients may look online. They also emphasized that measures should be equitable across sites.
- One TEP member noted that the California Joint Replacement Registry aimed to add ASC data to the registry since this would increase the number of facilities in the registry, increase the volume of patients per surgeon, and also volume of patients per group since many of the surgeons perform surgeries in more than one facility. Additionally, inclusion of ASC and HOPD data would mean that more clinicians and groups would be eligible for performance measurement. They suggested CORE look at each entity as a unit of accountability. The TEP member noted interest in identifying complications for patients across settings.
  - Dr. Suter noted that there are separate payment programs and measure sets for HOPDs and ASCs. However, the current measure is being developed for the Quality Payment Program (QPP), which is setting

agnostic. Dr. Suter agreed that identifying specific differences in patient demographics across settings is important.
    - Another TEP member expressed concern about separate public reporting of outcomes in PRO surveys across settings. They noted that patients may prefer to have procedures performed in the setting with the highest PRO survey improvement scores, even if that setting may be clinically inappropriate for that patient.
- Ms. Zribi discussed the recommendation to expand the postoperative PRO data collection window. Previously, the postoperative PRO data collection window was 9-12 months after the THA/TKA procedure. However, clinical stakeholders noted that this limits successful data collection, as follow-up appointments are oftentimes scheduled more than one year after a THA/TKA procedure. CORE recommends expanding the post-op data collection window to 10-14 months.
    - A TEP member noted their support for this recommendation and noted that follow-up appointments typically fall more than one year after the procedure. They noted allowing an extended timeframe may allow providers to capture more data.
    - Three TEP members strongly supported expanding the data collection period.
    - A TEP member noted that patients often reach their maximum improvement in pain and function within 6-9 months of the procedure but suggested further expanding the post-operative PRO data collection window to 9-24 months after the procedure to collect as much data as possible. They stated that the outcome will be the same at 12, 14, 16, and 24 months, without deterioration until 24 months.
        - Another TEP member agreed with this alternative window, noting that it would give leeway to collect data from populations that may move or may not be able to come back to their provider because of the COVID-19 pandemic. The TEP member noted that expanding the timeframe to two years could be problematic as it hinders actionable use of the data and agreed with the 10–14-month recommendation.
        - A TEP member agreed with the recommendation to expand the postoperative PRO data collection timeframe. The TEP member asked how the postoperative PRO data collection window would operate for patients who underwent staged procedures, since oftentimes those patients would make one follow-up appointment for both surgeries.
        - Ms. Zribi noted that CORE is actively discussing how staged procedures would be handled. Currently, staged procedures that occur within the measurement period are excluded from the measure due to the overlap of recovery of the procedures. She encouraged TEP members to provide feedback regarding the defined timeframe for staged procedures.
        - Dr. Balestracci stated that it can be challenging to match PRO data for staged procedures and identify the outcome for an individual procedure.

        - A TEP member noted that historically, 20 procedures are required to accurately measure a physician's performance based on PRO surveys. He stated that few surgeons perform 20 bilateral THA/TKA procedures and that bilateral procedures are becoming less common, so including them may not be worth the hassle.
    - A TEP member supported the expanded post-operative PRO data collection timeframe. They noted that the first six months after their procedure were problematic and felt that they would have been able to give accurate information about their recovery 10-14 months after the procedure.
- A TEP member reemphasized an earlier comment about the importance of using PRO data to support patients in making informed choices and also to drive clinical practice.

Next Step

- On behalf of CORE, Ms. Rodriguez thanked the group for their time and feedback. She noted that meeting minutes and TEP summary report will be drafted and distributed in the coming weeks, and that CORE will request TEP review of the summary report before public posting.

Appendix H. Summary of Email Correspondence After TEP Meeting #4

Email Correspondence After TEP Meeting #4

TEP Member Email

Good morning. I was hoping you could share this information with your group at Yale/CORE regarding PTs and outcomes measures data collection. I'm attaching an APTA document regarding PT's and MIPS participation. In the most general terms, it's primarily PTs in private practice who would participate in MIPS if certain criteria are met, or may have voluntarily participated. So, there may be some data, but could be limited. Certainly, it limits provider types, much like some discussions about varied hospital types/locations, this could possibly bias some information, as hospital-based outpatient programs, and smaller outpatient providers are not likely included.

Maybe something to investigate would be the APTAs outcome registry, which is approved by CMS as a Qualified Clinical Data Registry (QCDR). I can't speak to the specifics about it, but if there were any interest in investigating any usefulness of that data for comparison, I'd be happy to see if anything could be facilitated between the Yale/CORE team and APTA.

Thank you for allowing my participation in the TEP group!

CORE's Response

Thank you very much for sharing this information with our team. We will look into the APTA outcome registry to better understand the types of measures it supports. Thank you again for your ongoing support and contributions.

Appendix I. Detailed Summary of TEP Meeting #4

Orthopedic Technical Expert Panel (TEP) Meeting #4 Minutes

Tuesday, July 13, 2021, 5:30-7:00 PM ET

Participants:

- YNHHS/CORE: Kerry McDowell, Lisa Suter, Rachelle Zribi, Emma Turchick, Sheng Zhou, Kyaw Sint, ZQ Lin
- TEP Participants: Thomas Barber, Phyllis Bass, Benita Lattimore, Vinod Dasa, Rachel DuPré Brodie, Cheryl Fahlman, William Hamilton, Cynthia Jacelon, Barbara Lewis, Craig Miller, Michael Perskin, Nan Rothrock, Jonathan Schaffer, Adam Schwartz, Robert Sterling, Margaret VanAmringe, Kevin Woodward, Adolph Yates

CORE Action Items:

- CORE will share the specific CJR hospital distribution (200 hospitals) details with TEP members. *Complete*
- CORE will follow up after the meeting with TEP members who attended the call by phone to receive their face validity responses and any additional feedback. *Complete*

Detailed Discussion Summary

Welcome

- Ms. Kerry McDowell welcomed Technical Expert Panel (TEP) members, reviewed the goals for the meeting, and reminded members that all discussions must remain confidential until made public by the Centers for Medicare and Medicaid Services (CMS).

Measure Score Validity Approach and Results

- Ms. Rachelle Zribi reviewed The Center for Outcomes Research and Evaluation (CORE's) approach to empirical validity testing to assess whether measure scores indicate quality of care as required for the National Quality Forum (NQF). She explained for the empirical measure score validity testing, CORE compared the measure results of the clinician- and clinician group-level THA/TKA PRO-PM risk-standardized improvement rates (RSIRs) with the clinician- and clinician group-level Merit-based Incentive Payment System (MIPS) THA/TKA complication measure risk-standardized complication rates (RSCRs) using data from 2016 to 2018. She noted the complication measure assesses several complication outcomes that can occur within 90 days of an elective THA/TKA procedure. CORE examined the distribution of RSIRs compared to the RSCRs, as well as to the RSCR performance categories.
- Ms. Zribi noted that before testing, CORE expected there may be a relationship between these measure scores such that higher RSIRs, which indicate better outcomes, would be associated with lower RSCRs.

- Ms. Zribi acknowledged the challenges associated with this testing, including the patient-reported outcome-based performance measure (PRO-PM) is a unique and novel measure outcome, these two measures use different timeframes for the measure outcomes, and the differences in the severity of the measure outcomes. Lastly, she noted the group of patients included in the measure results are different, the PRO-PM results only include patients with complete patient-reported outcome (PRO) data, while the complication measure results include all eligible patients for that specific clinician or group within the measurement timeframe.
- Ms. Zribi reviewed the clinician-level correlation analysis results. For clinicians with more than 25 procedures with complete PRO data, there was a low negative correlation, which is small in magnitude and is not statistically significant between RSIRs and RSCRs.
    - A TEP member referred to slide 9 and expressed concern over combining low- and high-volume clinicians. Given the exclusion of clinicians with less than 25 procedures, they proposed clinicians who did fewer procedures could have a higher complication rate. They asked if CORE considered comparing the group of the low-volume clinicians to the group of the high-volume clinicians (although there would be a small sample size).
        - Ms. Zribi thanked the member for their suggestion. She referred to slide 10 and noted CORE looked at all clinicians in the dataset (including those without the 25-case volume restriction) and the results were similar. The CORE team provided the correlation details as correlation without volume thresholds: -0.034 (p=0.2316) at clinician-level and, -0.008 (p=0.8491) at clinician group-level.
        - Dr. Lisa Suter thanked the TEP member for the great suggestion and noted this was a good reminder of the published literature indicating the relationship between procedural volume and quality outcomes in lower extremity arthroplasty procedures, and acknowledged the complexities given the variable PRO response rates. Sometimes low volume doesn't necessarily indicate a low-volume provider, rather that they submitted a low volume of PRO data, and CORE would need to look into the overall denominator cases. Dr. Suter noted CORE will advise CMS to consider procedure volume using claims data and stratifying by decile or quintile of volume then comparing to the PRO-PM results.
    - A TEP member asked for clarification on the meaning of the clinician results and asked if that means the complication rates did not have a relationship with the RSIRs.
        - Ms. Zribi confirmed this interpretation.
- Ms. Zribi reviewed the clinician-level results of RSIRs for all clinicians categorized by RSCR national performance categories. This comparison included all providers with complete PRO data in the Comprehensive Care for Joint Replacement (CJR) dataset and

is not limited to providers with at least 25 procedures with complete PRO data. Results showed few clinicians were in the outlier groups with one clinician in the "better than the national rate" and five clinicians in the "worse than the national rate." The majority of clinicians fall into the "no different than the national rate" or "number of cases too small" categories, and for each of these categories there is a wide range of RSIRs, making it difficult to identify a relationship between improvement rates and these performance categories.

- Ms. Zribi reviewed the correlation analysis results for the clinician group level. Similarly, for the clinician groups with more than 25 procedures with complete PRO data, there was a low negative correlation, which was small in magnitude and not statistically significant between RSIRs and RSCRs.
- Ms. Zribi reviewed the results of RSIRs for all clinician groups categorized by the RSCR national performance categories. This comparison included all providers with PRO data and is not limited to clinician groups with at least 25 procedures with complete PRO data. The results showed there were ten clinician groups in the "better than the national rate" and seven clinician groups in the "worse than the national rate." The majority of clinician groups fell into the "no different than the national rate" or "number of cases too small" categories. Therefore, it was difficult to identify a relationship between the improvement rates and these performance categories.
- Ms. Zribi summarized the correlation statistics for these comparisons as low and not statistically significant. She noted that CORE is not surprised by the results given a priori expectations. Specifically, for the comparisons to the performance categories, there were very few providers performing in the better or worse categories and she acknowledged the differences in the outcome rates may impact these correlation analyses as the complication measure assesses an outcome that has a very rare occurrence (around 3% for clinician- and clinician group-level), while PRO-PM improvement rates are much higher (around 64%). She explained these results are influenced by, and potentially limited by, the testing dataset as these PRO data were collected and submitted as part of the CJR model at the hospital level. There were moderate PRO response rates (32% for clinicians and clinician groups) due to the collection incentives provided in the CJR model. Finally, low variation in the complication measure results make it hard to detect statistically significant differences. She acknowledged the limitations of the testing dataset and noted CORE will advise CMS retest measure score validity in a larger dataset in the future.
- Dr. Suter added there are approximately 200 hospitals participating in the voluntary CJR PRO data collection, which is a small proportion of the approximately 4,000 hospitals that perform THAs/TKAs. She reconfirmed there are few statistical outliers on the complication measure. She acknowledged that CORE recognizes that although directionally it's reassuring, they're not seeing the alignment they would like to. The TEP

member suggestion to look at overall provider volume as another metric is a great idea. CORE plans to continue exploring these issues.

- A TEP member asked if there was another measure other than the complication measure that could have been chosen to potentially show a better correlation, possibly utilization or Medicare Spending Per Beneficiary (MSPB).
    - Ms. Zribi thanked the member for the suggestions of utilization and MSPB. She noted the CORE team considered potentially looking at Consumer Assessment of Healthcare Providers and Systems (CAHPS) for MIPS. The CAHPS for MIPS were specific to primary care providers, and it was unlikely clinicians in this measure would report on that measure. CORE reviewed the Agency for Healthcare Research and Quality's (AHRQ) CAHPS Surgical Care Survey, and the concepts pertained to patient experience (e.g., communication with nurses or care team) and CORE chose not to pursue it given the survey focused on patient experience rather than outcomes. She reiterated the challenges associated with this novel measure and unique outcome.
- Another TEP member asked about the distribution of the 200 hospitals in CJR and if the hospitals were similar or diverse in terms of mostly urban, rural, and academic characteristics, or if they were all clustered around a similar type of hospital that's reporting the PRO data.
    - Ms. Zribi confirmed that the CORE team looked at the overall distribution, and the distribution across different performance years, showing there wasn't a large trend of mostly teaching or urban hospitals, rather there were various proportions of hospital types and locations.
        - **Action Item:** CORE will share the specific CJR hospital distribution details with TEP members.
- Another TEP member noted the hospital characteristics in the CJR data were based on metropolitan statistical areas (MSAs) and as such contain mostly urban and suburban hospitals, which may result in some bias. They suggested using the THA/TKA readmission measure (NQF #1551) to correlate readmissions as an alternative to the complications because readmissions might correlate with the patient's sense of recovery in their outcomes.
- The same TEP member agreed with the CORE team and other members that these measures and timeframes are very different, and there's potential for selection bias. The variable data collection periods between complications (7, 30, and 90 days), and PROs (between 9-12 months) may influence the results. Some patients have complications soon after their procedure that do very well at the one-year follow up, so a lot of variability could be subtracted out. In addition, the standard deviation of the complication measure is incredibly narrow, around 2-3%, so there's a small window of complications versus improvement in PROs,

which occur across more patients, and for some of these patients, just because they had a complication recorded doesn't mean they weren't satisfied with their joint replacement. Lastly, they noted CJR incentivized PRO data collection incrementally, but it's important to be mindful that hospitals that weren't ready for this type of data collection may have selectively reported their PROs. Also, patients with complications may be less likely to fill out the PROMs and report that they didn't do as well, because they may still be in the midst of recovering from their complication. It is also likely they may never have been given a PROM to fill out.

- Another TEP member agreed and noted they didn't expect any correlation with this measure for all the reasons previously stated. First, the time difference is significant and when measuring these PROs at about one year people with complications likely already recovered. Secondly, the PROs are highly correlated with the expectations of the patient and matching expectations with the provider, which is not going to be reflected in the complication results. They expressed some concerns about the premise of matching the complication measure with the PRO-PM and suggested consideration of length of stay measurements, which may result in better correlations. They also noted that the 1-day stays or the outpatient THA/TKA are not captured in either measure results and that may impact the results.
- Another TEP member suggested consideration of measures of length of stay or discharge to skilled nursing facilities for future validity testing.
- A TEP member asked if other disciplines may be reporting THA/TKA measures, for example, physical therapy, that CORE could potentially consider.
    - Ms. Zribi noted she was unaware of any measures and asked for input from the group.
- Another TEP member noted there was reporting for physical therapists in private practice but that stopped with the change to MIPS. Outpatient utilization (versus skilled nursing settings) may be correlated as it's used for THA/TKA patients. They planned to look further into physical therapy measures through the American Physical Therapy Association. Ms. Zribi thanked the TEP members for the great feedback.
- Another TEP member agreed with previous members' comments and was not surprised there were no correlations. Regarding the RSCR categories, they asked if there was a way to look at certain types of complications because certain outcomes such as a superficial wound that's taking an antibiotic such as Keflex would be captured in the National Surgical Quality Improvement Program (NSQIP) as an infection, but they're not infected. They noted the differences in severity of those patients with someone with a dislocation outside the complication measurement timeframe who may be dissatisfied and have lower

functional scores at the one-year follow-up timeframe. This TEP member asked, with limiting at 30-, 60-, 90-day timeframes, is there a way to parse out the types of complications that would potentially be pertinent to THA/TKA, and ones that are more medically related that would be somewhat temporary and then would improve?

- Dr. Suter thanked the TEP member and noted CORE could potentially look at mechanical complications and prosthetic joint infections. The good news is the complication measure eliminates those superficial Keflex complications. A patient must be readmitted and have a debridement or go to the operating room to be captured as a significant surgical bleed or a significant infection, so, there is a threshold of severity to be considered in the complication measure outcome. She noted that hearing from everyone about the discordance in timing, as well as other input to look at provider volume and try to better correlate the procedures with the timeline of the PRO-PM are really helpful suggestions.
- Another TEP member suggested separating mechanical complications and infections, as defined in THA/TKA complication measure. It would be very valuable to run those analyses separate from the other complications captured because the patient probably doesn't associate a heart attack that they recover 100% from with the outcome of their THA/TKA. By narrowing it down to mechanical complications and infections, this may show some correlation that otherwise doesn't show up.
- Dr. Suter thanked the TEP member for their helpful suggestion.

Summary of Measure Specifications

- Ms. Zribi reviewed the summary of the final measure specifications and reminded TEP members the measure specifications are aligned with hospital-level measure and reflect the feedback from collaboration with the TEP over the last year. The data sources of the measure include the patient-reported outcome measures (PROMs), Medicare administrative claims data, and additional risk variables collected such as health literacy and mental health. The cohort includes Medicare fee-for-service (FFS) patients aged 65 and older that undergo elective primary THA/TKA. She noted the current dataset is limited to inpatient procedures based on the CJR data. Based on TEP feedback from the last meeting and other stakeholder feedback, CORE plans to test the measure using THA/TKA procedures that occur in the outpatient and ambulatory surgical center (ASC) setting in the future. The measure outcome is attributed to a clinician or a clinician group, and this approach is based on the algorithm that is used by the existing clinician- and clinician group-level THA/TKA complication measure with a refinement to limit to clinicians that are responsible for the procedure. The two PROM surveys included in the measure are the Hip Dysfunction and Osteoarthritis Outcome Score for Joint

Replacement (HOOS JR.) and the Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS JR). Both of these tools are short, validated surveys that ask questions about patients' pain and function. PRO data collection timeframes preoperatively, are up to 90 days before a procedure and postoperatively, 10-14 months after the procedure. She noted the TEP members supported the shifted postoperative timeframe from the initial 9-12 months. The patient-level outcome is defined as the substantial clinical benefit (SCB), which is a threshold of 22-point improvement for hip patients using the HOOS, JR, and 20-point improvement for knee patients using the KOOS, JR

- Ms. Zribi reviewed the approaches to risk adjustment. The clinically derived and empirically tested risk model included 19 factors such as health literacy, back pain, pain in non-operative lower extremity joint, and the baseline Patent-Reported Outcomes Measurement Information System (PROMIS) Global Mental Health subscale score. CORE used the approach of stabilized inverse probability weighting to account for potential non-response bias. Lastly, she explained that the overall measure results assess the risk-standardized proportion of patients that undergo elective primary THA/TKA, who meet or exceed the SCB thresholds. The measure results could be calculated either at the clinician level or the clinician group level.
- Ms. Zribi reviewed the ongoing reevaluation and surveillance items that CORE plans to investigate in the future which are based on previous TEP input. CORE will be retesting the risk model prior to mandatory measure use, potentially through voluntary national data collection, and continue investigation of the relationship between social risk, response rates to surveys, and overall measure results over time. Finally, she noted CORE will assess the impact of including procedures performed in the outpatient and ASC settings, again, through possible voluntary national data collection.
    - A TEP member noted two observations. First, the mental health subscale score as defined using the PROMIS Global-10 or Veteran's RAND 12 Item Health Survey (VR-12) provided good risk adjustment for this measure and the hospital-level measure. Consistent with the original intent of using the HOOS, JR, KOOS JR, and the short Oswestry joint score to keep the clinician reporting burden down, it would be beneficial to only require reporting of the mental health subscale score instead of the entire PROM. Secondly, it would be valuable to incentivize providers to report PROM data through the American Joint Replacement Registry (AJRR) and make this fundamental, by rewarding within MIPS or quality points through CMS, such that reporting to AJRR becomes more routine. Some hospitals are already incentivized to participate in AJRR to qualify for being a center of excellence through The Joint Commission (TJC). This could help CORE and CMS because, within this specialty, participation in a joint registry database informs risk-adjustment variables that may be of value to refine and improve the measures.

        - Ms. Zribi thanked them for the valuable feedback and will share the suggestions on reporting burden and reporting through AJRR for the clinician-level measure with CORE's Strategic Planning Implementation team and CMS.

- Ms. Zribi explained TEP members would be polled to rate measure face validity and CORE planned to include their responses to these two questions in the NQF forms, as required for new measures going through initial endorsement. CORE will also include these results in the measure methodology report. TEP members were asked to rate their level of agreement on a six-point scale, ranging from strongly agree to strongly disagree, with the following two statements:
    2. The clinician-and clinician group-level hip/knee PRO-PM, as specified, will provide a valid assessment of improvement and functional status and pain following elective primary hip or knee replacements.
    3. The clinician- and clinician group-level hip/knee PRO-PM, as specified, can be used to distinguish between better and worse quality care among clinicians and clinician groups.
- Ms. Zribi thanked the TEP members for participation in the poll.
    - **Action Item:** CORE will follow up after the meeting with TEP members who attended the call by phone or who were unable to join the meeting to receive their face validity responses and any additional feedback.

Next Steps

- On behalf of CORE, Ms. Zribi thanked the group for their time and valuable feedback. She noted the CORE team will draft and distribute meeting minutes and TEP summary report in the coming weeks, and that CORE will request TEP review of the summary report before public posting. She noted this TEP is also supporting other orthopedic measures and teams at CORE. The CORE Episode Payment Model (EPM) THA/TKA Complication measure team will likely hold another TEP meeting with this group in spring 2022. Also, CORE's Strategic Planning Implementation team may contact the TEP for input on measure implementation topics later this year.
    - Dr. Suter thanked the TEP members for their important contributions and noted their work related to the hospital-level PRO-PM led to the measure being signaled by CMS in the Hospital Inpatient Prospective Payment System rule, and CMS proposed the measure with a postoperative PRO data collection window that better captures the 1-year follow up of 10-14 months.