Summary of Technical Expert Panel (TEP) Evaluation of Measures: Total Hip and/or Total Knee Arthroplasty (THA/TKA) Patient-Reported Outcome Hospital- and Eligible Professional-Level Performance Measures

March 7, 2014

This material was prepared by Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE) and Booz Allen Hamilton, under contracts to the Centers for Medicare & Medicaid Services (CMS) and the Office of the National Coordinator for Health Information Technology (ONC). The contents presented do not necessarily reflect CMS policy.

Contents

List of Tables

Background

Yale-New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (CORE) and Booz Allen Hamilton are under contract with the Centers for Medicare & Medicaid Services (CMS) and the Office of the National Coordinator for Health Information Technology (ONC), respectively, to develop patient-reported outcome (PRO)-based performance measures following total hip arthroplasty (THA) or total knee arthroplasty (TKA). Because the CORE and Booz Allen Hamilton teams are developing similar measures, the two teams are collaborating to choose which instruments to include in the measures, the timing of pre- and post-surgery administration of the instrument(s), and potentially on the definition of improvement. By collaborating, we hope to a) reduce provider reporting burden and patient burden and b) improve the feasibility and usability of these measures.

As part of this collaborative effort, CORE and Booz Allen Hamilton obtained expert and stakeholder input on the proposed PRO performance measures. The CORE and Booz Allen Hamilton measure development teams met regularly with their respective consultants and experts. Additionally, CORE and Booz Allen Hamilton convened a single national Technical Expert Panel (TEP) of clinicians, consumers, purchasers, and experts in quality improvement to provide input on key methodological issues.

This report summarizes the feedback and recommendations provided by the TEP regarding the proposed measures.

Measure Development Teams

The CORE and Booz Allen Hamilton measure development teams include measure development, clinical, statistical, policy, and project management experts who provide a broad range of perspectives and expertise. The team participates in all discussions and facets of measure development.

The CORE Development Team

The CORE new measure development team is led by Dr. Lisa Suter. Dr. Suter is a health services researcher, practicing rheumatologist and Assistant Professor of Medicine at Yale School of Medicine with experience in outcomes research and orthopedic outcome measure development specifically. See Appendix A for the full list of the CORE development team.

The Booz Allen Hamilton Development Team

The Booz Allen Hamilton new measure development team is led by Mike Sacca, Program Manager for Electronic Measure Development, and includes subcontractors from the National Committee for Quality Assurance (NCQA) and The Dartmouth Institute. See Appendix B for the full list of the Booz Allen Hamilton development team.

The Technical Expert Panel

In alignment with the CMS Measures Management System (MMS), CORE and Booz Allen Hamilton released a two week public call for nominations and convened a TEP. Potential members were solicited

via email per recommendations by the experts in the field, stakeholder groups, CMS hospital listservs, and through a posting on CMS's website.

The TEP was asked by the measure development teams to provide feedback on key methodological and clinical questions. The TEP is comprised of individuals with diverse perspectives and backgrounds and includes clinicians, patients, purchasers, and experts in quality improvement. The appointment term summarized in this document was through February 2014.

Specific responsibilities of TEP members included:

- Reviewing background materials provided by CORE and Booz Allen Hamilton prior to each TEP meeting
- Participating in all TEP meetings to the extent possible
- Providing input to CORE and Booz Allen Hamilton on key methodological, clinical, and other technical questions
- Providing feedback to CORE and Booz Allen Hamilton on key policy or other non-technical issues
- Reviewing TEP summary report prior to public release

Table 1. TEP Members

Name	Affiliation (Title)
Peter G. Allen, MS **	Food and Drug Administration (FDA) (Regulatory Scientist/Biomedical Engineer), Silver Spring, MD
David C. Ayers, MD	University of Massachusetts (UMass) Medical School (Professor of Orthopaedics), Worcester, MA
Thomas C. Barber, MD	Kaiser Permanente (Associate Physician in Chief), Oakland, CA
Daniel J. Berry, MD	Mayo Clinic (Chairman of Department of Orthopedic Surgery), Rochester, MN
Vinod Dasa, MD	Louisiana State University Health Sciences Center (Associate Professor, Department of Orthopaedic Surgery), New Orleans, LA
Cheryl Fahlman, PhD, MBA, BSP	Premier, Inc. (Principal Research Scientist), Washington, DC
Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN	Association of Rehabilitation Nurses; University of Massachusetts Amherst School of Nursing (Associate Professor), Greenfield, MA
Courtland G. Lewis, MD	Hartford Hospital (Director of Orthopaedic Surgery), Farmington, CT
Patient*	Patient
Michael H. Perskin, MD	The American Geriatrics Society; New York University School of Medicine (Associate Chair of Clinical Affairs and Assistant Professor in the Department of Medicine), New York, NY
Jonathan L. Schaffer, MD, MBA	The Cleveland Clinic Foundation (Managing Director, eCleveland Clinic Information Technology Division), Cleveland, OH
John H. Seiverd, PT, DPT, CCCE	James A. Haley Veterans' Hospital (Physical Therapy Center Coordinator of Clinical Education, Orthopaedic and Neurologic PT Residency Program Director), Tampa, FL
Lyle S. Sorensen, MD	Virginia Mason Medical Center (Chief of Orthopaedics and Sports Medicine), Seattle, WA
A. Christopher Strenta, PhD*	Dartmouth College (Associate Dean, Finance and Operations), Hanover, NH
Margaret A. VanAmringe, MHS	The Joint Commission (Vice President, Public Policy and Government Relations), Washington, DC

*Recent Recipient of a THA or TKA
**Observer only

TEP Meetings

CORE and Booz Allen Hamilton conducted two TEP meetings during the 2012-2013 contract year and have conducted one TEP meeting to date during the 2013-2014 contract year (see Appendix C for TEP meeting schedule). The TEP meetings followed a structured format consisting of presentation of key issues encountered in measure development and the teams' proposed approaches to addressing the issues, followed by open discussion of these issues by the TEP members. A high-level summary of the content presented to the TEP at the three TEP meetings is provided below and the specific recommendations of the TEP are summarized in Table 1 below.

During the first TEP meeting, the measure developers reviewed the measure development process and presented the goals of the measures, the importance for measuring patient-reported outcomes, and solicited feedback from the TEP about which PRO instruments would be most appropriate to use to define the measure outcomes.

- The measure developers summarized progress to date, which included preliminary definitions of the measure cohorts.
- Future measure development work not discussed during this TEP will include defining risk models for case mix adjustment and measure testing.
- The measure developers presented the following list of candidate PRO instruments that were identified through literature reviews and discussion with clinical and methodological experts as valid, reliable, and responsive assessments of patient-reported outcomes in patients undergoing THA/TKA:
    - Patient Reported Outcomes Measurement Information System (PROMIS) Global
    - PROMIS-29
    - EuroQOL-5D (EQ-5D)
    - MOS Short Form (SF)-36 and VR-36
    - MOS Short Form (SF)--12 and VR-12
    - MOS Short Form (SF)--8
    - CARE C – Section II: D Pain and E Mobility
    - Oxford Hip and Knee Scores (OHS and OKS)
    - Hip Osteoarthritis Outcome Score (HOOS) and Knee Injury and Osteoarthritis Outcome Score (KOOS)
    - Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)
- The measure developers then asked the TEP to provide feedback on the advantages and disadvantages of using these instruments to measure PROs in patients undergoing THA/TKA.
- The measure developers reviewed a list of questions that would be sent to the TEP after the first meeting using an online survey tool. The goal of the survey was to narrow the list of PRO instruments that would be used during measure development.

The second meeting focused on discussing the survey results, narrowing the selection of candidate PRO instruments for further testing to two generic and two condition-specific, and defining the timeframes for PRO instrument data collection pre- and post-surgery.

- CORE and Booz Allen Hamilton reviewed results from the PRO instrument survey that was administered after the first TEP meeting.
- The measure developers also presented options for the timing of pre- and post-operative PRO data collection. These options were based on timeframes suggested by published evidence and current programs that are measuring PROs in patients with THA/TKA.
    - The options for pre-operative data collection were:
        - Within three months prior to surgery
        - Within one month prior to surgery
    - The options for post-operative data collection were:
        - Three-six months after surgery
        - Six-nine months after surgery
        - Nine-12 months after surgery
- The measure developers also proposed that the TEP extend TEP membership to September 2014.
    - This would likely include two additional TEP meetings, most likely in the fall 2013 and spring/summer 2014.
    - The discussion topics could include:
        - Outcome definitions
        - Risk adjustment
        - Testing and validation of the measures
        - Public comment on the measures

During the third TEP meeting, the measure developers reiterated the goals of measure development and risk adjustment specifically. The meeting focused on selecting candidate risk variables for case mix adjustment and narrowing the list of candidate outcome definitions.

- The measure developers presented criteria for selecting candidate risk variables for risk model testing.
    - The criteria for selecting THA/TKA PRO-PM candidate risk variables were:
        - Evidence-based
        - Feasibility
        - Scientific validity and reliability
- The measure developers presented a list of candidate risk variables were presented that were tiered according to the selection criteria.
    - The three tiers of candidate risk variables were indicated as the following:
        - High priority
        - High priority, but not feasible now

        - Low priority
- The measure developers then asked the TEP to provide feedback on specific risk variables that would be most important to test for case mix adjustment.
- The measure developers also presented a list of candidate outcome definitions.
    - The list of candidate outcome definitions included:
        - Mean post-surgery PROM score
        - Mean change in PROM score
        - Post-surgery PROM score threshold
        - Mean change in PROM score threshold
        - Minimal clinically important difference
        - Minimal clinically important improvement
        - Patient acceptable symptom state
- The measure developers then asked the TEP to provide feedback on the usability and interpretability (for patients and surgeons) of each outcome.
- The measure developers informed the TEP that a follow-up survey on candidate risk variables and outcomes will be distributed for completion.
- The measure developers also proposed that the TEP meet again in the summer of 2014 to discuss results of measure testing.

Table 2. Key Issues Discussed and TEP Feedback

Key Issues Discussed	TEP Feedback
Goals of the TEP The goals of the TEP included recommending generic and condition-specific patient reported outcome (PRO) instruments for both the hospital- and physician-level measures, to discuss PRO data collection timing for the measures, to recommend outcome definitions for the measures, and to discuss risk adjustment variables for the measures.	The TEP was supportive of the goals of the TEP, the TEP charter and timeline, and the importance of the measures under development.
Measure Development Progress To Date The measure developers proposed harmonizing cohort definition by excluding revision and non-elective THA/TKA procedures, such as those associated with fractures. Booz Allen Hamilton will separate THA and TKA procedures into two independent measures; CORE had not finalized their approach but acknowledged the importance of the different rehabilitation patterns of these two procedures.	While discussing the goals and decisions in measure development so far, a TEP member questioned whether the focus of the measures would include only primary total joints or if partial knee replacements would be included. The measure developers clarified that the intention was to include only total joint arthroplasty procedures, although the cohort specifications have not yet been finalized and it remains unclear if these procedures can be differentiated. The TEP was supportive of excluding non-elective and revision procedures from the measure cohorts and separating THA and TKA procedures.
Performance Measures and PRO Instruments The measure developers described that performance measures can assist in clinical care decision-making as well as informing quality improvement efforts. **Importance of measuring patient-reported outcomes for patients undergoing TKA/THA** CORE and Booz Allen Hamilton explained the importance of measuring patient-reported outcomes (PROs). PROs allow patients to report their experience with their health and health care directly. Patient experience includes, pain, function, satisfaction, and quality of life. The measure developers went on to explain that total hip and total knee replacements are very common procedures with over 500,000 procedures performed annually among Medicare beneficiaries. The	The TEP was supportive of clinical quality measures that evaluate PROs for patients undergoing THA/TKA.

Key Issues Discussed	TEP Feedback
development teams added that other countries have been measuring these outcomes for some time and their experiences confirm that there is variation in these outcomes not entirely explained by patients' clinical characteristics.	
Selection of PRO instruments CORE and Booz Allen Hamilton presented a summary of the evidence for a list of candidate generic health and condition-specific PRO instruments. The measure developers sought feedback from the TEP on the burden of using PRO instruments in patient care and for quality measurement (burden on both the patient and data collection level). The measure developers clarified that both projects would like to collect postoperative data collected after the patient is discharged as this better reflects the functional results of THA/THA surgery. To reduce burden on the TEP members, the measure developers asked the TEP if there were any PRO instruments that could be removed before the survey was distributed. The TEP asked to remove the EQ-5D and the CARE-C instruments as they lacked data regarding mental and emotional health. The TEP also requested the Oxford instruments be removed from the survey as they do not separate pain and mobility symptoms and as such offer less clinical utility than other instruments.	The TEP raised a concern that PRO measurement is a moving target as the field is quickly evolving and new instruments (or revised versions of existing instruments) are being released. The TEP emphasized that PRO instruments for the THA/TKA population need to be able to distinguish between pain and function scores. The TEP was supportive of using PRO data to inform patient care (both pre-operatively and post-operatively). The TEP suggested that pre-operative scores could be used by physicians for risk adjustment as well as to identify patients who might need additional peri-operative care or services. The TEP members discussed burden on physicians' offices as these measures would be collected by physicians post-surgery. The TEP cautioned that patients would be the ones to fill out these instruments and that too many questions could be a burden on them. One TEP member added that there could be regional and socioeconomic differences in filling out these instruments as literacy and language issues could vary. The TEP agreed that pre- and post-operative PRO data collection for this measure should employ both a generic instrument that assesses mental and emotional health and overall quality of life in addition to a condition-specific instrument that assesses outcomes relevant to THA/TKA, such as pain and mobility.
Review of PRO survey results The measure developers reviewed the criteria for selecting a PRO instruments which included: • Provides meaningful information • Useful for care, quality improvement, and	The TEP suggested that if Computer Adaptive Testing (CAT) is available, they would recommend the PROMIS-29 over the PROMIS Global. However, since CAT is not available for these measures, the TEP preferred the PROMIS Global over the PROMIS-29 (due to its brevity). The TEP was also concerned that many of the questions on

Key Issues Discussed	TEP Feedback
performance assessment • No undue burden • TEP supportive The discussion included reviewing results from the PRO instrument survey that was distributed to the TEP members after the first TEP meeting. Using these survey results, the measure development teams facilitated a discussion for recommendations for the following instruments: • Generic PRO instruments: o PROMIS Global o PROMIS-29 o SF-36/VR-36 o SF-12/VR-12 o SF-8 • Condition-specific instruments o Oxford Hip/Knee Score o HOOS/KOOS o WOMAC	the PROMIS-29 were duplicative of questions in the condition-specific instruments. The TEP theorized that familiarity with the SF instruments may have been an explanation for the better survey results for the SF instruments when compared to the VR instruments, despite these instruments being essentially identical in content. TEP members expressed a preference for instruments that were no or low cost to physicians and that were easy to license. One TEP member added that the SF instruments are actually more understandable to those with lower literacy levels. Next, the TEP discussed the condition-specific instruments. The TEP reiterated that it was important to distinguish between pain and physical function with these instruments, and expressed concern with using the Oxford Hip and Oxford Knee scores. The TEP also added that it would not make sense to include both the HOOS/KOOS instruments and the WOMAC for further testing, since the HOOS/KOOS contain the WOMAC content. The TEP suggested including the non-proprietary PROMIS Global and VR-12 instruments in further measures testing efforts. The TEP recommended including the HOOS/KOOS condition-specific instruments in further measure testing.
PRO Pre-Operative Data Collection Timing The measure developers presented the **pre-operative timing** research and options for consideration. The pre-operative options included: • Within 3 months prior to surgery • Within 1 month prior to surgery The measure developers clarified that both the condition-specific and generic health instrument would be collected during the same time frame.	The TEP was in favor of the 3 month timeframe as it would provide flexibility for smaller hospitals or practices in collecting this data. The TEP did not believe there would significant differences in score results collected three months prior to surgery compared to results collected one month prior to surgery. The TEP recommended using the 3-month pre-operative timing as this could be used to evaluate patients prior to surgery and would give flexibility for institutions collecting this data.

Key Issues Discussed	TEP Feedback
PRO Post-Operative Data Collection Timing The measure developers also presented **post-operative timing** option for consideration including: • 3-6 months after surgery • 6-9 months after surgery • 9-12 months after surgery	TEP members felt that 3 months was too early for post-operative data collection as individuals may recover at different speeds, and there is clinically significant improvement that occurs after three months. However, one TEP member cautioned that providers in less-advantaged areas may experience much lower response rates after three months. Many TEP members agreed that a 12-month follow-up period for data collection would be ideal for the outcomes being measured. While some felt that 6 months would be an adequate data collection time, 12 months would be preferred. The TEP discussion varied; however, there seemed to be consensus that the best timeframe for post-operative data collection would be between 6 and 12 months.
Candidate Risk Variables The measure developers presented a list of candidate risk variables were presented that were tiered according to the selection criteria in the following groups: • High priority • High priority, but not feasible now • Low priority	The TEP expressed concern that some risk variables with important face validity to orthopedists were not being sufficiently prioritized due to the absence of available data for measure testing and development. The TEP expressed support for further data collection to examine these risk variables.
Candidate Outcome Definitions The measure developers also presented a list of candidate outcome definitions. The list of candidate outcome definitions included: • Mean post-surgery PROM score • Mean change in PROM score • Post-surgery PROM score threshold • Mean change in PROM score threshold • Minimal clinically important difference • Minimal clinically important improvement • Patient acceptable symptom state	The TEP recommended eliminating mean post-surgery PROM score and post-surgery PROM score threshold from further consideration as these outcome definitions do not assess the change in PROM score before and after surgery. The TEP also agreed with eliminating the patient acceptable symptom state from further consideration as it requires additional data collection.
Next Steps: Next steps will be to continue with measure development. This will include finalizing measure	The TEP members expressed interest in meeting again after the developers conduct measure testing and produce results.

Key Issues Discussed	TEP Feedback
specifications and testing the measures. CORE will hold an interim public comment on the hospital-level measures.	

Conclusion

TEP feedback was instrumental in shaping the approach to measure development. Their input allowed us to narrow the list of acceptable PRO instruments, informed our choice of pre- and post-operative survey timeframes, and narrowed the potential ways we will use the PRO survey results to calculate the outcome. In addition, CMS, ONC and the measure developers heard the concerns of the TEP regarding the burden to patient, surgeons, and hospitals of collecting PROs as well as the need for additional data collection to allow evaluation of a more comprehensive list of candidate risk variables for risk adjustment. We will consider these concerns carefully as measure development proceeds.

CORE and Booz Allen Hamilton will continue to consult with both expert consultants and the TEP as the measures are further developed and refined.

Appendix A. CORE Measure Development Team

Table 3. CORE Team Members

Name	Title/Affiliation	Contact Info
Harlan Krumholz, MD, SM	Director, CORE	harlan.krumholz@yale.edu
Lisa Suter, MD	Measure Lead / Associate Director	lisa.suter@yale.edu
Zhenqiu Lin, PhD	Lead Analyst	zhenqiu.lin@yale.edu
Elizabeth Drye, MD, SM	Director, Quality Measures	elizabeth.drye@yale.edu
Susannah Bernheim, MD, MHS	Director, Quality Measures	susannah.bernheim@yale.edu
Michael Araas, MPH	Project Coordinator	michael.araas@yale.edu
Weiwei Zhang, MPH	Analyst	weiwei.zhang.wz227@yale.edu
Rana Searfoss, BA	Research Assistant	rana.searfoss@yale.edu

Table 4. CORE Working Group Members

CORE Working Group Members	Title/Affiliation
Kevin Bozic, MD, MBA	William R. Murray, MD Endowed Chair in Orthopaedic Surgery Professor and Vice Chair University of California, San Francisco Department of Orthopaedic Surgery Core Faculty, Philip R. Lee Institute for Health Policy Studies Visiting Scholar, Harvard Business School

Appendix B. Booz Allen Hamilton Measure Development Team

Table 5. Booz Allen Team Members

Booz Allen	Title/Affiliation	Contact Info
Mike Sacca	Program Manager, Electronic Clinical Measure Development	Sacca_Michael@bah.com
Pamela Edison, MHS, PMP	Deputy Project Manager, Electronic Clinical Measure Development	Edison_Pamela@bah.com

Table 6. NCQA Team Members

NCQA	Title/Affiliation	Contact Info
Phyllis Torda, MA	Vice President, Quality Solutions Group	Torda@ncqa.org
Bob Rehm, MBA	Assistant Vice President, Performance Measurement	Rehm@ncqa.org
Jenna Williams-Bader, MPH	Assistant Director, Performance Measurement	Bader@ncqa.org
Daniel Roman	Senior Healthcare Analyst, Performance Measurement	Roman@ncqa.org
Kat Sobel	Healthcare Analyst, Performance Measurement	Sobel@ncqa.org

Table 7. Dartmouth Institute for Health Policy and Clinical Practice Team Members

Dartmouth Institute for Health Policy And Clinical Practice	Title/Affiliation	Contact Info
Melanie Mastanduno, BSN, MPH	Director, Population Health Measurement Center for Population Health	Melanie.P.Mastanduno@hitchcock.org
Eugene Nelson, DSc, MPH	Director Population Health Measurement Program Director Population Health and Measurement Dartmouth-Hitchcock Medical Center	Eugene.C.Nelson@hitchcock.org

Appendix C. Technical Expert Panel Call Schedule

TEP Meeting #1

Tuesday, July 23, 2013, 2:00-4:00 PM EST

TEP Meeting #2

Tuesday, August 27, 2013, 5:00-7:00 PM EST

TEP Meeting #3

Friday, January 17, 2014, 3:00-5:00 PM EST

Appendix D. TEP Meeting #1 Minutes

Patient-Reported Outcome (PRO) Following Elective Total Hip and/or Total Knee Arthroplasty: Hospital- and Eligible Professional-level Performance Measures

Technical Expert Panel (TEP)

Summary Call #1

Tuesday, July 23, 2013, 2:00-4:00 pm ET

Participants

TEP: Peter G. Allen, MS; David C. Ayers, MD; Vinod Dasa, MD; Cheryl Fahlman, PhD, MBA, BSP; Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN; Courtland G. Lewis, MD; Patient; Michael H. Perskin, MD; John Seiverd, PT, DPT, CCCE; Lyle Sorensen, MD; A. Christopher Strenta, PhD; Margaret VanAmringe, MHS;

Booz Allen Hamilton/National Committee for Quality Assurance (NCQA)/Dartmouth Institute for Health Policy & Clinical Practice Overview: Mike Sacca; Pamela Edison, MHS, PMP; Melanie Mastanduno, BSN, MPH; Jenna Williams-Bader, MPH;

Yale New Haven Health Services Corporation – Center for Outcomes Research (CORE): Elizabeth Drye, MD, SM; Lisa Suter, MD; Zhenqiu Lin, PhD; Susannah Bernheim, MD, MHS; Jaymie Potteiger, MPH; Kanchana Bhat, MPH; Rana Searfoss, BA; Smitha Vellanky, MSc; Lori Schroeder, JD, LLM; and Kevin Bozic, MD, MBA;

The Centers for Medicare & Medicaid Services (CMS): Lein Han, PhD; Karen Nakano, MD, MS; Elizabeth Ricksecker, MA

The Office of the National Coordinator for Health Information Technology (ONC): Kevin Larsen, MD; Lauren Richie, MA; Julie Grouse; and Corette Byrd

SUMMARY	Action items
• The TEP was introduced to the projects, goals of the meetings, and the measure developers • The TEP approved the TEP charter and reviewed development milestones • The team presented why Patient-Reported Outcome (PRO) measures are important and the NQF criteria that should be used to determine an appropriate instrument • The TEP reviewed the PRO instruments under consideration for the performance measures	**Submit survey results by Monday, August 5th at 5:00 EST**

	GENERAL
Welcoming Remarks and Introductions	Jaymie Potteiger, MPH gave a brief introduction for the projects. The Centers for Medicare and Medicaid Services (CMS) has asked CORE to develop one to two measures for the total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) using patient-reported outcomes for potential use in performance measure reporting on hospital quality. The Office of National Coordinator for Health Information Technology (ONC) has contracted with Booz Allen Hamilton to develop two PRO-based electronic clinical quality measures to assess improvement following THA or TKA that can be used for eligible professional-level performance measurement. Ms. Potteiger asked that everyone at the meeting keep all personal opinions and experiences confidential. She reviewed the agenda which included an overview of measure development work, introductions from the TEP, summary of TEP role and TEP charter, brief description of the measure development timeline, and a discussion of the PRO instruments. Dr. Lein Han, the Government Task Leader (GTL) for the CORE contract gave an introduction to the CMS project. Dr. Han described the difference between the CMS and ONC projects stating that CMS would like to develop a risk-adjusted patient self-reported measure for profiling hospital performance. She added that this will be developed during two phases over a two year timeline. Dr. Kevin Larsen, Medical Director for Meaningful Use, gave an introduction to the ONC project. Dr. Larsen added that from the ONC perspective, the goal of the joint TEP was to have as much alignment between the two projects as possible. Mike Sacca, Program Manager for ONC HITECH at Booz Allen Hamilton, introduced the measure development teams. These teams included the Center for Outcomes Research (CORE), Booz Allen Hamilton, NCQA, and Dartmouth.
Goals of the TEP and TEP introductions	Dr. Suter reviewed the goals of the TEP meetings from slide six: • Recommendations regarding the generic and condition-specific patient reported outcome (PRO) instruments for use in the hospital- and provider-level measures • Recommendations regarding preoperative and postoperative data collection timing (2nd meeting) • Recommendations about the outcome definition (2nd meeting) o Dr. Suter clarified that this would include recommendations of exactly how to measure change in functional status following surgery, such as the number of patients that achieve a minimally important difference • Discussing risk adjustment and issues specific to eligible professional-level Electronic Health Record (EHR) measures o This may be scheduled for a 3rd TEP meeting in September or August Dr. Suter asked each TEP member to introduce themselves, listing their organization or affiliation and disclosing any conflict of interest that may have changed since they submitted a

	TEP application. There were no new conflicts identified.
TEP Charter and Timeline	Rana Searfoss, BA, explained the TEP role. The purpose of the TEP is to provide stakeholder and technical input. This process ensures transparency through the measure development process. Ms. Searfoss described the purpose of the TEP charter which outlines the responsibility of the TEP. The TEP approved the TEP charter. Ms. Searfoss described the measure development timeline. She added that the TEP would soon be sent a survey regarding the PRO instruments up for discussion. She said there would be a TEP meeting in August. CORE's recommendations to CMS are due September 14, but measure development will continue past that date.

Topic	DISCUSSION	Action item
Development Milestones	Dr. Suter described the research processes for these measures so far, which included systematic literature reviews and discussions with clinical experts, and structured interviews with several providers. Dr. Suter added that there appears to be no consistent use of PRO instruments, data collection, or timing for data collection. She said that many groups are using more than one instrument to collect data.	None
Previous Key Decisions	Dr. Suter described the key decisions so far in measure development. The teams have agreed to focus on elective total hip and total knee replacement. She added that they will also exclude patients who have fractures or mechanical complications. **One TEP member asked if these measures could focus on only primary total joints and asked about partial knee replacements.** Dr. Suter responded that this was the intention. She added that with ICD-9 codes, it is difficult to differentiate some procedures. Dr. Suter asked if there were strong feelings about the cohort to please email the teams at cmshipkneeprom@yale.edu. Dr. Suter said that the ONC measures will separate total hip and total knee procedures into two different measures because of the difference in rehabilitation. The CMS measures may measure these separately but combine scores due to sample size. Dr. Suter added that there are many measure development steps that will be taken after the TEP meetings. This future work includes defining the risk adjustment models.	None

	Dr. Suter added that patient-reported outcomes are important because they can assist in clinical care decision-making as well as inform quality improvement efforts. She said they can be used to monitor medical devices and can evaluate the quality of care provided to patients. She added that this measurement is about the patient experience which includes pain, function, satisfaction with the outcomes of care, and quality of life. Dr. Suter discussed the importance of the patient-reported experience as it comes directly from the patient and is not filtered or interpreted through a provider. She added that the data suggested that better PRO quality of life is associated with better survival and with lower future healthcare costs for some conditions. Total hip and total knee replacements are very common procedures with over 500,000 procedures performed annually among Medicare beneficiaries. She added that other countries have been measuring these outcomes for some time and their experiences confirm that there is variation in outcomes.	

Topic	DISCUSSION	Action item
NQF Criteria	Ms. Searfoss discussed the National Quality Forum (NQF) criteria for the performance measures. She explained that the NQF is a national organization that endorses different types of measures, including process and outcome measures. She added that there are four NQF criteria that are described on page 12 of the TEP packet. These criterion include: • Impact, opportunity and evidence • Reliability and validity • Usability • Feasibility One TEP member asked for an expanded explanation regarding the NQF criteria for feasibility. Dr. Suter responded that the NQF criteria are designed to guide NQF endorsement. Feasibility includes avoiding undue burden which addresses a number of perspectives including patient burden, but also the burden of the healthcare provider.	None
PRO Instruments Background	Jenna Williams-Bader, Assistant Director, Performance Measurement at NCQA, introduced the list of PRO instruments under discussion for the TEP meeting. She described the criteria that should be used in choosing the PRO	None

Topic	DISCUSSION	Action item
	instrument. Next, she discussed the high-level questions for discussion regarding the PRO instruments. • Generic health vs. condition-specific • Recommendations on generic health vs. condition-specific for hospital- or eligible professional-level • Which instrument do you recommend She added a brief summary of how generic instruments cover a broad range of patient's health, which includes both mental and physical health. However, they may be less responsive to change from a specific intervention. She discussed the issue of burden which can be a balance between collecting more information from the patient and provider standpoint, and putting too much burden on the provider and patient. **One TEP member asked about the sampling methodology that was going to be used for the performance measures.** Dr. Suter responded that at this point in the development process, the teams have not yet made decisions regarding sample size. She added that the goal of the meeting was to identify possible PRO instruments that would be valuable for measurement. Dr. Larsen added that for the Meaningful Use program, all patients in a single surgeon's practice who meet the criteria are included in the measure denominator. He added that this is typically an all-patient in a measurement year program measure. **One TEP member added that sample size is an important aspect to consider, especially when discussing burden.** **Another TEP member added that data collection for this type of information will happen in the physician's office and this could affect the patient flow and staffing of the office.** Dr. Kevin Bozic, a working group member for CORE, added that this call should focus on the PRO strengths and limitations. He added that there will be plenty of time to comment on strategies for implementation going forward.	
PRO Instruments: Details	Ms. Williams-Bader reviewed Table 1 at a high-level, beginning with the generic PRO instruments. She added that the teams have focused on instruments that are reliable and valid and that the instruments have many similarities.	None

Topic	DISCUSSION	Action item
	Next, she discussed the responsiveness for the PRO instruments. While there are some instruments without data for this population under discussion, all instruments are considered responsive. She added that 'floor effect' meant that there would be patients entering at the bottom of the scale and 'ceiling effect' would mean that there would be patients at the very top of the scale. She added that floor and ceiling effects could indicate several things, but wouldn't necessarily mean that an instrument isn't a good measurement tool. **One TEP member asked a question regarding electronic versions of PRO instruments and if those versions of the tools would be considered in this discussion.** Dr. Suter responded that the instruments being discussed today were paper, telephone, or online surveys, but they are a fixed set of questions. **Another TEP member discussed that this type of project would be a moving target since many groups are working to improve PRO instruments. The TEP member added that the critical criteria that this work should consider would be cost, accessibility, ease of use, and burden of implementation.** **A TEP member discussed the need to have a comparative group for the performance measures. He added that this should include information on how patients with a wide spectrum of disease score on the performance measures.** **A TEP member also added that it is important for tools to be able to distinguish between pain and function scores, which some tools may not be able to do.** **Another TEP member added that it is important to establish the baseline for the population from the beginning.** **Another TEP member added that the information gained from preoperative scores, such as anxiety or depression, could be very important before actually doing surgery on a patient.** **A TEP member added that it was important to remember that even for the hospital-level information, physicians would be collecting the PRO data as there would not be useful information directly after the surgery in the hospital.**	
	Dr. Suter responded that both ONC and CMS are looking at downstream post-operative data collection, not inpatient data collection. She added that the mechanism for collection is still to be determined.	

Topic	DISCUSSION	Action item
	Dr. Suter also added that the CMS measure these measures will likely measure relative performance, giving an example of the current total hip and knee readmission measures that allow hospitals to see and understand readmission rate differences compared to the national average. Dr. Suter then asked the TEP if it would be sufficient to collect a generic instrument preoperatively and collect a condition-specific instrument preoperatively and postoperatively. **One TEP member responded that they would like to see both a generic health and condition-specific instrument collected so that the field can learn more about recovery as technology changes.** **Another TEP member added that the impact of these procedures is such that without a condition-specific measure, it may be hard to differentiate between the impact of the procedure. This TEP member then added that, from a surgeon's standpoint, 90 days would be a global period for payment but that additional time may be required for an optimal surgical response to be achieved.** **Another TEP member added that it would be best to collect both a generic and condition-specific instrument, in an ideal world, adding that knowing emotional health is very important for risk stratification.** **An additional TEP member said that psychosocial factors on the generic instruments touch more broadly on the issues of mental and emotional health that were discussed during the meeting. This TEP member agreed strongly with the use of a generic instrument.** **Another TEP member added that the information contained in generic instruments, that address overall quality of life and mood, is more relevant to patients than the information contained in the condition-specific instruments.** **An additional TEP member agreed with the importance in understanding the overall condition of the patient.** **One TEP member added that it is important for everyone to keep in mind that a patient would need to actually answer all of the surveys which could represent considerable burden.** **Another TEP member added that it appears generic questionnaires are more attuned to patient's views while the condition-specific instruments may be what the surgeons find important.** **One TEP member added that they agreed with the concept that both instrument types are important. They added they felt that a patient's willingness to fill out the survey would be tied directly with the**	

Topic	DISCUSSION	Action item
	relationship to the physician or physician's assistant. **Another TEP member said that there may be some regional differences for filling out the survey due to educational levels and other factors. They added that there will be socioeconomic status issues with the modes for filling out these instruments.** **Another TEP member added that, from experience, there are many issues that may make instruments complex such as literacy issues or language issues.** **One TEP member added that, from experience, 20% of English speaking patients asked for assistance for the SF-36, WOMAC, and the EQ-5D. The TEP member added that, in their experience, such patients did not necessarily have worse clinical outcomes than individuals with no difficulty filling out PRO instruments.** Dr. Suter added that if we were to institute a performance measure that collected patient-reported outcome data, and there was a systematic group of patients that were unable to consistently fill out the surveys, if that patient population overall was at greater risk for poorer outcomes, it may impact the performance measurement of the surgeon or hospital being assessed. **One TEP member added that in underserved patient populations, many do not get physical therapy. They added that many patients cannot afford it. Therefore, the TEP member expressed concerns that this may make their hospital look worse based on the patient population.** **One TEP member expressed that the SF instruments are proprietary, which put them at a disadvantage, while the VR and PROMIS instruments are not proprietary. However, they said that the SF instrument has more data and has been used for a wide spectrum of disease. They added that once PROMIS receives more appropriate validation, this issue could be corrected.** **One TEP member asked if the VR surveys were moving towards the CAT type of environment.** The team members responded that they were unaware of this. Dr. Suter added that prior to the TEP call, the teams reached out to all of the PRO instrument developers for the instruments discussed today. The teams asked for information on their measures and many participated and shared insights, references, and thoughts on the instruments. She added that the teams did not hear a response for the VR measures. Dr. Suter asked if there were any strong feelings for the condition-specific	

Topic	DISCUSSION	Action item
	instruments that were presented in the materials. **One TEP member added that the WOMAC is probably comparable to the SF-12 and SF-36 in terms of being widely used and tested, but has the disadvantage of being proprietary.** **Another TEP member added that unless the developer of the WOMAC is bought out, making arrangements to use this tool has been a challenge. This TEP member also added that the HOOS and KOOS were fairly easy to use.** **One TEP member expressed concern about the number of questions patients will need to fill out if we require both a generic and condition-specific instrument.** **Another TEP member added that it was important to include that there is ongoing research to combine a generic instrument and a condition-specific instrument into one instrument.** **One TEP member added that different populations may score differently on the generic vs. condition-specific instruments, showing that there may be different items that are not being captured in specific instruments that are needed and some that are not needed. The TEP member added it would be important for research to be done to streamline this effort.** **A TEP member added that it would be helpful to have validity data for all of the instruments shown.** Dr. Suter said that there would be validity testing of the performance measures and that the reason ONC and CMS are considering existing patient-reported instruments was that it was a priority to collect data using already validated surveys. She added that when you pull certain questions out of a survey, in some ways, you can potentially invalidate the instrument. **One TEP member added that the surveys did not ask the question "would you recommend this surgeon/hospital for this procedure to someone else."** **Another TEP member said that they have concerns about the timeframe for responding to combined instruments with questions for pain and functional status. This TEP member elaborated that for measuring pain, one month could be fine, while functional status may need at least one to three months to improve. However, assessing quality of life would need to be nine to 12 months after surgery to be adequate. The TEP member added that this could be difficult when adding PRO instruments together.**	

Topic	DISCUSSION	Action item
PRO Instruments: Selection for Survey	Dr. Suter described the survey that was sent to the TEP members. She said it is nine questions for each instrument and asked if the TEP members would vote to remove any instruments that were not wanted on the survey. The TEP members decided to remove the following instruments from the survey: • EQ-5D • CARE-C • Oxford Hip and Knee Scores Dr. Susannah Bernheim, the Director at CORE, suggested that, when thinking about these instruments, to notice if the questions being asked have to do with importance, meaningfulness, and understandability of these instruments for the purpose of evaluating care. She added that CMS and ONC understand that the implementation challenges are not small, but that it is a unique opportunity to help in the early phases of deciding how to gather this information. Dr. Suter added that there would be comment boxes for each instrument to provide any other opinions that may not be reflected in the survey questions. Ms. Potteiger reviewed the next steps for the TEP which included the completion of the TEP survey and preparation for the 2nd TEP meeting. During the 2nd TEP meeting, the discussion will be to review survey results, begin discussions for timing and outcome definitions. Ms. Potteiger ended the call by thanking all of the TEP members for their time and feedback on behalf of the teams at Booz Allen Hamilton and CORE.	None

	NEXT STEPS	
Next steps	• TEP members to complete PRO instrument survey by **Monday, August 5th at 5:00 EST** • The teams will send a Doodle poll to the TEP for scheduling the 2nd TEP meeting	

Appendix E. TEP Meeting #2 Minutes

Patient-Reported Outcome (PRO) Following Elective Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA): Hospital- and Eligible Professional-level Performance Measures

Technical Expert Panel (TEP)

Summary Call #2

Tuesday, August 27, 2013, 5:00-7:00 pm ET

Participants

TEP: Peter G. Allen, MS; David C. Ayers, MD; Thomas Barber, MD; Daniel Berry, MD; Vinod Dasa, MD; Cheryl Fahlman, PhD, MBA, BSP; Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN; Courtland G. Lewis, MD; Patient; Michael H. Perskin, MD; Jonathan Schaffer, MD, MBA; John Seiverd, PT, DPT, CCCE; Lyle Sorensen, MD

Booz Allen Hamilton/National Committee for Quality Assurance (NCQA)/Dartmouth Institute for Health Policy & Clinical Practice Overview: Mike Sacca; Pamela Edison, MHS, PMP; Melanie Mastanduno, BSN, MPH; Kathy Carluzzo; Phyllis Torda, MA; Bob Rehm, MBA; Jenna Williams-Bader, MPH; Daniel Roman; Kat Sobel

Yale New Haven Health Services Corporation – Center for Outcomes Research (CORE): Lisa Suter, MD; Elizabeth Drye, MD; Zhenqiu Lin, PhD; Jaymie Potteiger, MPH; Rana Searfoss, BA; Kevin Bozic, MD, MBA

The Centers for Medicare & Medicaid Services (CMS): Lein Han, PhD; Elizabeth Ricksecker, MA

The Office of the National Coordinator for Health Information Technology (ONC): Kevin Larsen, MD, Lauren Richie, MA

SUMMARY	Action items
• The team discussed results from the PRO instrument survey • The TEP supported the use of separate assessments of pain and function • The TEP supported the use of the Patient Reported Outcomes Measurement Information Systems (PROMIS) Global instead of the PROMIS-29, as well as the Veterans Rand 12 Item Health Survey (VR-12) instead of the Veterans Rand 36 Item health Survey (VR-36), because of the reduced burden in data collection • The TEP supported the use of non-proprietary tools, such as the VR-12, instead of the 12-Item Short Form Health Survey (SF-12)	**Review TEP Summary report**

SUMMARY	Action items
• The TEP supported the use of the Hip dysfunction and Osteoarthritis Outcome Score (HOOS) and Knee injury and Osteoarthritis Outcome Score (KOOS) as the only condition-specific assessment instruments • The TEP supported the use of a 3-month pre-operative assessment timeframe • Some TEP members supported the use of a 6-month post-operative assessment timeframe whereas others favored a 12-month post-operative assessment timeframe • The TEP approved a suggestion to consider extending TEP membership through September 2014 • (Note: Below, "provider" indicates an eligible professional, which in the case of THA/TKA would likely be the surgeon)	

	GENERAL
Welcoming Remarks and Introductions	Jaymie Potteiger, Project Coordinator at CORE, gave a brief introduction for the projects. CMS has asked CORE to develop one to two measures for THA and/or TKA using PROs for potential use in performance measure reporting on hospital quality. ONC has contracted with Booz Allen Hamilton to develop two PRO-based electronic clinical quality measures to assess improvement following THA or TKA that can be used for eligible professional-level performance measurement. Ms. Potteiger asked that everyone at the meeting keep confidential all personal opinions and experiences shared with the TEP. She reviewed the agenda, which included a discussion of the results of the PRO survey, a discussion of options for timing of administration of PRO instruments, and a review of post-meeting action items. She discussed the meeting objectives, which included finalizing recommendations on PRO instruments for testing, reviewing results of the PRO survey, discussing timing of pre-operative and post-operative PRO assessments, and extending TEP membership through September 2014. Dr. Kevin Larsen, Medical Director for Meaningful Use, reviewed the objective of this TEP meeting: to select instruments and pre-operative/post-operative timing for assessments of patients who have had THA or TKA, which will be used for quality measures at the provider (eligible professional) and hospital level. Alignment of measure development at these two levels of measurement will reduce burden by avoiding duplicative assessments for different measurement purposes.

Topic	PRO Instruments and Timing	Action item

Topic	PRO Instruments and Timing	Action item
PRO Survey Recommendation	Dr. Lisa Suter, Associate Director at CORE, reviewed the criteria for selecting PRO instruments: that they provide meaningful information, are useful for care and performance assessment, do not place undue burden on anyone in the care process, and have been supported for use in these quality measures by the TEP. Dr. Suter also reviewed the questions included in the survey for evaluating the candidate PRO instruments (TEP completed surveys prior to this meeting). She emphasized that the ultimate goal of these projects is to highlight variations in quality of care, and that provider-level data will be aggregated to provide data at the hospital level. Dr. Suter listed the generic and condition-specific PRO instruments under consideration. She reminded the panel that our objective is to ultimately recommend only two generic and two condition-specific instruments for testing. Of the two generic instruments, the TEP should select one of the PROMIS instruments, in order to include a generic instrument that has the ability to incorporate computer adaptive testing (CAT), as well as one of the Short Form (SF)/Veterans Rand (VR) instruments. Dr. Suter sought input on the survey results regarding the PROMIS instruments. On a question regarding the use of instruments for informing quality improvement, the PROMIS-29 received more negative than positive responses. Dr. Suter also clarified that this question aims to assess whether the instrument provides information that is useful for quality improvement rather than performance assessment at the physician or hospital level. **A TEP member expressed concern that while more information can be obtained from lengthier instruments, their practicality without CAT is limited. The PROMIS-29 may have better statistical precision as well as subscores, which the PROMIS Global does not have.** **A TEP member stated that pain intensity questions in the PROMIS-29 may be duplicated in the condition-specific instruments.** **A TEP member stated that the objective would be to capture patients who are major outliers after surgery. This TEP member spoke in favor of using the PROMIS Global, in spite of the lower precision, because of easier implementation. Another TEP member stated that these measures would be looking at system-level changes where additional precision would not be helpful.** **A TEP member cautioned that there is a bimodal, as opposed to normal, distribution of post-operative scores on assessments such as the EQ-5D; an average may not demonstrate where quality**	None

Topic	PRO Instruments and Timing	Action item
	improvement is needed. **A TEP member asked how patients will be providing this information in most settings; measure development staff stated that we will discuss later how to collect this data in the least burdensome manner. The TEP member emphasized that the level of burden is a significant consideration in selecting a tool and a few other TEP members affirmed this statement.** **A TEP member stated that assessing pain is critical to assessing the outcomes for this patient population, and that if there is not a pain score in the generic instrument to make sure that one is included in the condition-specific instrument. It would also be critical to assess pain relief separately from function for risk stratification of these measures. Without the availability of CAT, the PROMIS Global would be preferable to the PROMIS-29.** **A TEP member stated that instruments any longer than a dozen questions result in a significant drop in compliance rates.** Dr. Suter stated that data collection at the provider level would be aggregated to the hospital level and therefore sought rationale behind differences in survey ratings between applying instruments at the provider and hospital level. Aggregation of data would likely be done nationally. **A TEP member cautioned that some surgeons practice at more than one hospital.** Dr. Suter asked the TEP for further feedback from the TEP on the PROMIS surveys and any differences between using this instrument at the hospital level and provider level. **A TEP member stated that the PROMIS Global and PROMIS-29 would convey critical information: a patient's mental status prior to surgery, since that perception affects outcomes. This TEP member also favored using fewer questions to obtain this information.** Dr. Suter pointed out a discrepancy in survey ratings between the PROMIS Global's usefulness in a hospital-level measure and usefulness in a provider-level measure. A TEP member responded that this assessment may be more appropriate at the provider level because some surgeons are skeptical that the hospital has an impact on patients' overall quality of life after surgery. **A TEP member expressed concern about how patients with multiple conditions should be handled with regards to outcomes. Dr. Suter responded that it would be beyond the scope of this project to**	

Topic	PRO Instruments and Timing	Action item
	consider these assessments as applied to other conditions, although future measures of patient-reported outcomes could address this. **A TEP member stated, for the hospital-level measure, that assessing global health may be more useful in a retrospective manner than a prospective manner, as hospitals may not have access to pre-operative scores and may not be involved in pre-operative planning and assessments. Another TEP member supported this statement.** Dr. Suter summarized the discussion up to this point: for the generic measure, as long as there was adequate data capture in the generic instrument, a shorter instrument would be more valuable because of the reduced burden in data collection. Dr. Suter asked the TEP for feedback regarding the SF and VR measures, which capture the same information but received different ratings in the survey. **A TEP member responded that the more familiar tool received a higher rating. Another TEP member responded that a distinguishing factor – that the SF is proprietary – justified a lower survey rating. Another TEP member expressed support for using non-proprietary instruments.** Dr. Suter asked the TEP whether the SF-12/VR-12 was more favorable than the SF-36/VR-36. **A TEP member expressed support for the 12-question version of this instrument, which another TEP member also supported.** **A TEP member pointed out the SF tool is better for patients with lower literacy, although the tool is proprietary.** Dr. Suter asked the TEP for recommendations on selecting condition-specific instruments, which include the Oxford Hip Score/Oxford Knee score, the HOOS/KOOS, and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). She reminded the TEP that the Oxford scores do not distinguish between pain and mobility, which the TEP indicated as important for quality improvement, care management decisions, and performance measurement. The Oxford scores are proprietary, although other subject matter experts had expressed support for these instruments. **A TEP member stated that the HOOS and KOOS are derived from the WOMAC and that the WOMAC is proprietary; the HOOS/KOOS would be preferable to the WOMAC. Another TEP member supported this.** **A TEP member recommended using the Oxford scores and HOOS/KOOS if we are going to select two condition-specific**	

Topic	PRO Instruments and Timing	Action item
	instruments. **A TEP member emphasized that the Oxford scores cannot distinguish between pain and physical function. This issue becomes more prominent if the paired generic assessment instrument does not include a separate pain score. Another TEP member supported measuring these variables separately.** **A TEP member supported use of the HOOS/KOOS because they are non-proprietary and that there is the potential to examine subscores for these instruments in the future.** **A TEP member stated that the use of the HOOS and KOOS is supported by the Osteoarthritis Research Society International, which covers multiple specialties and is therefore a strong endorsement.** **A TEP member suggested moving forward with only the HOOS and KOOS. Other TEP members supported this statement.** Dr. Suter thanked the TEP for their insights and sought confirmation from the TEP that they were in support of proceeding with only the HOOS and KOOS instruments. **A TEP member added that the function metric becomes very critical in elderly and frail patients.**	
Timing of PRO Collection	Jenna Williams-Bader, Assistant Director, Performance Measurement at NCQA, reviewed the available options and rationale behind a 3-month and 1-month pre-operative timeframe for PRO assessment. The 3-month timeframe allows some flexibility and aligns with other measurement efforts. The 1-month timeframe may reflect the patient's baseline at the time of surgery more accurately. **A TEP member asked for confirmation that the timeframes indicate that the instrument was administered any time within the 3-months before the surgery as opposed to occurring at 3 months. Ms. Williams-Bader affirmed that this was the case.** **A TEP member asked what the timing of the generic and condition-specific assessments would be. Ms. Williams-Bader clarified that there will be one pre-operative timeframe that would apply to both the provider-level measure and the hospital-level measure. Phyllis Torda, Vice President, Quality Solutions Group at NCQA, further clarified that there would be one pre-surgery assessment (with the generic and condition-specific instruments) and one post-operative administration of the generic and condition-specific instruments.** **A TEP member stated that the implementation of this is important. If this would take place at the pre-operative visit, that would be within a**	None

Topic	PRO Instruments and Timing	Action item
	month. If this would take place at the surgeon's office, the longer 3-month window would be more practical. The Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) allows for 3-month window because the patients are from smaller hospitals and private practices, so a 3-month window was chosen for flexibility. **A TEP member stated that the rate of change within the 3 months before surgery, once someone has decided to have surgery, is minimal. The TEP member expressed support for the longer timeframe.** **A TEP member stated that the setting in which the assessment was completed may strongly affect the patient's anxiety level; filling out an assessment after an anesthesiologist speaks to the patient about the risk of paralysis may affect survey answers.** **Another TEP member agreed that where the survey is taken should be accounted for since it will influence the answers.** **A TEP member pointed out that many patients need a reminder in order to complete a pre-operative assessment and that a longer timeframe would allow for this reminder.** **A TEP member asked if the survey could be completed as a part of the education process, while another TEP member stated that this should be completed during earlier assessments because education takes place within the month before surgery. Another TEP member stated that this education process may bias the assessment instrument results.** **A TEP member stated that having a summary of pain scores pre-operatively helps form an understanding of the patient's quality of life as the patient is going through this decision making process.** **A TEP member asked whether there is potential for abuse or gaming by getting patients to give lower scored responses in order to achieve higher deltas on the assessment instruments.** **A TEP member expressed support for using a 3-month pre-operative timeframe.** Dr. Suter reminded the panel that a disadvantage of a wider window is that there may be factors that can influence pre-operative variation. She asked the TEP if there were any disadvantages to a narrower timeframe, in addition to the fact that a physician may not see a patient in the month prior to surgery. **A TEP member stated that at 1 month there is less of an opportunity**	

Topic	PRO Instruments and Timing	Action item
	for patients to change their mind about going through with the surgery. **Another TEP member asked whether this assessment could be completed at the physician's office or the hospital, as the setting and format matter. Dr. Suter acknowledged that the setting and format for the assessment instrument are important but are outside the scope of this discussion. Ms. Torda emphasized that we have not yet developed the testing protocol.** **A TEP member asked whether this project would aim to assess all patients or only a certain percentage of patients at a practice or hospital, which would affect whether a 1-month or 3-month timeframe was better. Ms. Williams-Bader clarified that the denominator for the measures would only include patients who have had the surgery.** **A TEP member suggested that there are other factors affecting specificity besides this timeframe and that allowing for a longer pre-operative timeframe would be more likely to yield a high response rate. This TEP member endorsed the 3-month timeframe. Another TEP member supported the 3-month timeframe for completion rates.** **A TEP member asked what the drop off rate was for patients who decide not to have surgery.** Dr. Suter stated that our goal would be to measure everyone and to get as many patients who have the surgery to have the pre-operative and post-operative assessment. Dr. Suter also clarified that we will be excluding revisions and fractures; these measures will only include primary elective procedures. **A TEP member explained that filling out an assessment survey, and thinking through those kinds of questions, may affect a patient's decision to have surgery or not. This may result in a higher quality decision for the patient. A few TEP members supported this as part of the shared decision making process.** Ms. Williams-Bader thanked the TEP for their input and confirmed that the TEP had expressed support for the 3-month pre-operative timeframe. Ms. Williams-Bader reviewed the options under consideration, including 3 to 6 months, 6 to 9 months, and 9 to 12 months after surgery, and clarified that this includes a buffer period to allow patients to respond. Shorter timeframes may reduce loss to follow-up. England's National Health Service uses a 6-month goal for collecting their PRO data. There is published evidence for statistically and clinically	

Topic	PRO Instruments and Timing	Action item
	significant improvement between 3 and 6 months after surgery as well as between 6 months and 12 months after surgery, although the curve in improvement flattens after 6 months. **A TEP member suggested separating the generic and condition-specific questionnaire, as they may be valid only at different intervals. Ms. Williams-Bader clarified that we would aim to choose only one timeframe for the post-operative assessments and that we would also include a buffer or tolerance in the allowed timeframe.** **A TEP member stated that 3 months would be too early and the assessments within that time would detect different rates of recovery but would not capture ultimate outcomes from the procedure. There are some, but not many, patients still experiencing change 6 months after the procedure. There is most likely going to be an encounter at 12 months. This TEP member favored 12 months because that is most likely when patients would be meeting with their physicians again, and this would minimize burden.** **A TEP member cautioned that some physicians will discharge the patient at 3 to 6 months and never hear from them again; there is a difference between academic centers and research centers compared to smaller groups doing a small number of joint replacements a year.** **A TEP member asked for an estimation of the proportion of patients that have shown significant improvement, but not necessarily plateaued, by 3 months. Another TEP member suggested that instead it would be important to consider time point at which patients had plateaued.** **A TEP member suggested that, in their experience, when comparing patients at 3 months, 6 months, and 12 months, 3 months would be too early, whereas by 6 months, hip replacement patients have experienced more than 90% of their ultimate improvement, and the change that occurs between 6 months and 12 months is minimal. This TEP member added that there is continued improvement between 6 months and 12 months for knee replacements, a higher percentage than hips. The vast majority of improvement, but not all, has taken place by 6 months. This TEP member collects these assessments electronically so an additional visit is not required. This TEP member also noted that the implementation of this project would be important to how these measures are structured. This TEP member endorsed 6 months as a practical post-operative assessment goal. This TEP member added that a one-year follow-up is common although it may vary across the country.** Dr. Kevin Bozic, a member of CORE's measure development working group, agreed that, in his experience, most of the improvement is seen	

Topic	PRO Instruments and Timing	Action item
	by 6 months for hip patients but not knee patients, and is therefore in favor of the 12-month timeframe. Dr. Bozic asked about the American Association of Hip and Knee Surgeons (AAHKS) intention to recommend a shorter timeframe for functional assessments about THA/TKA. However, the final AAHKS letter to CMS endorsed 180 to 365 day timeframe. **Another TEP member stated that many bundled payment initiatives that are occurring in orthopedics are tied to 90 days. Dr. Bozic noted, however, that the delta within the first 90 days may produce an inaccurate snapshot.** **Two TEP members expressed support for "the longer, the better" timeframe, especially for knee replacement patients. Another TEP member stressed that it is important to separate economic simplicity and validity, and that the longer timeframes would be better for validity. This TEP member added that AAHKS's suggestion is economically-driven, not scientifically-driven.** Dr. Suter emphasized that the discussion should focus on PRO data collection and to not confuse more specific clinical issues. She stated that knowing the difference between hip and knee replacement recovery is valuable, and that 6 months may be sufficient for hip replacement but not total knee replacement. However, it would be ideal from a measurement standpoint to have hip and knee replacement on the same timeframe. Dr. Suter reminded the TEP that we are looking for sufficient improvement to indicate the ultimate outcome for the patient but may not capture all improvement. She asked the TEP whether there would be specific clinical reasons for not collecting data at 6 months, knowing that the knee replacement patients may not reach their full potential by 6 months. **A TEP member stated that, in an underinsured population, there is significant drop-off in the post-operative assessment when compared to more affluent communities.** **A TEP member responded that the timeframe should be longer than 6 months. Another TEP member agreed and said that, if the intent is to look for a delta or outcome, then 12 months would be preferable.** Ms. Williams-Bader clarified that we have not yet defined the kind of outcome that we would like to use for these measures, and that we will get feedback from the TEP at another time. The choice of outcome may influence which timeframe seems most appropriate. If we were to choose minimal important difference (MID) as the outcome, we would not need the 12-month timeframe because most patients will have reached the MID by 6 months, if they are going to reach it. She also	

Topic	PRO Instruments and Timing	Action item
	clarified that it might be helpful to determine when we could identify outliers, as opposed to every single point for every patient; that would matter most if we were choosing average amount or overall amount of change as the outcome definition. **A TEP member stated that, when this is rolled out to the surgeon community, we need to make the argument that we chose the timeframe based on how we are going to measure it, otherwise the rationale becomes convoluted. If this timeframe is set to a point where surgeons know that their patients have generally plateaued, it will make more sense to surgeons and you will get better acceptance.** Dr. Suter asked the panel to confirm that a year would be the ideal clinical timeframe. A number of TEP members expressed their support for the 12-month timeframe. Dr. Suter then asked the panel for feedback on the 6-month timeframe. A TEP member stated that 6 months is adequate, although surgeons have been trained that it takes a full year to get optimal improvement. Dr. Suter asked the TEP if adding a window around 6 months would be sufficient. A TEP member suggested "9 months plus or minus 3" or "6 to 12 months" since the difference between 6 and 12 months is minimal. Mel Mastanduno, Director, Population Health Measurement Center for Population Health at Dartmouth Institute for Health Policy and Clinical Practice, stated that Dartmouth-Hitchcock, though only representing one provider organization, has experienced best response rates at 3 months and substantial drop-off at 6 months and 12 months. Dr. Suter acknowledged that we have so far discussed the clinical issues regarding timing and then asked the TEP about logistical issues of collecting this data at various points in time. She summarized discussion to this point: longer intervals are better but there are concerns about variation in data collection and when physicians are following up with patients. **A TEP member stated that patients are more activated to participate in PRO surveys earlier in their recovery. This TEP member was interested in considering the data with MID at earlier timeframes, as it was not a concept that the TEP member had thought about previously. Patients do move or have other things happen as the year goes on.** **A TEP member pointed out a potential floor effect if measured early and a potential ceiling effect if measured later; which effect is desired affects when measurement should occur.**	

Topic	PRO Instruments and Timing	Action item
	Dr. Suter asked the TEP how rapidly it becomes evident to clinicians that a patient will not do well. A TEP member responded that most manipulations occur within the first 3 months although maximal improvement will not be apparent until 9 to 12 months. **A TEP member stated that a patient who has a manipulation of their knee requires another month of recovery, although most patients would be satisfied by 6 months.** **A TEP member stated that it is important to have a distinction between pain relief and function; most patients will have a majority of pain relief by 3 months, but it takes 6 months to get full functional improvement. To be able to look at pain and function, 6 months is the better timeframe. To look only at pain relief, 3 months is a meaningful timeframe.** Dr. Suter asked the TEP if they would be comfortable with a measurement window that starts at 6 months, in order to compromise and balance response rates with a clinical plateau. A TEP member asked for clarification on the end of that measurement window. Dr. Suter asked the TEP as to how long that window should be, understanding that sometimes narrower is better and that a broader window means more variation in data collection. A TEP member stated that we should be able to look at the differences over this timeframe after we have data. **A TEP member asked whether the location would influence patients' responses, such as one performed at 6 months in an office as opposed to 8 months at home. Dr. Suter responded that there is data to suggest that mode of data collection does affect response rates in other PRO data.** **A TEP member asked for the rationale behind focusing on only one post-operative data point, as it may be interesting to get a high capture rate at 3 months and then a high delta at 6 to 12 months; one measurement point may not be enough. Ms. Torda responded that, from a measurement perspective, we would need to look at one post-operative measurement time point for feasibility. Ms. Torda further clarified that if we used a longer timeframe, such as 6 months, there is nothing that precludes anyone from measuring at 3 months as well. Often in measurement we find that organizations supplement the literal requirements of the measure to collect more information.** Dr. Suter sought additional comments or questions from staff and the TEP. Ms. Williams-Bader summarized discussion on timing to this point. Regarding pre-operative timing, 3 months will allow for more flexibility and there will not likely be large differences in score between 1 month and 3 months. Regarding post-operative timing, there is support for 6	

Topic	PRO Instruments and Timing	Action item
	months and 12 months. CMS/ONC might want to consider a range, which the team will discuss. Ms. Potteiger reviewed the next steps for the TEP, which included finalizing the measure specifications (selecting PRO instruments, specifying timing for PRO data collection, and determining outcome definition and risk adjustment), testing and validation of the measures, and public comment on the measures. Ms. Potteiger discussed the TEP summary report, which will be available for review by the TEP and will later be posted during public comment. Dr. Suter indicated there is interest in extending the TEP membership through September 2014, in order to include discussions of outcome definition, risk adjustment, and testing results. The TEP expressed interest in this extension of membership. Ms. Potteiger ended the call by thanking all of the TEP members for their time and feedback on behalf of the teams at Booz Allen Hamilton and CORE.	

	NEXT STEPS
Next steps	• The team will prepare the TEP summary report • The TEP will review the TEP summary report • The team will revise measure specifications based on TEP input regarding the PRO instruments and timing of data collection

Appendix F. TEP Meeting #3 Minutes

Patient-Reported Outcome (PRO) Following Elective Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA): Hospital- and Eligible Professional-level Performance Measures

Technical Expert Panel (TEP)

Summary Call #3

Friday, January 17, 2014, 3:00-5:00 pm ET

Participants

TEP: Peter G. Allen, MS; David C. Ayers, MD; Thomas Barber, MD; Daniel Berry, MD; Vinod Dasa, MD; Cheryl Fahlman, PhD, MBA, BSP; Cynthia S. Jacelon, PhD, RN-BC, CRRN, FAAN; Courtland G. Lewis, MD; Patient; Michael H. Perskin, MD; Jonathan Schaffer, MD, MBA; John Seiverd, PT, DPT, CCCE; Lyle Sorensen, MD; A. Christopher Strenta, PhD

Booz Allen Hamilton/National Committee for Quality Assurance (NCQA)/Dartmouth Institute for Health Policy & Clinical Practice Overview: Mike Sacca; Melanie Mastanduno, BSN, MPH; Kathleen Carluzzo; Phyllis Torda, MA; Jenna Williams-Bader, MPH; Daniel Roman; Katherine Sobel

Yale New Haven Health Services Corporation – Center for Outcomes Research (CORE): Lisa Suter, MD; Elizabeth Drye, MD, SM; Zhenqiu Lin, PhD; Michael Araas, MPH; Rana Searfoss, BA; Weiwei Zhang, MPH; Susannah Bernheim, MD, MHS

The Centers for Medicare & Medicaid Services (CMS): Lein Han, PhD; Karen Nakano, MD, MS

The Office of the National Coordinator for Health Information Technology (ONC): Kevin Larsen, MD, Lauren Richie, MA; Jennifer Wolff

SUMMARY	Action items
• The TEP reviewed and approved the updated TEP charter. • The developers reviewed candidate risk variables for THA/TKA PRO performance measure risk model testing. The developers grouped variables in the following categories: o High priority and to be considered for model testing (evidence-supported, feasible, low burden, and supported by orthopedists) o High priority to orthopedists, but lacking either evidence, feasibility, or low burden (i.e., they required additional consensus building and/or data collection to include in the risk model)	• The measure developers will survey the TEP on candidate risk variables and outcome definitions. • The developers will prepare the TEP meeting minutes and TEP summary report. • The TEP will review the TEP summary report.

SUMMARY	Action items
o Low priority due to lack of evidence, feasibility, and low burden • The TEP expressed concern that the proposed candidate list of risk variables did not prioritize key variables that provide important face validity to orthopedists. • The TEP requested that variables in the second category be included in the risk model. • The developers reviewed the rationale for risk adjustment and clarified that the goal is to level the playing field and not to maximize prediction of patient-level outcomes. • The developers also noted that variables with non-standardized or unreliable definitions present methodological problems if used in outcome measures. • The developers assured the TEP that it would fully consider the recommendations received in the meeting. • The developers proposed to follow up with an email survey to ensure the TEP's recommendations are fully captured and considered. • The developers reviewed candidate measure outcome definitions for the measures. • The TEP supported using an assessment of change in patient-reported outcome measure (PROM) score rather than a threshold PROM score. • The TEP supported assessing patient improvement from the preoperative state to the postoperative state. • The TEP recommended removing the following three outcome definitions from the candidate list: o Mean postoperative PROM o Threshold PROM o Threshold PROM delta • The developers proposed to follow up with an email survey to obtain more detailed feedback on candidate measure outcome definitions. • The developers notified the TEP of CMS's plans of holding a March 2014 public comment period for the hospital-level measure.	

	GENERAL
Welcoming Remarks and Introductions	Michael Araas, MPH welcomed the group to the third TEP meeting for the development of the Patient-Reported Outcomes following Elective Total Hip and/or Total Knee Arthroplasty (THA/TKA): Hospital-level Performance Measures and Eligible Professional-level Electronic Clinical Quality Measures. Mr. Araas delivered introductory remarks: • Asked TEP members to report any new disclosures since signing their nomination/disclosure forms

	GENERAL
	• Reminded participants that the materials and the specifics of the meeting are to be kept confidential • Noted this project is funded by contracts with the CMS Center for Clinical Standards and Quality and Office of the National Coordinator for Health Information Technology Mr. Araas reviewed the agenda for the call and the TEP meeting's main goals: • Review and ratify updated TEP charter • Seek TEP's input on candidate variables for risk adjustment • Seek TEP's input on candidate outcome definitions • Address outstanding questions and review next steps
TEP Role and Charter	Mr. Araas reviewed the TEP charter. The goal of convening this TEP is to obtain stakeholder and technical input. This process ensures transparency through the measure development process. Mr. Araas described the responsibilities of the TEP. The TEP approved the TEP charter.
Defining Common Terms	Mr. Araas reviewed common terms used during the meeting; CMS's and ONC's measure contractors use the terminology advanced by the National Quality Forum (NQF): • A patient-reported outcome (PRO) is the concept of a patient-reported outcome • A patient-reported outcome measure (PROM) is a survey instrument that captures patient-reported outcomes • A patient-reported outcome performance measure (PRO-PM) is a performance measure that uses patient-reported outcome data to define the measure outcome

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
Introduction to Risk Adjustment	Lisa Suter, MD thanked TEP members for continuing their TEP membership and for their valuable input. Dr. Suter discussed risk adjustment in order to frame the discussion on candidate risk factors for risk model testing. The goal of measuring outcomes is to improve patient care and decrease variation in outcomes due to poor care quality. Healthcare providers can directly control the care they provide their patients and some of the elements that affect outcomes. The healthcare system can also indirectly influence outcomes (i.e., through policy). In contrast, healthcare providers cannot change the fact that a patient may present with a history of diabetes; however, providers take this into account by determining a management plan for that patient. In the same way, outcome measures must account for patients' clinical status at the time they undergo surgery. Dr. Suter discussed that to compare providers who might care for patients with different disease severity the developers must identify relevant risk factors to include in the measure risk adjustment model. Dr. Suter reviewed that some risk factors are both more challenging to measure and to change (e.g., density of healthcare providers in a particular area or patient factors such as literacy that

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	do not inherently affect outcomes but are often associated with outcomes). Dr. Suter discussed that while these risk factors are not a focus of this TEP meeting, they are a concern for the measure developers and for ONC and CMS. The impact of such factors, like race or socioeconomic status (SES), on the performance assessments resulting from these measures will be thoroughly explored during measure development. At this time, based upon guidance from both CMS and NQF, it is not the developer's intention to risk adjust for such factors because this would obscure any existing disparities. Dr. Suter explained that we will discuss this topic more during the next TEP call when the developers have results from analyses conducted using existing data. Dr. Suter provided an overview of risk adjustment: • Provides statistical approach to make valid comparisons among providers • Accounts for patient severity of illness at the time of hip or knee replacement • Should not account for hospital or healthcare system influences or results of care (providers should get credit for providing good care and be held accountable for poor care)
Introduction to Candidate Risk Variables Discussion	Dr. Suter discussed that CORE identified candidate risk variables for inclusion in the measure risk model through an environmental scan (a scan of existing programs that measured PROs after THA/TKA) and a systematic review of the published literature. Dr. Suter reviewed the criteria for identifying possible candidate risk variables: • Evidence based: o Does the risk factor independently predict outcomes? (CORE looked at published literature that investigated multi-variable models and identified risk factors statistically and significantly associated with PROs.) o Do clinicians consider them important risk variables? • Feasibility o Is the data available for measure development and testing? o Is the data easy for patients and surgeons to collect? • Scientific validity and reliability. o Is the risk variable reproducible (i.e., can it be performed by a number of different people and get a reliable result)? o Do different providers define the variable differently, which would make it a less reliable and valid risk variable? Dr. Suter explained that after considering these criteria, CORE tiered the risk factors into three tables: 1. Table 1 presents the high-priority risk variables that will be included in development of the risk model. These variables may not end up in the final risk model, but they will be considered for inclusion. These risk variables met all three criteria. 2. Table 2 presents variables that are identified as high priority, but are not necessarily feasible for this current measure development process. These variables do not meet all

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	three criteria, but the orthopedic community identified them as a priority. At the time of the TEP call, the developers recommended deferring further investigation of these variables until they are more amenable to measure development and testing. Dr. Suter discussed that the developers are asking the TEP to help prioritize among these risk factors in Table 2 for further testing. 3. Table 3 presents low-priority risk factors that had little evidence to support their use, or they were difficult to collect, or not reliably collected. This can be because of the way these risk factors were defined or because they were duplicative when compared to other, more reliable, evidence-based, or valid risk variables. **A TEP member suspected that there will be buy-in on Table 1 and Table 3. This TEP member noted that the orthopedic community will be concerned around the deferral of Table 2, because a number of the diagnoses and problems in Table 2 are not rare; orthopedic surgeons know that these factors have a powerful effect on outcomes even if there may not be a lot of data to demonstrate this in multi-variable models. TEP member warned that there will be great uproar among the orthopedic community if the developers ignore the risk variables in Table 2.** **A TEP member agreed. This TEP member believes the orthopedic community would be willing to partner with the developers, ONC, and CMS to collect the necessary information to investigate Table 2 risk variables; through collection, these risk factors could be feasible for inclusion in risk model testing.**
Discussion of Candidate Risk Variables	Dr. Suter reviewed Table 1 which presents the high-priority risk factors that are feasible to include in risk measure model development. These risk factors represent a mix of factors that are strongly endorsed by the orthopedic community (Dr. Thomas Fehring, the President of the American Academy of Hip and Knee Surgeons [AAHKS], elicited feedback from orthopedic professional societies on risk variables across all types of outcomes after THA/TKA [not specifically patient reported]). Dr. Suter discussed that because the ONC measures under development are predominantly electronic health record (EHR)-based measures, it is important to consider EHR feasibility and what is already being incentivized for meaningful use collection. Dr. Suter asked TEP members if they have concerns about any of the risk factors listed in Table 1. Dr. Suter reminded TEP members that these are candidate risk factors that will be tested for inclusion in the risk model; this is not a final risk model. **A TEP member responded that vital signs are not collected by orthopedic surgeons and including vital signs will change clinical workflow. Patients' vital signs are typically evaluated during preoperative medical and anesthesia evaluation prior to going to the operating room; this would be the most appropriate place for vital signs to be collected.** Dr. Suter asked this TEP member if there is strong clinical utility or prediction risk from vital signs; because this is an elective surgery, Dr. Suter would expect vital signs to be less variable preoperatively among this group of patients. **TEP member is not aware of prediction risk of vital signs in the orthopedic literature. TEP**

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	member discussed that if somebody is hypertensive they will get stopped along the pathway to surgery because the patient's data is captured by other medical professionals, but not by the orthopedic surgeon. **A TEP member asked if the vital signs represent a singular snapshot or whether there is the ability to trend them. TEP member inquired if a patient comes into a provider's office hypertensive because of pain, how does this impact the measure if they are normally not hypertensive?** Dr. Suter responded that vital signs would capture the immediate preoperative assessment. The way meaningful use has incentivized collection would not allow trending of preoperative vital signs. **A TEP member asked for clarification on whether the team is referring to when patients go to the orthopedic surgeon's office for evaluation or the day of surgery.** Dr. Suter responded that this is mostly determined by the meaningful use criteria, but the developers anticipate that this would occur at the hospital on the day of presentation. Dr. Suter reviewed the input received on vital signs: these are not routinely collected at the time of surgery by surgeons; they are usually performed at multiple time points during preoperative assessments; and patients with abnormal vital signs do not often receive an elective surgery. Jenna Williams-Bader, MPH reminded TEP members that there are two different measures under development using different data sources. For the meaningful use measure, which uses EHR data, vital signs will be included from the preoperative assessment (when the surgeon meets with the patient in his or her office prior to the day of surgery). When CORE performs analysis of risk-adjustment variables, CORE will likely be using vitals that are collected at the hospital. Ms. Williams-Bader discussed that as we think about how vital signs might have an impact on outcomes and whether they need to be incorporated into this model in the future, the team will think about the vitals collected by the surgeon in the office. Kevin Larsen, MD agreed with Ms. Williams-Bader. Dr. Larsen explained that the meaningful use requirement is referring to CMS's EHR incentive program that asks for routine collection of blood pressure at clinical visits for providers receiving meaningful use incentive payments. Dr. Larsen anticipates that as people adopt the meaningful use program, there will be broader adoption of routine blood pressure collection. **A TEP member discussed that AAHKS brought up to ONC that this issue needs to be looked at from an orthopedic standpoint. TEP member explained that at the TEP member's institution blood pressure and vital signs are not collected prior to an orthopedic evaluation.** **A TEP member responded that regardless of whether vital signs are collected or incorporated in meaningful use, the idea that vital signs have any substantial effect on a THA/TKA PROM is very low. Patients are screened medically appropriately, which will be illuminated by a measure of surgical complications, but for patient-reported outcome measures it is not an important variable.**

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	A TEP member inquired if measure developers will include race and SES in the risk model. This TEP member discussed that these risk factors are difficult to measure and most orthopedists do not collect SES. Dr. Suter responded that this PRO-PM may illuminate that literacy and SES impact both the response rate to the survey, as well as potentially, the actual responses on the PRO survey. Dr. Suter agreed that measuring SES is very challenging. In prior assessments for other CMS outcome measures, measure developers used multiple measures of SES (e.g., median income of the zip code in which the patient lives and whether the hospital serves as a safety net hospital). **This TEP member raised the concern that these measures of SES are blunt instruments to measure something that can have a really big impact.** Dr. Suter discussed that NQF is struggling with guiding measure developers about this issue. SES risk adjustment for outcome measures has been a topic of debate for a long time. At this point, we are following the existing guidance from CMS and the NQF which is not to include these variables in the risk adjustment because this would adjust away the effect. While in certain situations that might be an appropriate goal, these measures are likely to be measures of relative performance amongst peers; therefore, if you give hospitals that serve a lower SES an advantage in the measure, then providers that serve patients of high SES will be penalized. Dr. Suter ensured the TEP that the measure developers will evaluate the measures to see the impact that any available, rigorous, and valid measure of SES has on hospital or the eligible provider performance assessment. Elizabeth Drye, MD added that CORE has looked at how providers with different proportions of low-SES patients perform on the measure score. For the THA/TKA hospital-level complication measure currently in public reporting, there is not much difference on average between hospitals that have a high proportion of low-SES patients and those that do not. Dr. Drye discussed that there exists a range of performance across hospitals with no low-SES patients and hospitals with a high proportion of low-SES patients, which suggests there are both good and less strong performers in those groups. Currently for THA/TKA readmission and complication measures, SES has not been a big factor in the way that providers look on the outcome measures. The starting assumption is that where there is a difference, the measure developers will not adjust it away in order to see the disparities, because it would be dealt with more in the use of the measure than the building of the measure. Dr. Drye stated that developers will examine this further at a later time. **A TEP member disagreed, explaining that the TEP member sees many Medicaid patients and indigent care and the TEP member's experience is counter to what Dr. Drye explained. This TEP member discussed that if this measure is implemented without risk adjusting for SES, it may limit the number of Medicaid patients some surgeons might see for fear of penalization.** **A TEP member mentioned a recently published article that indicated hospitals that had greater levels of dual-eligible patients had a greater readmission penalty rate because these are high-risk patients.**

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	A TEP member recently presented Oxford Knee Scores stratified by insurance status—Medicare versus private payer versus Medicaid. The Medicaid patients did much more poorly relative to the private payer and Medicare group. **A TEP member discussed that Dr. Matsen at the University of Washington has published the same type of information (outcome versus insurance status and Medicaid status).** Dr. Suter thanked TEP members for their input and concerns and suggested that the measure developers bring the results of their analyses, including disparity analyses, to the TEP at a future date. It would be helpful to have this discussion with data and address implications for this particular measure. Dr. Suter summarized TEP member concerns: vital signs are not uniformly collected among surgeons and not a meaningful clinical predictor of PROs; and SES and race important risk variables for inclusion in risk model. **A TEP member asked for further detail on the preoperative PROM score in Table 1.** Dr. Suter responded that there is no further detail at this point. Dr. Suter reviewed that the TEP recommended two generic PRO surveys (VR-12 and the PROMIS Global) and two condition-specific PRO surveys (HOOS and KOOS). Dr. Suter anticipates the generic and a condition-specific PRO will be collected preoperatively and postoperatively. The team still has to investigate how the outcome and risk variables are defined and linked, so the measure developers are leaving the definitions open at this point; measure developers will present the results to the TEP for their input. **A TEP member discussed that the PROMIS Global or the VR-12 is important because it can be divided into a physical summary score as well as an emotional summary score (or mental component score [MCS]). The MCS score is a summary score for many of the risk factors listed on the candidate comorbidities table, including depression, mental health, psychiatric disease, and anxiety. The MCS score summarizes what can be subclinical disease and can help define a pre-existing risk factor for a patient who may not carry a diagnosis for anxiety or depression in claims data. Likewise, the physical component score (PCS) from either the PROMIS Global or the VR-12 is a summary score about the preoperative physical status of the patient and provides important information about the entire musculoskeletal system, disease in other joints, cardiac and pulmonary disease, and other important medical conditions that can affect a patient's function preoperatively. Preoperative MCS and PCS are important for risk adjustment because they provide a wealth of information.** Dr. Suter agreed that individual PROM scores provide a wealth of information that may be useful either as individual components or as an aggregate measure and will be explored. Dr. Suter reviewed that the comorbidities were identified by the systematic literature review as well as by orthopedists. The asterisks represent those variables highlighted by the orthopedic community. The measure can incorporate a wide range of comorbidities by looking at all claims data (at least inpatient administrative claims) for the 12 months prior to their surgery. Dr. Suter acknowledged concerns about comorbidities captured in claims. There are limitations to

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	capturing comorbidities regardless of data source; patient-reported comorbidities may not be consistent and the problem list in the EHR may not be universally standardized or adopted at this point. The measure developers will use standardized ways of identifying comorbidities and test all the comorbidities available to the developers, prioritizing those that have been highlighted. The measure developers will also explore the different components of the PROM score as a TEP member identified. Dr. Suter discussed not including the risk variables in Table 2 in immediate risk-adjustment testing. Dr. Suter highlighted that there are limitations in these measures, in addition these variables may or may not prove to be important for ultimate inclusion in a PRO-PM. Dr. Suter recommended the team move ahead with a parsimonious list of evidenced-based risk variables that we have the ability to assess. Dr. Suter asked TEP members to identify, within Table 2, whether there are individual risk variables that ought to be investigated further. Dr. Suter echoed a TEP member's point about the value of the information incorporated in the mental and physical components of the generic health status PROMs. Dr. Suter discussed that when you aggregate administrative claims at the level of a hospital, and potentially the level of the provider (measure developers have not investigated provider level yet), they perform similarly to clinical-based information. They may not be able to predict risk at the patient level, individually, but aggregated data over the hospital can be very powerful. Dr. Suter discussed that PROM scores for every patient may be an even more predictive risk variable since they offer a wealth of information about how these patients live, their quality of life, and their health status; these variables may capture some aspects of other risk variables that we all think are critically important, like weight, education, or workman's compensation. Dr. Suter acknowledged that this measure will need to evolve over time. Dr. Suter agreed that the variables in Table 2 are important and invited the TEP to identify among which are the most important so that CMS and ONC could hear the TEP's concerns. **A TEP member expressed concern over the correlation between PROs and other risk variables. This TEP member asked if Dr. Suter feels comfortable in generalizing that PRO findings correlate with other variables in every situation at the hospital level. TEP member discussed that the correlation may be stronger in hospitals that are high volume and lower in hospitals that are low volume.** Dr. Suter clarified that measure developers will investigate this concern and do not have empiric data at this point. The measure developers will inform the TEP if there is a strong correlation at the hospital level between the PROs and other risk variables. Dr. Suter discussed the caveat that some of the risk variables in Table 2 that have been prioritized by the orthopedic community do not have standardized definitions. Dr. Suter discussed the lack of evidence supporting many of these risk variables because they are either not universally collected or they are collected and assessed by physicians in a non-standardized manner. This is a problematic issue in measure development if the variable cannot be reliably collected across different providers and different settings. **The TEP member responded that the definitions have significant meaning when looking at**

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	cross-correlation between underlying clinical issues and PROs. This TEP member noted that the orthopedic community has definitions, but they are not necessarily consistently used. **A TEP member asked if the team worries that some patients will not receive treatment if some of these risk variables are not included in risk adjustment.** **A TEP member agreed. This TEP member discussed that every orthopedic surgeon knows what these risk variables (e.g., congenital deformity, angular deformity) are and challenged the notion that the measure developers cannot define these variables and use definitions that people can understand.** **This TEP member expressed concern that if the team does not include these variables in an early model, people in the orthopedic community will dismiss the value of this model and be upset about the risk model. TEP member discussed that surgeons viscerally understand that these variables have a clear impact on any surgery that they do, more than factors such as age.** **A TEP member expressed concerns that not including these variables in the risk model will affect access to care for patients on which providers would have operated; because surgeons are worried about poor outcomes, they may not operate on these patients.** Dr. Suter responded that measure developers will include these factors if the orthopedic community has evidence that these risk variables can be collected uniformly and have inter-rater reliability of acceptable levels. Dr. Suter discussed that many of these risk factors (not necessarily congenital deformities, but other factors) may be captured by other risk variables that can be tested in the near term. Dr. Suter expressed that her experience with patients with workman's compensation is that there are a variety of ways in which the impact of workman's compensation can be assessed at the patient level. Dr. Suter discussed that it would be helpful to have evidence from the orthopedic community about the reliability of these variables, and if they offer incrementally more information than the risk variables that are widely and more easily collected by surgeons, physicians, patients, and hospitals in terms of the measure development process. **A TEP member asked if the measure developers would be open to looking at literature from other orthopedic specialties (foot and ankle surgeons have data looking at the impact of workman's compensation on calcaneal fracture outcomes).** **A TEP member responded that it is well proven in orthopedic literature that workman's compensation has a drastic impact on THA/TKA outcomes.** Susannah Bernheim, MD discussed that the measure developers have run into this issue before where there are variables presented in published literature that at a patient level have been shown to be very predictive of outcomes and seem critical to risk adjustment and yet are not available. Dr. Bernheim reflected on what CORE has learned over 10 years of doing this work. Dr. Bernheim acknowledged that there are two separate projects under development, but reflected only on the hospital measure. The goal is not to assess each individual patient's risk perfectly, rather the aggregate risk of a group. Dr. Bernheim stated that we are often surprised

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	that even in the absence of a single variable that may be critical to predict outcome at the patient level, we can predict risk well at the aggregate hospital level. Dr. Bernheim stated this may not be true for this measure and we need TEP members' input on what variables are the most important. CORE does not know whether we need these variables, but Dr. Bernheim asked TEP members to keep their mind open about risk factors; we may learn that we can build a better-than-expected model that predicts risk of an entire population at a hospital with a different set of variables. Cardiologists could not fathom that we could understand acute myocardial infarction (AMI) outcomes without risk factors like blood pressure and shock at arrival. CORE performed a validation study that showed that the hospital outcomes using clinical variables for risk-adjustment matched with the hospital outcomes using the administrative data. For this PRO measure we will have richer data sources and data from EHRs. **A TEP member responded that if the measure developers do not include these risk factors in the risk model at the start, how can it be determined that these risk factors are not important.** Dr. Suter responded that we may need to conduct a validation study. A validation study may not be feasible in the current contract year, but CMS and ONC will hear the TEP's concerns. Measure development is a phased approach; at this point, the measure developers intend to investigate a preliminary set of candidate models and assess the viability of the measure. If the risk model does not perform as well as the developers' threshold of acceptance, then the next step would be to figure out how to collect additional variables and redevelop the risk model. If the developers can develop a risk model with some preliminary evidence that is satisfactory, the risk model needs to be validated against data that includes the risk variables that have been identified by the orthopedic community before it can be used for individual provider profiling. **A TEP member stated that the TEP member is disturbed by this discussion because it feels as though the measure developers do not want to hear the TEP's input. This TEP member requested these concerns be on record that the measure developers either feel they cannot collect the variables in Table 2 and are making justifications why they are not collecting them, or they do not wish to hear TEP concerns about why these risk factors might need to be considered.** **A TEP member agreed. TEP member stated that we are setting up a grading system for surgeons and hospitals and some hospitals serve a great majority of Medicaid patients. Those patients might have a higher incidence of chronic pain problems. Every orthopedic surgeon recognizes the difficulties of taking care of Medicaid and workman's compensation patients, and that these are risk factors for poor outcomes. It seems that if we do not include these risk factors, and a surgeon is trying to do his duty by taking care of those patients, a surgeon or hospital may have a bad score; patients that are looking at data will think a hospital is a poor performer when it may not be rather the hospital is taking care of a difficult group of patients.** Dr. Suter responded that the measure developers have heard the TEP's concerns. Dr. Suter asked if the TEP members feel that every single risk variable listed on Table 2 is essential for

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	proceeding or if there are there risk variables on this list that are less important. **A TEP member responded that infection, congenital deformity, and angular deformity are important factors. This TEP member does not think abductor deficiency is important because it would be uncommon in the preoperative state. TEP member reviewed the rest of the risk factors in Table 2: extensor mechanism deficiency is important but rare; range of motion is more important in the knee than the hip; workman's compensation is important; the group already discussed education; chronic pain management is probably important; and no strong opinion for previous surgery on the lower limb and previous hip injury in terms of importance.** **A TEP member asked CORE how many elective surgery outcome measures they have developed versus non-elective issues like AMI.** Dr. Suter responded that the current measure under development is CORE's first PRO measure; CORE has developed two hospital-level, elective THA and TKA readmission and complication measures that recently started being publicly reported on *Hospital Compare*. The elective THA/TKA readmission and complication measures have a similar cohort to the PRO measures under development (e.g., they similarly exclude revisions, fractures, and hardware removal). **This TEP member stated that the context in which we work is important. In other disease states (e.g., AMI, pneumonia, and chronic heart failure) there is little choice for patients and providers; this is the opposite for THA/TKA.** Dr. Suter agreed and discussed that since this is an elective procedure and patients have an opportunity to make a choice and physicians have an opportunity to offer that choice, we have a responsibility to ensure that we are not going to incur unintended consequences by measuring providers. Dr. Suter ensured the TEP that measure developers hear their concerns and are not dismissing their concerns. Dr. Suter asked if any other TEP members want to weigh in on the risk variables on Table 2. **A TEP member discussed that this requires a far more extensive discussion because the literature is extensive and none is conclusive. TEP member echoed previous concerns and expressed that giving this discussion short shrift is extremely inappropriate. This TEP member reviewed the risk factors on Table 2: previous infection impacts outcomes; preoperative range of motion for the knee predicts postoperative range of motion; chronic pain management is an issue; previous hip injury is an important predictor for hip surgery; previous knee injury is an important predictor for knee surgery, but less than hip injuries impact on hip surgery; and workman's compensation portends a bad result. TEP member discussed that education should not be on this list since the vast majority of factors are anatomic deformities and there are only a few socioeconomic factors.** Dr. Suter responded that the measure developers are trying to ensure they are inclusive in risk variable testing. Dr. Suter asked if this TEP member is concerned about items that are not included on Table 2 that should be included. **This TEP member responded that there are other variables that should be included and some**

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	variables that should not be on the list because they are not feasible now or in the future. TEP member repeated that giving short shrift to this topic is very disturbing. Dr. Suter responded that the team should hold follow-up conversations. Dr. Suter suggested surveying the TEP about the risk variables if TEP members think that this would be a more in-depth opportunity to weigh in on individual risk variables. **A TEP member supported a survey because some factors are of variable importance and a survey would provide TEP members with the opportunity to rate them in terms of their potential impact.** **A TEP member agreed. This TEP member discussed that Dr. Fehring, President of AAHKS, offered to assist in collecting additional data about these factors which the orthopedic community feels strongly about. While there is not a lot of data right now from available datasets, we can enlist the orthopedic community's help to address some of these issues and find ways to collect this information.** Dr. Suter discussed that measure developers want to ensure all relevant risk variables are included in the survey that the TEP thinks are valuable. Dr. Suter asked if we need to include all tables in the survey or if Tables 1 and 2 encompass a sufficient breadth and depth. **A TEP member asked if measure developers are planning on testing the items in Table 3.** Dr. Suter responded that Table 3 has very challenging measures in terms of feasibility. If the TEP feels that the measure is insufficient without considering Table 3, then the measure developers need to be aware of that. **A TEP member responded that it is minimal additional effort to survey all three tables.** Dr. Suter stated that measure developers anticipate sending out a survey in a couple weeks after the meeting. Dr. Suter thanked the TEP for being willing to invest extra time and effort to ensure their concerns are heard. **A TEP member asked if the survey can have an "other" field for additional items.** Dr. Suter answered yes. **A TEP member recommended including Dr. Fehring's comments in the tables because AAHKS and AAOS can provide expertise. TEP member suggested gaining additional feedback from orthopedic societies. TEP member requested grouping risk variables by topic because TEP members may recommend picking only one or two anatomic and social issues. TEP member stated that the anatomic issues and previous surgery are far more important factors than the social issues.** **A TEP member asked for the survey to include definitions for risk factors that will be tested.** Dr. Suter responded that the survey will include definitions. Measure developers included a glossary of definitions in the TEP materials. One of the challenges is that not every risk variable

	Candidate Risk Factors for Risk Model Testing
Topic	**DISCUSSION**
	has a standard definition. Measure developers will identify where there is not an accepted, standardized definition; TEP members can inform measure developers if they are aware of an appropriate standard definition. Dr. Suter thanked the TEP for expressing their concerns. The measure developers will adjust measure development to incorporate those concerns.

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
Candidate Outcome Definitions	Ms. Williams-Bader reviewed the objective of the outcome definition discussion: to review the different outcome definition options and attempt to narrow the outcome definition list for 2014 testing. Ms. Williams-Bader discussed that we do not expect to make a decision about the outcome definitions at this TEP meeting. Rather the measure developers will conduct analyses over the next few months on different outcome definitions and present the data at the next TEP meeting to discuss which outcome definitions might be the best outcome definition for the measures. Ms. Williams-Bader discussed that there are many options for defining the outcome. The measure developers presented in the TEP materials some of the most common and evidence-based outcome definitions for PRO measures; while not an exhaustive list, these are the best options for this type of measure. Ms. Williams-Bader reviewed key questions to ask when determining the outcome definitions for measure testing: • Who will be using the outcome definitions? o Is the outcome definition useful to patients? o Is the outcome definition useful to providers (surgeons and hospitals) for quality improvement and for other purposes? o Is the outcome definition useful to other stakeholders? • Does the outcome need to identify below-average performers, average performers, and above-average performers? Ms. Williams-Bader reviewed the different categories of outcome definitions: • Continuous outcomes versus dichotomous outcomes o A continuous outcome is on a numeric scale (e.g., scores between 1-50 points) o A dichotomous outcome is whether the patient is meeting a pre-defined standard (e.g., patient who has a score of at least 40 on their postoperative PROM score, so the measure will say either “yes” they achieved that threshold or “no” they did not) • Change in health status measures versus the final health status measures o Change in health status is a score that reflects how much the PROM score has changed pre- to post-surgery. o The final health status is a score only looking at the final post-surgery PROM

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	score. Ms. Williams-Bader reviewed the disadvantages of continuous outcomes using means or averages: there is no way to tell by looking at the score what patient outcomes drove that score. There can be one hospital or surgeon whose patients mostly did very well, but a few did poorly. On the other hand, a hospital or surgeon can do reliably well, with some patients that do not do as well—this hospital or surgeon is not achieving the very high outcomes of the first surgeon or hospital, but they are not getting the very poor outcomes. When averaging, the scores for those two providers could look the same. Ms. Williams-Bader reviewed another disadvantage of the continuous outcome measures: one cannot tell who has worsened or who has stayed the same. For instance if a hospital or surgeon had large improvements, but some patients worsened, their average might still look good, even though they had some patients that did not do well. Ms. Williams-Bader discussed that literature suggests that the baseline PROM score has an impact on postoperative PROM scores; depending on the outcome definition, the impact will be different. For example, patients with a lower pre-surgery PROM score will achieve greater changes and improvements in their PROM score if they have more room for improvement. Therefore, for change-over-time measures, hospitals or surgeons may achieve better outcomes if their patients have lower baseline scores. However, when looking at a threshold score, for example, then the lower baseline scores will have more of a negative impact on the surgeon's or hospital's scores, because a patient with a low PROM score might improve a lot and still not change enough to meet the threshold. Ms. Williams-Bader recommended removing anchor-based outcome definitions from consideration because they are not particularly viable at the current time for outcome measures. Anchor-based outcome definitions use an external anchor to help determine which patients have improved, which have stayed the same, and which have worsened. The anchor can be defined by many ways: a clinical anchor decided on by consensus; a physician-determined anchor defined by how much the physician thinks the patient has improved; or a patient-determined anchor which includes a question asking the patient to either determine how satisfied they are with the surgery or globally how they feel after their surgery. The three anchor-based outcome definitions the measure developers recommend removing are: minimal clinically important difference (MCID), the minimal clinically important improvement (MCII), and the patient acceptable symptom state (PASS). Typically with PRO measures, the anchor is based on a question asked of the patient. Because these definitions often require additional data, particularly from the patient or the clinician, the measure developers suggest that it is reasonable to remove these definitions from consideration. The literature does not show major differences between these anchor-based definitions and the other MCIDs and MCIIs that were presented to the TEP. Ms. Williams-Bader asked TEP members for their input on the recommendation to remove these outcome definitions from consideration. Dr. Suter added that one of the primary motivations behind removing these outcome

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	definitions is because the measure developers are trying to minimize the number of questions asked of patients or surgeons to reduce burden and there is likely no methodological advantage to using anchor-based methods. **A TEP member asked if there is the possibility to ask the patient what they expect from their surgery using these PROMs, so we capture a relative baseline for each patient from which we can gauge whether we met their expectations?** Dr. Suter clarified that there are measures of patient expectations. The measure developers had not considered those as part of this measure, in terms of the outcome, because the goal is to focus more specifically on the actual symptom state of the patient and their functional status preoperatively and postoperatively. Dr. Suter noted that the TEP member's question is intriguing, and it highlights the idea of shared decision-making and many different aspects that are important to this discussion. However, the measure developers considered baseline expectations as a separate issue from the proposed measures. Dr. Suter asked other TEP members about their thoughts on the patient expectations aspect of measurement. **This TEP member clarified by providing an example of surgeons who treat a fairly diverse culture. High flexion, or the ability to bend your knee, for example, for someone in an Asian culture may be significantly important compared to someone in a different culture. Likewise, the ability to go up and down stairs may not be as important to a patient who works on a construction site, for example. A measure score of 50 for one person may mean something different for a different person. This TEP member asked if there is a way we can parse out those potential differences preoperatively.** Ms. Williams-Bader agreed that research exists in that area of differences in patient expectations of surgery. The measure development team will continue to consider this TEP member's points moving forward, but Ms. Williams-Bader noted that the data sources that the developers will use for measure testing are retrospective and may not ask about patient expectations pre-surgery, which means the measure developers will not be able to conduct targeted analyses on patient expectations at this time. Dr. Suter responded that there is an opportunity to continue this conversation about how to engage patients and their expectations through measurement, and it is particularly relevant in outcome measures. Dr. Suter further explained that a measure of patient expectations may be a complementary measure in a future iteration of the measures being developed as these types of measures evolve into focusing on shared decision-making. This path begins by using the current measures under development as a pure assessment of clinical outcome. **A TEP member added that the conversation about measuring patient expectations is important, but we are trying to develop measures that can be widely distributed. Incorporating the factor of patient expectations may be unmanageable and does not fit the current scope of work. This TEP member favors using tools that require minimal input, particularly from the provider side, in order to get the kind of data we need.** Dr. Suter thanked TEP members for their input.

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	Ms. Williams-Bader reminded the TEP members that any comments they were not able to share on the call could be emailed to the measure developers. Ms. Williams-Bader presented a figure showing simulated data points for ten fake patients undergoing total knee replacement. The graph had squares to represent preoperative PROM scores and triangles to represent postoperative PROM scores. One patient had the same preoperative and postoperative PROM score. Two patients worsened after their surgery, which was represented by red arrows pointing down from their preoperative score to their postoperative score. All other patients improved after surgery, which was represented by black arrows pointing up from their preoperative score to their postoperative score. Ms. Williams-Bader pointed out that all of the patients start at different points and end at different points. Some patients improved or changed more than other patients and some patients showed major changes from their preoperative state to their postoperative state. Ms. Williams-Bader reminded the TEP members that the figure showed actual differences, but when groups of patients are compared over time, there is a chance that measurement error shows a difference that is not a real difference. Thus, it is important to assess statistically significant changes. Ms. Williams-Bader presented two more figures to show how some of the different outcome definitions might give different results for the same set of patients for the same provider. One figure showed a horizontal red line representing a post-surgery threshold score. If we want patients to achieve a certain threshold after their surgery, setting a threshold score can show patients who have achieved the threshold by meeting or exceeding the pre-set score. Three patients in the example figure achieved the threshold, so the provider score was 30%. The next figure highlighted patients who achieved a threshold for the change in PROM score from preoperative to postoperative status. Five patients from the same set of patients in the previous figure achieved the threshold for change in PROM score. Thus, using the second definition, the provider was 50%. Ms. Williams-Bader asked the TEP if they had any questions about the example figures. **One of the patients on the TEP commented on the sample provider report cards, which the measure developers created to show how comparing the measure results for a set of providers may look for each potential outcome definition. As a patient, the TEP member would like to see five performance categories and quintile scores. Without knowing something about the distribution of those scores, it is difficult to determine how good or poor provider scores are.** Dr. Suter stated that the TEP member's feedback is helpful. Dr. Suter presented a sample report card for a provider (i.e., hospital or surgeon), which showed the actual numeric score on one row and the performance category (i.e., average, below average, or above average) on a second row. For every hospital or surgeon, the measure will estimate a point estimate, which is a numeric score. The point estimate can be the mean postoperative value, the percent of patients achieving a threshold, or it can be defined in a different way. There is an amount of uncertainty around that actual point estimate; for other outcome measures, CMS reports the actual outcome rate (i.e., point estimate) for the hospital along with an interval estimate. An interval estimate is similar to a confidence interval and it

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	measures the uncertainty around the numeric point estimate. A performance category could hypothetically be created that assesses, in addition to the actual score, whether the score is statistically significantly different than the average score for all of the hospitals or all of the surgeons that are being measured. Dr. Suter reiterated the previous TEP member's point that three performance categories may not provide sufficient meaning for comparison. Dr. Suter asked other TEP member about the different information conveyed by the numeric score or the performance category or both, and how this might help choose or understand their own experience as a patient, or their patient's experience as a surgeon. **A TEP member responded that one of the problems of PROs is that patients do not easily understand them. The numbers which are used, regardless of which outcome definition is chosen, are not easily transferable. The PROMIS Global or the VR-12 requires patient education to help patients understand what the interval means. It is perhaps slightly easier to interpret the HOOS and the KOOS, which have a zero to 100 scale, but some of the numerical numbers, be it the change or the mean postoperative score, are not something that is readily understandable by patients. Having both, or having some way to make the PROM score more patient-friendly and more easily understandable is important. Both the numeric score and performance category are needed for interpretation of the measure's result.** **This TEP member continued, stating that all of the orthopedic work to date has been focused on the delta (i.e., change in PROM score from the preoperative state to the postoperative state). Just looking at the mean postoperative score has not been, from a research perspective, able to show improvement. One of the ways to get higher mean post-operative scores is to operate on people who have less severe disease and have better pre-operative scores. Mean postoperative score does not illuminate improvement. Patients want improvement, which is measured by the delta. There are also profound differences in terms of patient-reported outcomes looking at pain, which is a much more predictable outcome following joint replacement compared to measuring function. Function is much less predictable and much more variable, depending on other musculoskeletal and medical comorbid conditions. This TEP member suggested removing mean scores from consideration because it is most important and relevant to patient interests to assess the change from the preoperative state to the postoperative state.** **A TEP member agreed with recommending that the measure developers focus on delta. TEP member added that the performance categories need to be presented to patients and the public, whereas the numeric score, specifically the delta, would be important to surgeons. Both are important, but for different purposes.** **A patient on the TEP also agreed.** Ms. Williams-Bader discussed the options for measuring the delta. Ms. Williams-Bader described the mean change in patient score, which simply indicates a change. The mean change in patient score does not indicate how much change in scores is a good change or how much change is a meaningful change. Providers could still be categorized by whether their mean

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	change score is average, below average, or above average. **A TEP member asked whether the measure developers would be able to compare pain and function separately instead of combining the two.** Dr. Suter stated that it is important to realize that, as an example, when CMS reports to hospitals the results on their outcome measures – for example, the recent publicly reported rates of readmission and complications after elective hip and knee replacement – what gets reported on the public website *Hospital Compare* is a number with an error estimate around it, and a categorization. There are accompanying figures that show green if a hospital performed better than the national average; yellow if a hospital performed the same as the national average, statistically speaking; or red if a hospital performed worse than the national average. This provides a sense of how well the hospital performs relative to its peers. CMS also sends hospitals a detailed list of information that includes data on all patients in the measures so that hospitals can use the information for targeted quality improvement. Dr. Suter added that, similar to a previous TEP member's comments, some of the information is more important to be public and must be easily translatable and understandable by a wide variety of people, patients, and providers. Alternatively, some information may be more detailed and important at the surgeon or the hospital level to enable the surgeon and the hospital to improve their care. It is worth thinking about whether this measure may need multiple pieces of information. Dr. Suter asked if the patients on the TEP had any comments about the type of information that may be important to see on a public dashboard about a particular hospital or surgeon. **One of the patients on the TEP had joint replacement surgery about nine months ago. Before going into surgery, as a patient, it is important to know the surgeon's level of experience. Patients with more education are more likely to research the surgeon before surgery. Some patients are limited to what information they receive in the mail or what their insurance companies provide them. Two key considerations of patients are (1) being pain-free – which applies to all patients – and (2) being able to function well in daily activities, which vary across patients. The TEP member recommended reporting the type of patients who are being treated, specifically the number of patients for each surgeon, who are on government insurance versus private insurance.** Ms. Williams-Bader agreed that the team will need to discuss the type of data that should be reported, both for the public and for the providers. **Another patient on the TEP stated that it would be helpful to report quintiles along with performance categories. This would help people to interpret what a particular numeric score means. It is difficult to interpret the numeric performance score without knowing something about the distribution of scores.** **A TEP member agreed with the TEP member's comment about the distribution of performance scores. This TEP member described that the other necessary piece is defining the minimum of what is a clinically important change. The minimum clinically important**

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	improvement is an important outcome to consider. Minimum improvement is different for different outcome measures and they are not necessarily mutually exclusive. We want to assess a delta, and then we want to decide whether to assess means or thresholds. These should be defined in relation to what is a clinically important change. Relating the results of the measures to a minimally important change would be really helpful to both patients and physicians. Ms. Williams-Bader described the post-surgery PROM score threshold. Ms. Williams-Bader reminded the TEP that a threshold would show the percentage of patients who achieved at least a certain score on their PROM. One disadvantage of setting a pre-specified threshold is that it would need to be a consensus-based value, which may not be evidence-based. A second disadvantage that the patient's ability to achieve the threshold may not necessarily be based on how well the surgeon performed; it may be based on the patient's baseline score. Ms. Williams-Bader presented another threshold option – a threshold for the amount of change in PROM score. This option is called the mean change in PROM threshold. This assesses, out of the mean changes, the number of patients who achieved a certain amount of change. Lastly, Ms. Williams-Bader presented the options for measuring minimal clinically important difference and the minimal clinically important improvement. The reason why minimal clinically important differences and improvements are different from the other definitions is that the amount of change that is determined by setting a threshold is defined as the least amount of significant change (significant from the patient's perspective). There are several statistical ways of determining the result. A percentage is used to report performance results of the hospitals or surgeons. For example, a result of 85% means that 85% of the provider's patients achieved the minimal clinically important difference or improvement. The difference between the minimal clinically important difference and minimal clinically important improvement is that a difference means the patient either improved or worsened by a certain amount, whereas improvement is the percentage of patients who achieved a positive improvement. Ms. Williams-Bader asked the TEP if they had any comments or questions. **A TEP member responded that there is controversy about minimal clinically important improvement (MCII) and minimal clinically important difference (MCID) in terms of between groups and between individual changes. There have been significant issues when using MCII/MCID to assess differences between individuals versus groups, and then looking at those differences between groups. They cannot be used interchangeably.** Ms. Williams-Bader agreed that this is a concern that the measure developers will investigate when considering the MCII and MCID. Ms. Williams-Bader suggested that the measure developers should distribute a survey of the outcome definitions to give the TEP an opportunity to provide additional feedback for each outcome definition. **A TEP member agreed with a TEP member's earlier comment and prefers the MCII. The delta is important and the MCII focuses on improvement, which is a major reason for performing**

	Candidate Measure Outcome Definitions
Topic	**DISCUSSION**
	the surgery. The MCII may be most relevant for patients when they are trying to evaluate the measure's results because patients want improvement. Dr. Suter ended the discussion by reviewing the next steps. First, the measure developers will survey the TEP on candidate risk variables and outcome definitions. Second, Dr. Suter noted that the CMS measure developer, CORE, will post preliminary measure specifications of the hospital-level measure for a public comment period in March. This is another forum to solicit feedback on the risk variables and the outcome definitions. This will not be the only opportunity for the public to comment on the measure. The public comment period will provide the TEP with an opportunity to share the measure information with member organizations and constituencies. Mr. Araas ended the call by thanking all of the TEP members for their time and feedback on behalf of the teams at Booz Allen Hamilton and CORE.

	NEXT STEPS
Next steps	• The team will prepare the TEP meeting minutes and TEP summary report. • The TEP will review the TEP summary report. • The team will survey the TEP's input on candidate risk variables and outcome definitions.