Measure Methodology Report for Public Comment:

Clinician and Clinician Group Risk-Standardized Complication Rate Following Elective Primary Total Hip Arthroplasty and/or Total Knee Arthroplasty

Submitted by:

Yale New Haven Health Services Corporation-Center for Outcomes Research & Evaluation (YNHHSC/CORE)

Prepared for:

Centers for Medicare & Medicaid Services (CMS)

October 2018

Center for Outcomes Research & Evaluation Project Team

Jeph Herrin, PhD*– Project Lead
Faseeha K. Altaf, MPH – Transition Project Coordinator
Jeffrey Dussetschleger DDS, MPH – Project Coordinator
Lisa G. Suter, MD* – Project Director
Susannah M. Bernheim, MD, MHS – Contract Director
Shu-Xia Li, PhD – Lead Analyst
Yixin Li, MS – Supporting Analyst
Zhenqiu Lin, PhD – Analytics Director
Mariana Henry, MPH– Research Associate
Andreina Jimenez, MPH – Research Associate
Amanda Audette, BS– Research Assistant II
Lynette M. Lines, MS, PMP – Project Manager
Harlan M. Krumholz, MD, SM* – Contract Principal Investigator

*Yale School of Medicine

Acknowledgements

This work is a collaborative effort, and the authors gratefully acknowledge and thank our many colleagues and collaborators for their thoughtful and instructive input. All final decisions are made by the project team at the Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE). Acknowledgment of input does not imply endorsement.

Specifically, we would like to thank the national Technical Expert Panel (TEP) members who provided important insights and feedback throughout measure development. These members included:

Name, credentials, and professional role	Organizational affiliation	Location
Kathleen Blake, MD, MPH; Vice President, Healthcare Quality (cardiology)	American Medical Association	Washington, DC
John Birkmeyer, MD; Chief Clinical Officer (general surgery)	Sound Physicians	Tacoma, WA
Dale Bratzler, DO, MPH; Chief Quality Officer (internal medicine); Interim Chief Quality Officer	University of Oklahoma Physicians; Chickasaw Nation Department of Health	Oklahoma City, OK; Ada, OK
Daniel Brotman, MD, SFM, FACP; Professor of Medicine, Johns Hopkins University Director of Hospitalist Program, (internal medicine)	Johns Hopkins University School of Medicine; Johns Hopkins Hospital	Baltimore, MD
Tracy Cardin, ACNP-BC, SFHM; Director of Nurse Practitioner/Physician Assistant Services (nursing - inpatient)	University of Chicago Hospital Medicine	Chicago, IL
Cathy Castillo, BA	Patient or caregiver representative	Redwood City, CA
Bruce Chernof, MD; President and Chief Executive Officer (internal medicine)	The SCAN Foundation	Long Beach, CA
Donna Cryer, JD; President and Chief Executive Officer	Global Liver Institute	Washington, DC
Sherrie H. Kaplan, PhD, MPH; Assistant Vice Chancellor, Healthcare Measurement and Evaluation School of Medicine, Professor of Medicine and Anesthesiology & Perioperative Care	University of California, Irvine	Irvine, CA
Timothy Kresowik, MD, MS; Professor of Surgery - Vascular Surgery (vascular surgery)	University of Iowa Hospitals & Clinics	Iowa City, IA

Name, credentials, and professional role	Organizational affiliation	Location
Joshua Lapps, MA; Government Relations Manager	Society of Hospital Medicine	Philadelphia, PA
Frederick Masoudi, MD, MSPH; Professor of Medicine and Staff Cardiologist (cardiology)	University of Colorado Denver, University of Colorado Anschutz Medical Campus	Aurora, CO
Brian McCardel, MD; Orthopedic Surgeon/Board Member (orthopedics)	Sparrow Health System	Lansing, MI
James Moore, MD; Clinical Professor of Anesthesiology and Perioperative Medicine (anesthesiology)	University of California Los Angeles Health	Los Angeles, CA
Michelle Mourad, MD; Vice Chair for Clinical Affairs and Value, Medicine (internal medicine - hospital medicine)	University of California, San Francisco Health	San Francisco, CA
Juan Quintana, DNP, MHS, CRNA; Certified Registered Nurse Anesthetist (nursing - anesthesia)	American Association of Nurse Anesthetists	Winnsboro, TX
Carol Raphael, MA, MPH; Senior Advisor	Manatt Health Solutions	New York, NY
Charlene Setlow; Patient	Patient representative	Salinas, CA
Heidi L. Wald, MD, MSPH; Vice President for Clinical Performance	SCL Health	Aurora, CO

We also would like to thank the Clinical Workgroup members who provided feedback on the MIPS THA/TKA complication measure specifications. These members included:

Name	Society Workgroup member is representing	Organization (title)
James I. Huddleston, MD	American Association of Hip and Knee Surgeons (AAHKS)	**Stanford University Medical Center** (Associate Professor of Orthopaedic Surgery)
David S. Jevsevar, MD, MNA	American Academy of Orthopaedic Surgeons (AAOS/Academy)	**Geisel School of Medicine at Dartmouth** (Assistant Professor of Orthopaedics; Chairman, Department of Orthopaedics; Physician Compensation Task Force Member at Dartmouth)
Joseph T. Moskal, MD, FACS	The Hip Society	**Virginia Tech Carilion School of Medicine** (Chairman, Department of Orthopaedic Surgery) **Carilion Clinic**
James I. Huddleston, MD	American Association of Hip and Knee Surgeons (AAHKS)	**Stanford University Medical Center** (Associate Professor of Orthopaedic Surgery)

We also would like to thank Katie Balestracci, PhD; Elizabeth E. Drye, MD, SM; Rebekah Heckmann, MD, MPH, MP; Raymond Jean, MD; Kendall Loh, BA; Melissa Miller, MP; Ilana Richman, MD; Doris Peter, PhD; Katie Piwnica-Worms, MD; Sriram Ramanan, BS; and Victoria Taiwo, MHA from CORE for their contributions to this work.

Finally, we would like to thank Kevin Bozic, MD, MBA (Dell Medical School) for providing important insight throughout measure development; Sharon-Lise Normand, PhD (Harvard Medical School) for her methodologic expertise; and our Contracting Officer's Representative, Lein Han, PhD, at CMS.

Table of Contents

List of Tables

List of Figures

Executive Summary

In 2017, the Centers for Medicare & Medicaid Services (CMS) contracted with Yale New Haven Health Services Corporation-Center for Outcomes Research and Evaluation (CORE) to develop an eligible clinician-level and/or eligible clinician group-level outcome measure that reflects the quality of care for patients undergoing elective primary total hip arthroplasty and/or total knee arthroplasty procedures (THA and TKA, respectively). CORE is developing the measure for use under the Merit-based Incentive Payment System (hereinafter, MIPS THA/TKA complication measure). This measure is a re-specification of the existing, publicly reported hospital-level THA/TKA measure and has been designed to align with the cohort and methods of that measure. This report presents the approach to development, specifications, and testing results of the MIPS THA/TKA complication measure for public input.

Although THA/TKA elective procedures dramatically improve quality of life and function, serious complications do sometimes occur. For patients undergoing operations that are elective, the associated risks are particularly important to understand and weigh in their decision-making. Current clinician-level quality improvement measures for patients undergoing elective THA and TKA procedures are generally limited to evidence-based processes of care. Measurement of patient outcomes, such as complications, allows for a more comprehensive view of quality of care, capturing more complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment.

CORE developed this measure consistent with CMS's guidance for quality measurement development. The CORE Project Team included a multidisciplinary team of clinicians, health services researchers, and statisticians. CORE convened and consulted a national Technical Expert Panel (TEP) throughout measure development. The TEP consisted of physician and non-physician practitioners with experience providing care in the hospital setting as well as patient and caregiver representatives. Additionally, CORE convened a workgroup of orthopedic society representatives.

The target population of the measure is Medicare fee-for-service (FFS) beneficiaries undergoing inpatient THA and/or TKA procedures. The measure's outcome is any one of the specified medical or surgical complications occurring during the index admission or during a readmission except for death, which can occur anywhere as long as it is within 30 days of the start of the index admission. The measure is risk-adjusted, meaning it takes into consideration patients' age, sex, and clinical comorbidities in profiling eligible clinicians (ECs) and EC groups. Patient outcomes are attributed to the clinician who billed for the procedure, or the Billing Surgeon. The measure reports risk-standardized complication rates for MIPS eligible clinicians – those with unique Taxpayer Identification Number (TINs)/National Provider Identifier (NPI) combinations – or MIPS eligible clinician groups, or TINs.

Overall, we find that the median risk-standardized complication rate (RSCR) for MIPS eligible clinicians with more than 25 cases was 2.7%, with a minimum interquartile range (IQR) of 2.4% - 3.2%. The median RSCR for MIPS eligible clinician groups with more than 25 cases was 2.8%, with a minimum IQR of 2.5% - 3.1%. The measure has good 3-year reliability using the current specifications. In a formal survey of the technical expert panel, 81% agreed the measure scores were valid and useful measures of quality of care.

The MIPS THA/TKA complication measure, as specified, has the potential to illuminate these differences in quality, inform patient choice, drive quality improvement, and enhance care coordination. We look forward to your input on any and all aspects of the measure specifications during public comment.

1. INTRODUCTION

1.1 Overview of Measure Development

In 2017, the Centers for Medicare & Medicaid Services (CMS) contracted with Yale New Haven Health Services Corporation-Center for Outcomes Research and Evaluation (CORE) to develop an eligible clinician-level and/or eligible clinician group-level outcome measure that reflects the quality of care for patients undergoing elective primary total hip arthroplasty and/or total knee arthroplasty procedures (THA and TKA, respectively). Specifically, CMS asked CORE to re-specify its existing hospital-level measure, "Risk-Standardized Complication Rate (RSCR) Following Elective THA and/or TKA," that is currently publicly reported for hospitals, for use in assessing individual or groups of clinicians.[1]

This report presents the preliminary specifications of the MIPS THA/TKA complication measure for public input. The goal of this measure is to improve the quality of care delivered to patients undergoing elective primary THA and/or TKA procedures. The complication measure will inform quality improvement efforts targeted toward minimizing medical and surgical complications during surgery and the postoperative period. The premise is that improved quality of care, including coordination and communication among providers and with patients and their caregivers, can favorably influence performance on this measure.

1.2 Importance of a Complications Measure for Elective Primary THA/TKA

THAs and TKAs are commonly performed procedures that improve quality of life. Between 2005 and 2011, 855,899 THAs and 2,040,667 TKAs were performed.[2,3] In 2014 alone, knee arthroplasty was the most common procedure performed on patients aged 65-84, and total and partial hip replacements in the age group were the second most frequently performed procedures.[4] Although these procedures dramatically improve quality of life, they are costly.[5] In 2014, THA and TKA aggregate costs for hospitalizations when knee arthroplasties and total/partial hip replacements were first listed on the record were $11.8 billion and $8.3 billion, respectively.[4] Medicare is the single largest payer for these procedures, covering approximately two-thirds of all THAs and TKAs performed in the US.[6] Combined, THA and TKA procedures account for the largest procedural expenditure in the Medicare budget.[7]

Future utilization of THA and TKA is projected to increase significantly. By 2030, the demand for THAs is estimated to increase by 174% while the demand for TKAs is estimated to increase by 673%.[8] Complications increase costs associated with THA and TKA and affect the quality, and potentially quantity, of life for patients. Because these are commonly performed and costly procedures, it is imperative to address quality of care, especially with their projected growth.

Although complications following elective THA and TKA are not common, they are measurable and vary in prevalence across providers. Rates for periprosthetic joint infection following THA and TKA range across hospitals from 1.6% to 2.3%, depending upon the population.[9,10] Reported 90-day death rates following THA range from 0.7% [11] to 2.7%.[12] Rates for pulmonary embolism following TKA range from 0.5% to 0.9%.[12-15] Rates for wound infection in Medicare population-based studies vary between 0.3% and 1.0%.[12,13,15] Rates for septicemia range from 0.1%, during the index admission[16] to 0.3%, 90 days following discharge for primary TKA.[12] Rates for bleeding and hematoma following TKA range from 0.9% [16] to 1.7%.[17]

The variation in complication rates across hospitals indicates there is room for quality improvement. A quality measure to address complications following THA and TKA provides an opportunity to provide targets for efforts to improve the quality of care and reduce costs for patients undergoing these elective procedures. In the case of THA/TKA, individual clinicians, in particular surgeons, are the key implementers of quality improvement strategies used by hospitals to reduce complication risk, as they play the primary role in the procedure.

Lastly, public and private payers have supported the implementation of orthopedic outcome measures. In 2013, CMS began to publicly report on the RSCRs for THA/TKA for the nation's non-federal short-term acute care hospitals (including Indian Health Services hospitals) and Critical Access Hospitals as part of the Inpatient Quality Reporting (IQR) program and thereafter finalized the measure for the clinical care domain of the Hospital Value-based Purchasing (HVBP) program starting FY 2019[18,19] and the Comprehensive Care for Joint Replacement (CJR) model. Even more recently, through its implementation of the Medicare Access and CHIP Reauthorization Act (MACRA) and development of its Measure Development Plan (MDP), CMS identified orthopedic surgery as a priority area for specialty-focused clinician measurement when measuring quality of clinician care. A subsequent environmental scan identified clinician quality measure gap areas and specifically identified complications from orthopedic procedures as a specific subtopic that currently has no measures.[20] The MDP Technical Expert Panel (TEP) reviewed and agreed with the importance of measuring outcomes of orthopedic surgery as an important measure development area in the Quality Payment Program and clinician measurement under the MIPS.[21] Additionally, beyond current hospital payment and reporting programs, the Core Quality Measures Collaborative, a stakeholder group convened by America's Health Insurance Plans, that included purchaser, consumer, CMS, the National Quality Forum, and physician organization representatives, identified orthopedic quality measurement as one of seven core measure sets to support quality improvement. [22] [22]

1.3 Quality Payment Program Background

In April 2015, Congress passed MACRA, which marked a milestone in moving from paying clinicians based on volume of services towards paying clinicians for value of care. MACRA laid forth two pathways for physicians and other clinicians participating in CMS's Quality Payment Program (QPP): (1) the Merit-based Incentive Payment System (MIPS) or (2) an advanced Alternative Payment Model (APM). This work is informed by, and focuses on, several aspects of the MIPS requirements.

1.3.1 Eligible Clinicians and Eligible Clinician Groups

The first aspect of the MIPS which informs this work involves defining eligible clinicians (ECs). CMS has identified a set of clinicians based on Medicare provider specialty codes and Medicare Part B volume requirements for participation under the MIPS. The types of MIPS ECs include physicians, physician assistants, nurse practitioners, clinical nurse specialists, and certified registered nurse anesthetists who bill under Medicare Part B (81 FR 77036).[23] CMS describes clinicians who participate in MIPS as MIPS ECs. MIPS ECs may participate as a single clinician (identified by a unique combination of Taxpayer Identification Number [TIN] and National Provider Identifier [NPI] numbers), as a group (TIN with 2 or more clinicians), or as a virtual group (2 or more TINs of solo practitioners and small groups of fewer than 10 clinicians). CMS intends to use at least one outcome

measure (or other high priority measure) to assess the quality of care provided by MIPS ECs who choose full participation in MIPS to achieve higher payment adjustments (82 FR 30028).[24]

1.3.2 Outcome Measures

As part of the MIPS, participating clinicians must report at least six quality measures. Of these six, one measure must be an outcome measure. If no outcome measure is available, clinicians must select another high-priority measure in its place. If fewer than six outcome measures are available, clinicians must report on those available. Placing importance on outcome measures and in alignment with statutory requirements, CMS indicated its plans to increase the requirements for outcome measure reporting over time as more outcome measures become available for MIPS reporting (81 FR 77101, 82 FR 30097).[25,26] While CMS has not indicated whether some or all future risk-adjusted outcome measures developed for use under the MIPS would be optional or required for reporting, CMS will automatically calculate the first risk-adjusted outcome measure finalized for the MIPS, called the all-cause readmission measure, for groups of 16 or more eligible clinicians and score measure performance using a decile distribution (81 FR 77282 through 77284).[27] This measure development work is motivated largely by the prospect of adopting additional inpatient outcome measures for ECs.

1.3.3 Existing MIPS Attribution Approaches

An important consideration for this work was the attribution used by existing outcome measures under the MIPS. CMS published beneficiary assignment methods for the MIPS all-cause readmission and total per capita cost measures and implemented the attribution for the first year of MIPS (2019 MIPS payment year). The attribution methodology was adopted from the Value Modifier (VM) program, which uses outpatient claims to identify a primary outpatient provider during a 12-month performance period during the measurement year. Specifically, the two-step attribution methodology for the VM all-cause readmission measure assigns beneficiaries first to eligible clinicians based on a plurality of charges for delivery of primary care services by primary care physicians or, secondly, to the specialist with plurality of charges for such services if no primary care physicians provided any such services in the 12-month performance period during the measurement year. For the total per capita cost measure in MIPS, CMS modified the algorithm by removing the skilled nursing facility codes from the list of qualifying primary care services used for attribution (79 FR 67960 through 67964, 81 FR 77131).[28]

Hospital Quality as a Proxy for Clinician Quality under the Merit-based Incentive Payment System

In the program's first year (2019 MIPS payment year), CMS introduced its consideration to allow facility-based clinicians to use their institutions' quality and/or cost scores as a proxy for the MIPS EC's quality and/or cost performance scores (81 FR 77127).[29] CMS believes providing this option to clinicians will allow for clinicians to be assessed along the lines of the facilities in which they work and minimize reporting burden (82 FR 53753).[30] For the 2021 MIPS payment year, CMS has proposed adopting measures from the Fiscal Year 2020 Hospital Value-based Purchasing Program, including CMS's hospital-level THA/TKA

complication measure, for facility-based measurement under MIPS (83 FR 35960).[31] Attribution of a facility-based clinician would be to the hospital at which the facility-based clinician provides services to the most Medicare patients, and attribution of facility-based groups would be the hospital at which the plurality of facility-based clinicians were attributed. In contrast to such facility-based measures, this work created an EC- or EC group-level measure that is aligned with, but not identical to, the original hospital-level measure. This approach for some measures may be more aligned with stakeholder interests when measuring the quality of clinicians.

1.3.4 Measure Alignment

Finally, one of CMS's priorities in implementing MACRA is to align quality measures across federal programs, such as the MIPS and advanced APMs, settings, and payers. In November 2017, CMS finalized using benchmarks for MIPS quality measures for calculation of APMs (82 FR 53698).[32] CMS's future policies in this area will be important in guiding the attribution of patient health outcomes to clinicians participating in the QPP via the MIPS or advanced APM pathways.

1.4 Approach to Measure Development

CMS contracted with CORE to lead the re-specification of the hospital THA/TKA complication measure for use in MIPS under the guidance of CMS. The CORE Project Team consists of a multidisciplinary group of individuals with expertise in measure development, health services research, clinical medicine, statistics, and measurement methodology.

We developed this measure in consultation with national guidelines for publicly reported outcome measures. We followed guidance set forth by the CMS Measure Management System Guidance[33], the National Quality Forum (NQF)[34], and articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes".[35] Following these standards has ensured a transparent and comprehensive process with expert input throughout development (see Acknowledgements).

Below we review our approach to measure development.

1.4.1 Expert and Stakeholder Input

As part of measure development, CORE has obtained input on measure development from persons and families, clinical and methodologic experts, and other stakeholders.

As part of CMS's commitment to incorporating views of persons and families, CORE hosted two listening sessions to obtain feedback from persons and families about clinician quality measurement. The goal of the sessions was to obtain input from persons and families regarding quality measurement at the clinician level and attribution of selected outcomes to clinicians. We provided participants with the project's background and presented three scenarios for discussion. As part of these sessions, participants provided input for various scenarios, including to whom patient outcomes should be attributed for patients undergoing elective procedures. Feedback focused on concerns about holding clinicians accountable for events beyond their control and about identifying the true causes of adverse outcomes.

As is standard with all measure development processes, CORE also convened, through a public process, and obtained input from a national TEP throughout measure development. The TEP consists of clinicians, patient advocates, and other stakeholders. The TEP has provided input on approaches to measure re-specification including attribution and risk-adjustment methodology (see Acknowledgements for TEP roster).

Additionally, we also have obtained stakeholder input from professional societies and individuals with relevant expertise. We obtained input from a Clinical Workgroup made up of representatives of relevant orthopedic professional societies (see Acknowledgements for Clinical Workgroup roster). On an ad hoc basis, CORE has solicited input from additional individuals with expertise relevant to orthopedic quality measurement.

We will also be incorporating stakeholder input from this public comment period and a listening session with representatives of multiple specialty societies that we will host during this call for public comment.

1.4.2 Key Principles Driving Attribution Identification and Evaluation

As part of this development process, we identified five key principles to guide re-specification of hospital measures for measuring clinician quality and added a sixth identified by the TEP. Our approach to identifying and evaluating attribution rules reflects a set of principles that we derived from prior work on hospital measurement, policy goals, consultation with our TEP, the context of adapting existing measures, and the common features of those measures. Notably, these principles are specific to hospital measure re-specification and may not be applicable to attribution in general. In this section, we state these six principles explicitly and describe how they proscribed and informed our choices and findings.

Principle #1: Attribution is Specific to the Measure Outcome

Throughout this document, attribution refers to the assignment of the outcome of a patient episode of care to one or more clinicians for the purpose of assessing clinician quality. Attribution, therefore, is specific to the outcome. This is because the goal of attribution is for purposes of quality measurement rather than for assessing utilization or other characteristics of the relationship between the patient and the provider. For example, when a patient is admitted for elective surgery, it may be most sensible to attribute any complications of that surgery to the surgeon but any post-discharge readmission to the clinician who discharged the patient. For the MIPS THA/TKA complication measure, we considered attribution to ECs who might plausibly influence perioperative care as well as post-discharge follow-up care.

Principle #2: Adapted Measure Should Align with Original Hospital Measure

Our goal was to adapt the patient cohort, outcome, and risk-adjustment strategy that had been previously specified for hospital measurement for use in measuring clinicians. We took as a principle, then, that an adapted measure should align to the degree practical with the existing measures. We only considered attribution approaches that could be implemented using the same data sources that are used to measure hospitals with the same cohort and outcome definitions. The risk-adjustment variables and model

would be, when practical, similar or identical to those used for the hospital-level measure as risk prediction is a patient-level adjustment and should not be influenced by the attribution. Thus, for the current measure we adopted the original cohort, outcome, and the existing set of risk factors and then evaluated the model performance.

Principle #3: Clinician Quality Reflects Hospital Quality

This measure was originally developed to measure hospital quality. When measuring performance, it may be possible (if technically challenging) to isolate the components of quality at the clinician, group, and hospital levels. However, just as hospital quality measurement inherently reflects contributions from clinical staff, hospital systems, and community resources, we adopted the analogous principle here that clinician performance measurement also reflects other factors including hospital quality. Therefore, just as with CMS's hospital measures, we did not try to separate them when measuring performance. From the perspective of the patient, this means that when comparing providers, the performance reflects the hospital or outpatient environment in which the physician practices. From the perspective of the policymaker, this principle means that clinicians are held accountable in part for the quality of the hospital environment where they treat patients; since these are individuals perhaps best placed to identify systemic opportunities for improvement, this can drive improvement throughout the system of care.

Principle #4: Inpatient Outcomes May Be Most Reasonably Attributed to Inpatient Clinicians

We identified candidate attribution rules using four sources: 1) a literature review/environmental scan; 2) current CMS policies; 3) TEP and other expert input; and 4) claims patterns for measured patients. A hierarchy that arose from clinical and TEP input allowed us to identify key candidate attribution rules:

- Hospital physicians generally play the most important role in outcomes after hospitalization.
- The most central hospital physician depends in large part on the condition/procedure and outcome.
- Physicians caring for patients before and after a hospitalization may also play a role in post-hospitalization outcomes.

Principle # 5: Attribution Should Align with Policy Goals

Consistent with guidelines on attribution published by the NQF,[36] we adopted the principle that the choice of attribution rule should be ultimately determined by policy goals and informed by clinical sensibility and empirical findings. Thus, while empirical findings may illuminate what is feasible and practical, they cannot determine what is "right" or "appropriate." For example, empirical results may indicate that a readmission outcome after a surgical procedure can be feasibly attributed to either the surgeon or the discharging clinician, but the results cannot determine that one is "better" or "more sensible" than the other. The choice between the two attribution rules will need to be

based on clinical and policy considerations. Thus, our approach was not focused on discovering a single best attribution rule for attributing an outcome, but rather illuminating what is possible and what is meaningful.

Principle #6: Attribution Should Consider the Potential for Unintended Consequences

We prioritize the goal of improving patient care. One implication of prioritizing patient care is that we considered the incentives created or modified by each candidate attribution rule. An attribution rule could conceivably create lines of responsibility that result in a tradeoff between better patient care and better clinician scores. For example, any rule that can be manipulated after admission, allowing clinicians to avoid attribution of a patient's outcome once they have provided care for that patient, could create incentives for a clinician to 'shift' patients with poorer prognoses to another clinician, resulting in perhaps worse care for the patient but better measure scores for the first clinician. Therefore, we considered potential unintended consequences for each candidate attribution rule.

These six principles provided a framework for thinking about attribution of inpatient outcomes in a way consistent with CMS's policies and goals. These principles and sources were broad enough to identify and to evaluate all candidate rules identified by us and the TEP.

Finally, we only considered attribution to the types of clinicians that are eligible for the QPP. Currently, the types of clinicians which qualify for participation are physicians, physician assistants, nurse practitioners, clinical nurse specialists, and certified registered nurse anesthetists; this list may be expanded over time as directed by CMS.

1.4.3 Strategy for Adapting Inpatient Outcome Measures to Apply to Clinicians

Prior to developing this MIPS THA/TKA complication measure, we developed a general strategy for re-specifying existing hospital-level inpatient outcome measures to apply to ECs. This strategy consists of: an approach to identifying candidate attribution rules; methods for evaluating the candidate attribution rules; and criteria for reviewing the findings to inform the choice of a final attribution rule. The overall process for identifying, testing, and selecting algorithms ("attribution rules") for assigning patient outcomes to clinicians consists of three key steps:

1. *Identify candidate attribution rules:* Use literature and related publications, existing policies, claims patterns, clinician input, and expert opinions to identify a preliminary set of candidate attribution rules for the measure under consideration. These rules were chosen to be consistent with the six key principles described above. Descriptive data on claims patterns may also inform this set of candidate attribution rules. The aim of this step is to identify a set of attribution rules that are both clinically meaningful and relevant to policy.

2. *Implement candidate attribution rules on a common dataset and evaluate key characteristics of each implementation:* For each implementation, evaluate empirically the face validity, ability to differentiate among providers, reliability

and case volumes, and overlap of the candidate attribution rules. We refer to results of a random attribution for comparison. This step confirms feasibility and could lead to some candidate attribution rules being refined or dropped from further consideration.

3. *Use clinical judgement and weigh potential unintended consequences and policy considerations to select a final attribution rule:* We present the results of the evaluation to stakeholders for their input. Specifically, we held an in-person meeting of our nationally convened (TEP) that includes representation from a broad group of providers as well as patients. We presented the results to them to obtain their preferences and input on the candidate attribution rules. Quantitative evaluation of the candidate attribution rules can inform but not determine the final choice of attribution rule. This step synthesizes the results of the first two steps to clarify trade-offs to inform the adoption of any attribution rule.

This strategy was pre-specified and endorsed by stakeholders and CMS.

1.5 Aims of the Measure

The primary objective of this work was to develop a 90-day THA/TKA complication measure for clinicians that:

- Captures differences in complications experienced by patients who underwent elective THA/TKA procedures.
- Adjusts for clinician case mix.
- Assesses for relative performance of clinicians.
- Aligns with CMS's existing hospital-level THA/TKA complication measure, as appropriate.
- Provide targets to clinicians for efforts to improve the quality of care.
- Reduce costs for patients undergoing these elective procedures.

1.6 Purpose of the Public Comment Period

Outcome measures include several major components: cohort, outcome, approach to risk adjustment for case-mix differences across providers, and statistical modeling approach. In addition, an important consideration for this work was developing an appropriate strategy for attributing the patient outcomes (complications) to individual eligible clinicians or eligible clinician groups.

As part of the measure development process, and in alignment with CMS Measure Management System guidance,[33] we seek comment on the following:

1. Does the measure identify the appropriate EC or EC group responsible for complications following elective primary THA/TKA procedures?
2. What, if any, additional validity testing would be meaningful for this measure?

Instructions for submitting comments as an individual or an organization are available on CMS's public comment website.

2. METHODS

2.1 Overview

This measure reports the EC-level or EC group-level risk-standardized complication rate (RSCR) following elective primary THA/TKA procedures performed during inpatient admissions.

Consistent with CMS's hospital THA/TKA complication measure, this MIPS THA/TKA complication measure uses three years of data to assess EC or EC group performance. The measure identifies "index" admissions for inclusion in the measure using Medicare Part A inpatient claims from three years for hospitalized Medicare FFS beneficiaries. An "index" admission is any eligible admission to an acute care hospital for an elective primary THA and/or TKA included in the measure. The admission date of the index hospitalization is the starting point for all follow-up, and the clinician that performed the procedure is the one held accountable for the measure outcome (complication or no complication), regardless of whether a patient is transferred to another acute care facility following the procedure (see Section 2.5 for attribution details).

The measure calculates complication rates using a hierarchical logistic regression model to account for the clustering of patients within clinicians while risk-adjusting for differences in patient case mix. The measure calculates the clinician RSCR by producing a ratio of the number of "predicted" to the number of "expected" admissions with a complication for each clinician and then multiplying the ratio by the national unadjusted complication rate.

We calculated the RSCRs for ECs and EC groups and evaluated the measure results for reliability and validity.

2.2 Data Sources

For measure development and testing, we used Medicare administrative claims and enrollment information for patients with hospitalizations between April 1, 2013 and March 31, 2017.

- *Medicare Part A inpatient data* - contains final action claims data submitted by inpatient hospital providers for Medicare FFS beneficiaries for reimbursement of facility costs. Information in this file includes ICD-9/10 diagnosis codes, ICD-9/10 procedure codes, dates of service, hospital provider ID, and beneficiary demographic information. These data are used to identify index hospitalizations, complications, and comorbidities for risk adjustment. These data are also used for providing the identity of the operator.

- *Medicare Part A and Part B outpatient data* - contains final action claims data submitted by inpatient hospital providers for Medicare FFS claims paid for the facility component of surgical or diagnostic procedures, emergency room care, and other non-inpatient services performed in a hospital outpatient department or ambulatory surgical/diagnostic center. These data are used to identify comorbidities for risk adjustment in the 12 months prior to index admission.

- *Medicare Enrollment Database* - contains Medicare beneficiary demographic, benefit/coverage, and vital status information. These data were used to determine FFS enrollment and post-discharge mortality status.

- *Medicare Part B inpatient claim line data from Integrated Data Repository (IDR)* - contains final action claims data for the physician services (regardless of setting) during the index admission, outpatient care, services, and supplies for Medicare FFS beneficiaries. The data also contain total aggregate amount of allowed charges for each EC (TIN/NPI combination) over measure period. For purposes of this project, Part B services included only face-to-face encounters between a care provider and patient. Therefore, the measure does not include information for services such as laboratory tests, medical supplies, or other ambulatory services. Each line in the claim file includes details of services rendered, the identity of the rendering clinician, and the payment the clinician received for each line of service. These data are used to identify the clinician who billed for the procedure and on rare occasions to identify and link the NPI and TIN.

For measure development and testing, we created and used datasets from the April 1, 2013 to March 31, 2017 data as follows:

- To test patient-level model reliability, we used the multiple datasets above, containing data from April 2013 to March 2015. We randomly split the two years of data into two equal samples (Development Sample and Validation Sample).

- To test patient-level model validity/reliability from a temporal perspective, we used data from April 1, 2015 to March 31, 2017 (Temporal Validation Sample).

- To test measure score reliability, we again used multiple datasets:
    - For test-retest reliability, we used data from April 1, 2013 to March 31, 2017. We randomly split the four years of data into two equal samples (Reliability Split Sample 1 and Reliability Split Sample 2). We compared measure scores from the two split samples to calculate reliability.
    - For signal-to-noise reliability, we used a 3-year sample from April 1, 2013 to March 31, 2016 (Medicare Full Sample).

- To assess model performance, calculate measure scores, and calculate performance category results for ECs and EC groups, we used a 3-year sample from April 1, 2013 to March 31, 2016 (Medicare Full Sample). Consistent with the hospital-level measure, this is representative of the amount of data (three years) that would likely be used to calculate the measure under MIPS.

2.3 Cohort Definition

The target population for this measure is Medicare FFS patients undergoing elective primary hip and/or knee procedures (see Table A1 in Appendix A for ICD-10-PCS procedure code list). Both THA and TKA procedures are performed in clinically-similar patient cohorts and for similar indications (osteoarthritis); hospitals typically develop protocols for lower extremity total joint arthroplasty, rather than for THA or TKA individually; the same surgeons frequently perform both procedures; and outcomes are similar.

2.3.1 Inclusion Criteria

Patients eligible for inclusion in the measure are those aged 65 years and older, electively admitted to non-federal acute care hospitals. An index admission is the hospitalization

during which the THA and/or TKA procedure was performed and to which the complication outcome is attributed. Eligible index admissions are identified using ICD-10-PCS procedure codes in Medicare Part A inpatient claims data. For risk adjustment and outcome assessment, patients must have continuous enrollment in Medicare FFS for 12 months prior to the procedure and 90 days after it. The flowchart depicting cohort selection is presented in Figure 1.

The datasets we used for measure development and testing spanned multiple years of data. We therefore used ICD-9-PCS and/or ICD-10-PCS procedure codes, depending on the dataset used. Since the measure would be implemented using data from after October 1, 2015 and use only ICD-10 codes, we provide the ICD-10 codes used to identify eligible THA/TKA procedures. The ICD-10 codes for discharges on or after October 1, 2015 that are used to identify a THA/TKA procedure as non-elective or non-primary and disqualify the admission from cohort inclusion are in Table A1 in Appendix A.

2.3.2 Exclusion Criteria

To identify a homogeneous cohort of patients undergoing elective primary THA and/or TKA procedures, we excluded admissions for patients who, on the index admission, had a principal discharge diagnosis indicative of a non-elective arthroplasty (e.g., hip fracture, mechanical complication). We also excluded patients who had a procedure code for an arthroplasty procedure that was not an elective primary arthroplasty (e.g., partial hip arthroplasty, revision procedures) or represented a different procedure (e.g., hip resurfacing, removal of implanted device).

In order to identify a cohort of elective THA and/or TKA procedures, the measure excludes admissions for patients:

1. With a femur, hip or pelvic fracture coded in the principal discharge diagnosis field for the index admission.
   Rationale: THA procedures are not elective in these patients, and these patients represent a higher risk category for mortality, complication, and readmission.

2. Undergoing partial hip arthroplasty (PHA) procedures (with a concurrent THA/TKA).
   Rationale: Partial arthroplasties are primarily done for hip fractures and are typically performed on patients who are older, frailer, and have more comorbid conditions.

3. Undergoing revision procedures (with a concurrent THA/TKA).
   Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and represent a higher risk category for mortality, complication, and readmission.

4. Undergoing resurfacing procedures (with a concurrent THA/TKA).
   Rationale: Resurfacing procedures are a different type of procedure involving only the joint's articular surface. Resurfacing procedures are typically performed on younger, healthier patients.

5. With a mechanical complication coded in the principal discharge diagnosis field for the index admission.
   Rationale: A complication coded as the principal discharge diagnosis suggests the procedure was more likely the result of a previous procedure and indicates the complication was present on admission. These patients may require more technically complex arthroplasty procedures and may be at increased risk for complications, particularly mechanical complications.

6. With a malignant neoplasm of the pelvis, sacrum, coccyx, lower limbs, or bone/bone marrow or a disseminated malignant neoplasm coded in the principal discharge diagnosis field for the index admission.
   Rationale: Patients with these malignant neoplasms are at increased risk for complications, and the procedure may not be elective.

7. With a procedure code for removal of implanted devices/prostheses.
   Rationale: Elective procedures performed in these patients may be more complicated.

After excluding the above admissions to identify elective primary THA/TKA procedures, the measure also excludes admissions for patients:

8. Who were transferred to the index hospital.
   Rationale: If the patient is transferred from another acute care facility to the hospital where the index procedure occurs, it is likely that the procedure is not elective, or that the admission is associated with an acute condition.

9. Who leave the hospital against medical advice (AMA).
   Rationale: Hospitals and physicians do not have the opportunity to provide the highest quality care for these patients.

10. With more than two THA/TKA procedure codes during the index hospitalization.
    Rationale: Although clinically possible, it is highly unlikely that patients would receive more than two elective THA/TKA procedures in one hospitalization, and this may reflect a coding error.

After applying the exclusion criteria above, we randomly select one index admission for patients with multiple index admissions in a split year (e.g. April 2013 to March 2014 or April 2014 to March 2015). We therefore exclude the other eligible index admissions in that year. Finally, the measure then excludes admissions:

11. Not attributable to an eligible provider.
    Rationale: Only patients with adequate clinician claims for attribution should be included in risk-adjustment model and the measure.

For a list of the ICD-9/10 codes for the exclusion categories described above, please see Table A1 in Appendix A.

2.4 Outcome Definition

The measure assesses a dichotomous yes or no outcome of whether each admitted patient experiences one or more of the complications defined below. Complications other than mortality are counted in the measure only if they occur during the index admission or require a readmission. The measure does not count complications that occur in the outpatient setting and do not require a readmission. The outcome is aligned with CMS's hospital-level THA/TKA complication measure.

2.4.1 Definition of THA/TKA Complications

The measure defines a "complication" as:

- Acute myocardial infarction (AMI), pneumonia, or sepsis/septicemia/shock during the index admission or a subsequent inpatient admission that occurs within seven days from the start of the index admission;
- Surgical site bleeding or pulmonary embolism during the index admission or a subsequent inpatient admission within 30 days from the start of the index admission;
- Death during the index admission or within 30 days from the start of the index admission; or,
- Mechanical complication or periprosthetic joint infection/wound infection during the index admission or a subsequent inpatient admission that occurs within 90 days from the start of the index admission.

Examples of how the measure assesses the complication outcome are:

- Patient is admitted for THA/TKA on January 1, discharged on January 6, and a pulmonary embolism occurs on February 5. The measure will not capture the pulmonary embolism as a complication (as it falls outside of the 30-day time window).
- Patient is admitted for THA/TKA on May 15, remains hospitalized, experiences an AMI on May 25, and is discharged on May 27. The measure will capture the AMI as a complication because it occurred during the index admission (regardless of the seven-day time window).

During development of CMS's hospital-level THA/TKA complication measure, clinical experts agreed with the approach of capturing complications that occur during the index admission but after the defined time window, as such complications likely represent the quality of care provided during the index admission.

Complications are identified using index admission claims and claims for subsequent hospitalizations at short-term acute care hospitals and critical access hospitals, with the exception of death, which is captured through the Medicare Enrollment database.

See Table A2 in Appendix A for hyperlinks to the specific ICD-10 code lists used to define the complications in claims for discharges on or after October 1, 2015, as well as other specifications used to define the complication outcome. The ICD-9 code lists for discharges prior to October 1, 2015 can be found in the 2016 hospital-level THA/TKA complication measure updates and specifications report posted on *QualityNet.*

2.4.2 Outcome Timeframe

The complication-specific follow-up periods are based on the input of clinical experts informed by analyses of 90-day trends in complication rates post-procedure during development of CMS's hospital-level THA/TKA complication measure[1] (Figure 1):

- The follow-up period for AMI, pneumonia, and sepsis/septicemia/shock complications is seven days from the admission date of the index admission because these conditions are more likely to be attributable to the procedure if they occur within the first week after the procedure. Additionally, analyses indicated a sharp decrease in the rate of these complications after seven days.

- Death, surgical site bleeding, and pulmonary embolism are followed for 30 days from the start of the index admission because clinical experts agree these complications are still likely attributable to the hospital performing the procedure during this period and rates for these complications remain elevated until roughly 30 days post-admission.

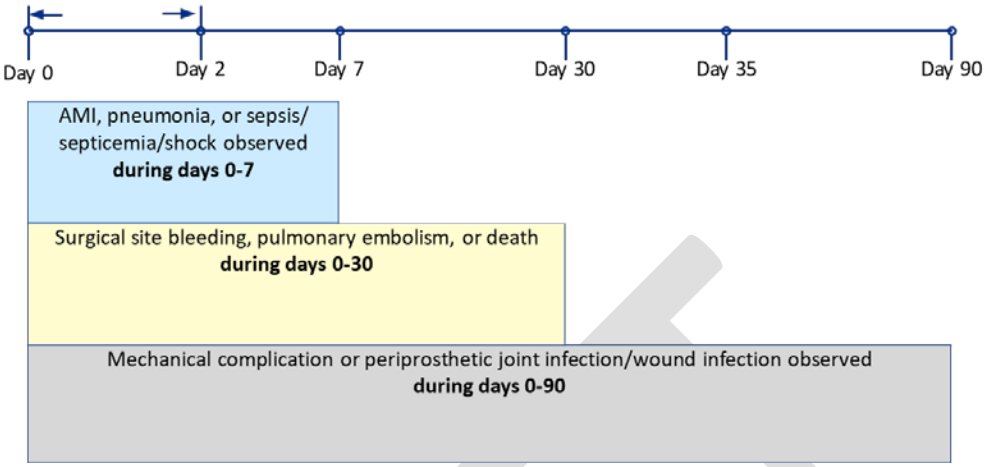
- The measure follow-up period is 90 days from the start of the index admission for mechanical complications and periprosthetic joint infection/wound infection. Experts agree that mechanical complications and periprosthetic joint infection/wound infections due to the index THA/TKA occur up to 90 days following THA/TKA.

Note that the measure captures all complications occurring during the index admission, regardless of when they occur, with the exception of complications that are coded as present on admission (POA). Not capturing complications coded as POA prevents classifying a condition as a complication of care if it was present at the time the patient was admitted as an inpatient.

In determining whether a complication other than death occurred during a subsequent admission within the complication-specific follow-up periods described above, the measure uses the claim "FROM" date from the subsequent admission claim, which is the date that admission started (that is, the date the patient first received care at that hospital within three days of that admission). Thus, in the case where a patient began their subsequent admission with an ED visit, observation stay, or care received in another outpatient location within the same facility, the case was converted to inpatient admission by that hospital within three days of that outpatient encounter, and the care is combined into one claim, the date the outpatient care started would be used to determine the timing of the subsequent admission.

Figure 1. Outcome timeframe for MIPS THA/TKA measure.

Day 0 = date of admission to the hospital



2.4.3 Outcome Attribution for Sequential Elective Primary THA/TKA Procedures

The measure randomly selects one index admission per patient per split year. However, it is still possible for two procedures to have overlapping outcome periods; in this case the outcome is assigned to the later procedure. For example, a patient is admitted on November 15, 2013 for an elective primary THA, and this patient is admitted again on January 15, 2014 for another elective primary THA. The patient is then readmitted for a mechanical complication on January 25, 2014. To avoid assigning the mechanical complication to both index admissions (since it falls within the specified follow-up period for both), the measure will assign the mechanical complication to the second index admission (January 15, 2014). This assignment of the complication outcome is only applicable in cases where a complication occurs after a second elective primary procedure but occurs within the follow-up period for both the first and second index admissions. If a complication occurs during the index admission, it will be assigned to that index admission. In other words, when two index THA/TKA admissions occur but their admission dates are fewer than 90 days apart and the second index admission is followed by one or more readmissions, any complications associated with those readmissions will only be attributed to the second index admission.

2.5 Attribution

2.5.1 Approach to Attribution

The measure attributes the outcome for each patient in the cohort to a single clinician. Each patient is attributed to the clinician who bills for the Part B Physician/Supplier claim for the THA or TKA procedure (hereafter, the Billing Surgeon) during the index admission. Conceptually, this is the clinician with the primary responsibility for the procedure and procedure-related care. In practice, however, patients may have different claims for the same procedure. In order to resolve such ambiguities, the Billing Surgeon is assigned through an algorithm (Figure B1 in Appendix B) as described below.

The algorithm uses billing claims to identify clinician(s) who bill for a THA (CPT® code 27130) or TKA (CPT®® code 27447 or CPT® code 27446 if CPT® code 27447 not billed) (steps 1-3 below). These CPT® codes are representative of the THA and/or TKA procedures included in the measure cohort.

1. If only one clinician bills for a THA (CPT® code 27130) or TKA (CPT® code 27446 or 27447) for a patient, the algorithm identifies and assigns this individual as the Billing Surgeon.

2. If two or more clinicians bill for THA/TKA procedures (CPT® 27130, 27447, or 27446), the algorithm seeks to identify a 'key' physician among them. The algorithm identifies and excludes assignment to clinicians who were assistants-at-surgery (assistant surgeon with CPT® modifier 80 or 82, minimum assistant surgeon with CPT® modifier 81). In this step, the algorithm assigns the Billing Surgeon as the clinician who billed for a THA or TKA procedure and is not an assistant-at-surgery.

3. If a single clinician who is not an assistant-at-surgery could not be identified for assignment, then the algorithm identifies whether there is a single clinician who was an orthopedic surgeon (Medicare Specialty Code 20) and assigns this as the Billing Surgeon.

4. If the algorithm cannot identify a Billing Surgeon, it identifies whether an Operator is listed on the institutional claim. The algorithm then defaults assignment to the Operator listed on the institutional claim. To identify the unique TIN/NPI combination for the Operator, the Operator's NPI is matched to the TIN with the most Part B allowed charges during the index admission or during the measurement year if the EC did not bill during the index admission.

Finally, if a Billing Surgeon or Operator cannot be identified with the steps above, the patient is not assigned to a Billing Surgeon and excluded from the measure.

Clinical experts and the TEP supported attribution to the Billing Surgeon using this algorithm. For a description of alternative candidate attribution rules considered but not selected, please see Appendix B.

2.5.2 Eligible Clinicians and Eligible Clinician Groups

For the purposes of development and testing we have defined 'eligible clinicians' (ECs) as unique combinations of NPI and TIN. Thus, a single clinician may be measured two or more times if they file Medicare claims under two or more TINs. Each attribution rule includes an algorithm for identifying a unique TIN/NPI combination.

The unique TIN/NPI combinations can be directly aggregated into groups of clinicians with the same TIN. We refer to these as MIPS EC groups. It should be noted that these only approximately align with practice groups. Note also that patients can only be assigned to groups by way of an EC (a TIN/NPI combination), and thus these are by default groups with at least one EC. Within the MIPS, an EC "group" must include two or more ECs, at

least one of which participates in MIPS. Because we cannot identify non-attributed ECs at each TIN, we report all TINs regardless of the number of attributed ECs.

2.5.3 Volume Requirements

It is impractical to measure outcomes for EC or EC groups which are assigned a small number of patients; though technically it is feasible to construct estimates based on as few as one patient, practically we would want to measure only those entities with adequate volume to construct moderately reliable estimates. Thus, we used reliability estimates to determine a minimum reporting volume of 25 cases for ECs and EC groups. This minimum reporting volume of 25 cases is consistent with the cutoff used for CMS's hospital-level THA/TKA complication measure and demonstrates acceptable preliminary reliability; further details about reliability testing results and minimum case volumes are included in the Results section below.

Some stakeholders expressed concern that not measuring low volume providers will fail to incentivize optimal care for all patients, as low surgical case volume might be associated with poorer quality care. While a minimum reporting volume of 25 cases does decrease the number of ECs and EC groups captured, it does not significantly decrease the number of patients captured. Even with a 25 case minimum, 93% to 98% of all patients are retained in the measure (for ECs and EC groups, respectively). Further, this measure assesses only Medicare FFS beneficiaries and does not include patients with other insurance types. Therefore, we cannot accurately determine whether or not clinicians are truly 'low volume' as they may have performed cases on non-Medicare patients not captured in our data.

To be consistent with the hospital-level measure and reflect that we cannot truly identify low volume clinicians, we used a minimum reporting volume of 25 cases for ECs and EC groups and did not adjust for volume in the measure calculation.

2.6 Risk Adjustment

The goal of risk adjustment is to account for patient age, whether the patient had one or two procedures, and comorbid conditions that are clinically relevant and have strong relationships with the outcome, while illuminating important quality differences between hospitals. The measure adjusts for case-mix differences based on the clinical status of the patient at the time of admission. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in the risk adjustment. Although they may increase the risk of mortality and complications, including them as covariates in risk adjustment could attenuate the measure's ability to characterize the quality of care delivered by hospitals. For each patient, risk-adjustment variables are obtained from inpatient, outpatient, and physician Medicare administrative claims data extending 12 months prior to the index admission and all claims data for the index admission itself.

In keeping with our key principle regarding alignment with the hospital-level measure, and because the hospital risk model was developed and validated at the patient-level using the same cohort adopted for this MIPS THA/TKA complication measure, we used the same risk factors as used by CMS's hospital-level THA/TKA complication measure.

Comorbidities for inclusion in risk adjustment are identified in administrative claims during the 12 months prior to and including the index admission. To assemble the more than 16,000 ICD-9 codes and 70,000 ICD-10 codes into clinically coherent variables for risk adjustment, the measure employs the publicly available CMS hierarchical condition categories (CCs) to group codes into CCs [37,38] and selects comorbidities for inclusion in risk adjustment on the basis of clinical relevance and statistical significance.

Table A3 in Appendix A lists the conditions not adjusted for if they only appear in the index admission and not in the 12 months prior to admission. The CCs outlined in this table are used to identify risk variables in claims for discharges on or after October 1, 2015 as well as discharges prior to October 1, 2015. The ICD-10 code lists referenced in the tables that are used to identify certain risk variables (e.g., Post traumatic osteoarthritis) in discharges on or after October 1, 2015 are posted on *QualityNet*. For a list of ICD-9 codes used to identify these variables for discharges prior to October 1, 2015, please refer to the 2016 procedure-specific complication measure updates and specifications report posted on *QualityNet*.

Additionally, the measure does not adjust for the patients' admission source or their discharge disposition (e.g., skilled nursing facility) because these factors are associated with the structure of the healthcare system, not solely patients' clinical risk factors. Regional differences in resource availability and practice patterns may exert an undue influence on model results. Moreover, the accuracy of these admission and discharge disposition codes is not known. The measure also does not adjust for socioeconomic status (SES), race, or ethnicity. Variation in quality associated with these characteristics may be indicative of disparities in the quality of the care provided to vulnerable populations and adjusting for these factors would obscure these disparities. The measure does not adjust for provider characteristics either, since this would hold different types of providers to different quality standards, and because such characteristics may exist on a causal pathway to the outcome, rather than act as confounders. This approach is consistent with CMS's hospital-level THA/TKA complication measure. The intent is for the measure to adjust for age and clinical characteristics while illuminating important quality differences. CMS's hospital-level THA/TKA complication measure was endorsed by the NQF without adjustment for patient-level SES factors. For more information about this decision, please refer to the *NQF website*.

2.7 Statistical Approach to Measure Calculation

The measure risk adjusts EC and EC group complication rates using a hierarchical logistic regression model. In brief, the approach simultaneously models two levels (patient and EC/EC group) to account for the variance in patient outcomes within and between ECs or between EC groups. The patient level models the log-odds of a complication adjusting for age, sex, selected clinical covariates, and an EC-specific or EC group-specific intercept. The second level models the EC-specific or EC group-specific intercepts as arising from a normal distribution. The EC-specific or EC group-specific intercept represents the underlying risk of a complication at that entity after accounting for patient risk. If there were no differences among ECs or EC groups, then after adjusting for patient risk, the EC or EC group intercepts should be identical across all entities.

After regressing the risk factors and the EC or EC group's specific intercept on the risk of a complication, the predicted probability of a complication is calculated by summing the estimated regression coefficients multiplied by the patient characteristics, adding the estimated

EC-specific or EC group-specific intercept, and inverse logit transforming this value. These are summed over each EC or EC group to get the predicted number of complications. The expected probability of a complication is obtained by summing the estimated regression coefficients multiplied by the patient characteristics observed in the EC or EC group mix, adding the estimated average EC or EC group intercept, and inverse logit transforming this value. These are summed over each EC or EC group to get the expected number of complications.

The risk adjusted complication rate is calculated as the ratio of the predicted number of complications to the expected number of complications, multiplied by the national unadjusted complication rate. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular EC or EC group's performance given its case mix to an average EC or EC group with the same case mix. A lower ratio therefore indicates a lower-than-expected complication rate or better quality and a higher ratio indicates a higher-than-expected complication rate or worse quality.

Please refer to Appendix C for further technical details.

2.7.1 Clinician and Clinician Group Performance Reporting

To categorize EC and EC group performance, we estimated each EC's or EC group's risk-adjusted complication rate and the corresponding 95% confidence interval (CI). We assigned EC or EC groups to a performance category by comparing each hospital's RCSR interval estimate to zero. For a full description of calculation, see Appendix C. Comparative performance for EC or EC groups with 25 or more eligible cases were classified as follows:

- "No Different than the National Rate" if the 95% interval estimate surrounding the EC or EC group's rate includes the national observed complication rate.
- "Worse than the National Rate" if the entire 95% interval estimate surrounding the EC or EC group's rate is higher than the national observed complication rate.
- "Better than the National Rate" if the entire 95% interval estimate surrounding the EC or EC group's rate is lower than the national observed complication rate.

3. MEASURE TESTING

We used the datasets described in Section 2.2 for measure development and testing. We report the measure scores and performance for three years of data because CMS would use three years of data for measure calculation and reporting; this is consistent with CMS's hospital-level THA/TKA complication measure.

3.1 Cohort

We provide the cohort derivation and index cohort based on three years of data because CMS would use three years of data for measure calculation and reporting; this is consistent with CMS's hospital-level THA/TKA complication measure. The dataset used was the Medicare Full Sample (3-year sample from April 1, 2013 to March 31, 2016).

3.2 Risk-adjustment Variables

We report the frequency of each risk variable for all datasets. This provides a description the patients included in the different samples, informing both face validity and reliability considerations. We present frequencies in the following datasets: Full Sample, Development Sample, Validation Sample, and Temporal Validation Sample.

3.3 Attributed Eligible Clinicians and Eligible Clinician Groups

For each attribution role, as well as for unique ECs (TIN/NPIs), we report the distribution of admissions assigned across ECs using the Medicare Full Sample. We also report the percent of admissions that could not be assigned and the total number of distinct ECs in that role. We replicate this for EC groups.

3.4 Model Parameter Estimates

We report the coefficient and variance estimates for the models using the Medicare Full Sample. Direction and magnitude of these provide face validity for the risk adjustment.

3.5 Point Estimates of Risk-standardized Complication Rates

After estimating the models reported in Section 3.4, we use the results to construct RSCRs for ECs and EC groups. We report the distributions of RSCRs for each entity using the Medicare Full Sample since this is representative of the amount of data that would be used to calculate the measure under MIPS. These data provide evidence of performance variation.

3.6 Distributions of Eligible Clinicians and Eligible Clinician Groups by Performance Category

After bootstrapping the RSCRs, we used the 95% CIs to identify ECs and EC groups which have RSCRs that are high (95% CI above the mean) or low (95% CI below the mean). The existence of such performance outliers provides additional evidence of performance variation and measure utility. We present results using the 3-year dataset, Medicare Full Sample, since this is representative of the amount of data that would be used to calculate the measure under MIPS.

3.7 Reliability

3.7.1 Data Element Reliability

The measure uses only those data elements from claims data that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of the CMS auditing and billing policies, and we seek to avoid variables which do not meet this standard.

In addition, CMS has in place several auditing programs used to assess overall claims coding accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential for payment.

We further assessed the reliability of the patient-level model by comparing coefficients from logistic regression models in the Development Sample, Validation Sample, and Temporal Validation Sample. We computed five summary statistics to assess model performance: calibration (a measure of over-fitting),[*] discrimination in terms of predictive ability, discrimination in terms of area under the receiver operating curve (ROC), distribution of residuals, and model chi-square.

Over-fitting refers to the phenomenon in which a model describes the relationship between predictive variables and outcome well in the development dataset but fails to provide valid predictions in new patients. Since the γ0 in the validation sample is close to zero and the γ1 is close to one in each of the models, there is little evidence of over-fitting.

Discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects. Therefore, we would hope to see a wide range between the lowest decile and highest decile, which these models show.

The c-statistic is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome. A c-statistic of 0.50 indicates random prediction, implying all patient risk factors are useless; a value of 1 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play no role in patients' outcomes. While higher C-statistic is desirable, we do not want to maximize c-statistic by adjusting for factors that should not

[*] Over-fitting Indices (γ_0, γ_1) provide evidence of over-fitting and require several steps to calculate. Let *b* denote the *estimated vector* of regression coefficients. *Predicted Probabilities* ($\hat{p}$) = 1/(1+exp{-Xb}), and *Z = Xb* (e.g., the linear predictor that is a scalar value for everyone). A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample; e.g., Logit(P(Y=1|Z)) = $\gamma_0 + \gamma_1 Z$. Estimated values of γ_0 far from 0 and estimated values of γ_1 far from 1 provide evidence of over-fitting.

be adjusted for; for example, we do not want to include in-hospital complications as a risk factor, even if it produces a higher c-statistic.

The model residuals are the difference between what the model predicts for each patient and the observed outcome. If they are not distributed symmetrically around zero, or if most values are not near zero, this indicates that the model assumptions are not met.

The model chi-square is a statistic which represents the degree to which the model explains the observed data.

3.7.2 Measure Score Reliability

We considered two notions of reliability when evaluating the MIPS THA/TKA complication measure. The 'test-retest' reliability is the degree to which repeated measurements of the same entity agree with each other. For measures of EC or EC group performance, the measured entity is naturally the EC or EC group, and reliability is the extent to which repeated measurements of the same entity give similar results. In line with this thinking, our approach to assessing reliability is to measure each EC or EC group once using a random subset of patients, then measure the same entity again using a second random subset exclusive of the first, and finally compare the agreement between the two resulting performance measures across all entities.[39]

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of the patients within each provider, calculated the measure for each provider, and repeated the calculation using the second half. Thus, each EC or EC group is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the provider, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (ICC[2,1]).[40] Because the split samples corresponded to fewer patients than would be typically included in a full measurement period (three years), we used the Spearman-Brown formula to adjust the ICC[2,1] for the reduced measurement period.[41,42] We assessed the testing results according to conventional standards (Landis and Koch, 1977).[43]

A second notion of reliability is that of signal to noise. The variation between entities ('signal') comprises the total variation ('noise' and 'signal') in the outcome. This is because the reliability of any one EC or EC group's measure score will vary depending on the number of patients. We use the formula presented by Adams and colleagues (2010) to calculate EC/EC group-level reliability scores.[44] To estimate the overall signal and noise, we used the estimated covariance from a hierarchical generalized linear model (HGLM) as the between-entity variance $\tau^2$and $\frac{\pi^2}{3}$ as within-entity variance σ_j^2 for each entity (EC or EC group) j. Then for each entity and entity group we calculated $\rho_j = \tau^2/(\tau^2 + \sigma_j^2)$. We then used the equation:

$$R_j = n_j\rho_j/(1+(n_j-1)\rho_j)$$

to calculate the reliability of each entity measurement; we report the mean R_j over all entities for different minimum volumes n_j; n_j is the number of observations under each entity or entity group.

3.8 Validity

3.8.1 Measure Score Validity

Validity of Attribution Rules

To inform selection of the attribution rule for this measure, we undertook several steps including: review of NQF recommendations, conducting an environmental scan of existing attribution methods used by CMS, reviewing the literature, understanding claims patterns, and consulting with clinicians and a national TEP (see details in Appendix B). We also held three TEP meetings to review the approach to identifying and evaluating candidate attribution rules and to obtain input on the selection of the final attribution rule.

Face Validity of MIPS Eligible Clinician and Eligible Clinician Group Measure Scores

Following presentation and review of the final measure specification, results, and testing, CORE surveyed the 19 members of the TEP regarding validity and usability of the MIPS HKC measure. We asked them to consider two statements

The risk-standardized complication rates obtained from the MIPS hip/knee complications (HKC) measure as specified:

1. *Are valid and useful measures of MIPS eligible clinician and MIPS eligible clinician group quality of care.*
2. *Will provide MIPS eligible clinicians and MIPS eligible clinician groups with information that can be used to improve their quality of care.*

They were asked to report their agreement with each statement on a 6-point scale, representing a range from Strongly Disagree to Strongly Agree.

4. RESULTS

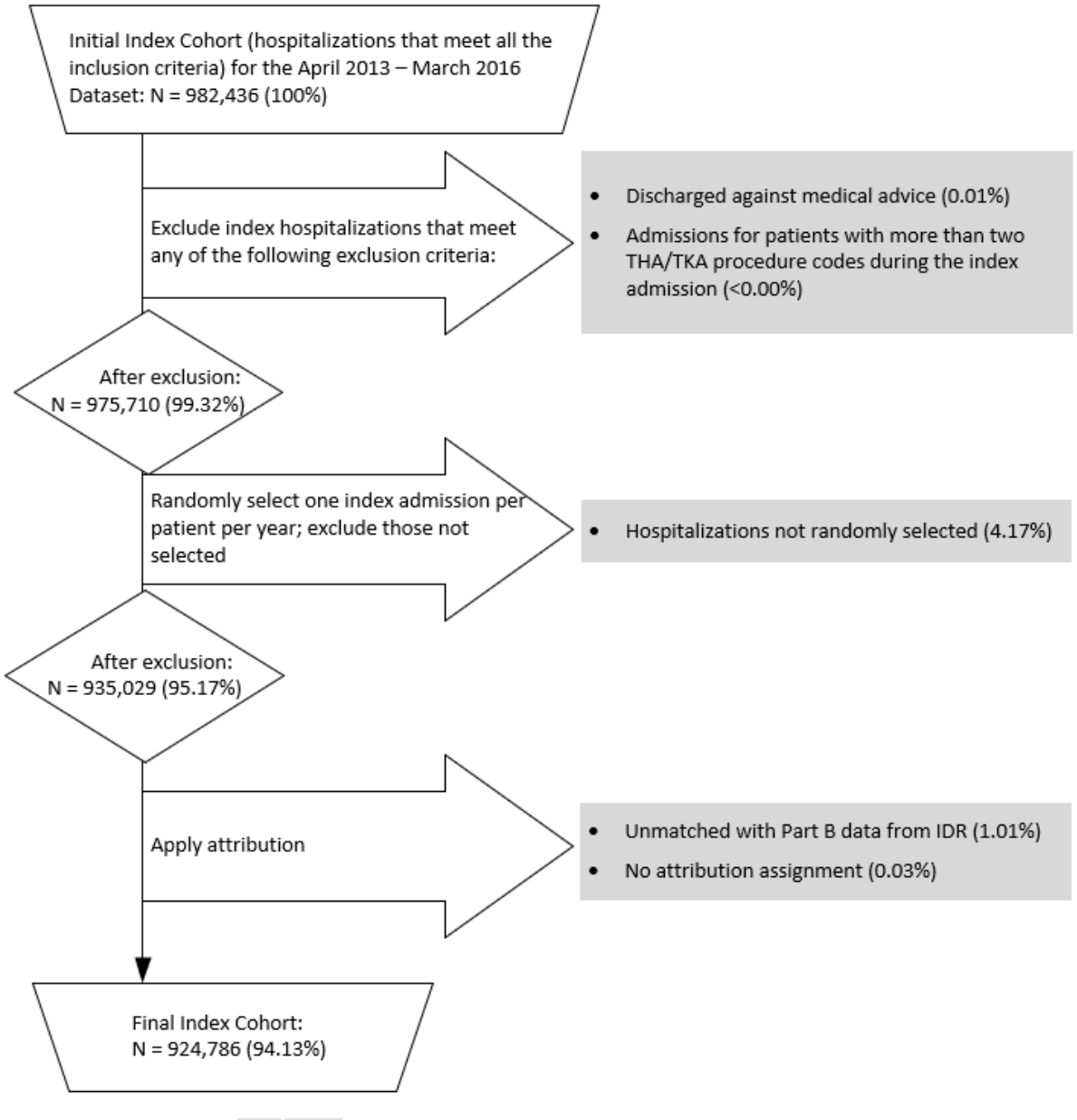
In this section, we present testing results for the MIPS THA/TKA complication measure.

4.1 Cohort

The exclusion criteria for this measure are presented in Section 2.3.2. The percentage of THA/TKA admissions that met each exclusion criterion in the 3-year dataset, Medicare Full Sample, is presented in Figure 2. We present the index cohort for this dataset since the measure cohort would be constructed using three years of index admissions in any future use.

Admissions may have been counted in more than one exclusion category because they are not mutually exclusive.

Figure 2. THA/TKA cohort exclusions (dataset: Medicare Full Sample [April 2013-March 2016])



4.2 Risk-adjustment Variables

Consistent with the CMS's hospital-level THA/TKA complication measure, the final risk-adjustment model included 33 variables. Table 1 shows the frequency of each risk variable in the Medicare Full Sample, Development Sample, Validation Sample, and Temporal Validation Sample. Compared to the Development Sample, the mean age of patients and the frequencies of risk-adjustment variables in the Validation Sample, Temporal Validation Sample, and Medicare Full Sample were similar.

Table 1. Frequency of THA/TKA model variables in full sample, development sample, validation sample, and temporal validation sample

Variable	Medicare full sample (April 2013 – March 2016)	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Number of admissions	924,786	302,561	302,519	653,598
Number of eligible clinicians	16,755	13,565	13,513	14,794
Number of eligible clinician groups	6,118	5,392	5,389	5,291
Unadjusted complication rate (%)	2.8%	2.8%	2.9%	2.5%
Mean age minus 65 (SD)	9.49 (6.0)	9.54 (6.0)	9.52 (6.0)	9.36 (5.9)
Male	37.2%	37.2%	37.1%	37.3%
Index admissions with an elective THA procedure	32.0%	31.6%	31.4%	33.0%
Number of procedures (2 vs. 1)	2.2%	2.3%	2.3%	2.0%
Metastatic cancer or acute leukemia (CC 8)	0.5%	0.5%	0.5%	0.6%
Other major cancers (CC 9-12)	12.9%	13.0%	12.9%	12.7%
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	18.0%	17.9%	7.8%	18.2%
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	28.1%	28.2%	28.4%	27.6%
Protein-calorie malnutrition (CC 21)	0.7%	0.7%	0.7%	0.7%
Bone/joint/muscle infections/necrosis (CC 39)	2.8%	2.7%	2.7%	2.7%
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	9.5%	9.5%	9.5%	9.8%
Osteoarthritis of hip or knee (CC 42)	96.3%	96.2%	96.3%	96.6%
Osteoporosis and other bone/cartilage disorders (CC 43)	23.9%	23.9%	24.0%	23.6%
Dementia or other specified brain disorders (CC 51-53)	4.1%	4.1%	4.2%	4.1%
Major psychiatric disorders (CC 57-59)	4.7%	4.8%	4.7%	4.7%
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	1.1%	1.0%	1.0%	1.2%
Cardio-respiratory failure and shock (CC 84), plus ICD-10-CM codes R09.01 and R09.02	2.5%	2.5%	2.4%	2.5%
Coronary atherosclerosis or angina (CC 88-89)	25.7%	26.3%	26.2%	24.5%
Stroke (CC 99-100)	2.1%	2.1%	2.1%	1.9%

Variable	Medicare full sample (April 2013 – March 2016)	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Vascular or circulatory disease (CC 106-109)	21.7%	21.7%	21.9%	21.5%
Chronic obstructive pulmonary disease (COPD) (CC 111)	12.7%	12.9%	13.0%	12.1%
Pneumonia (CC 114-116)	4.0%	4.0%	4.0%	3.9%
Pleural effusion/pneumothorax (CC 117)	1.4%	1.4%	1.4%	1.4%
Dialysis status (CC 134)	0.2%	0.2%	0.2%	0.2%
Renal failure (CC 135-140)	12.9%	12.6%	12.6%	13.7%
Decubitus ulcer or chronic skin ulcer (CC 157-161)	2.4%	2.4%	2.5%	2.2%
Trauma (CC 166-168, 170-173)	4.9%	4.9%	4.9%	4.8%
Vertebral fractures without spinal cord injury (CC 169)	1.2%	1.2%	1.2%	1.1%
Other injuries (CC 174)	27.9%	27.9%	28.0%	27.3%
Major complications of medical care and trauma (CC 176-177)	5.3%	5.2%	5.2%	5.1%
Morbid obesity (CC 22)	8.1%	7.8%	8.0%	8.8%
Other congenital deformity of hip (joint) (ICD-9 code 755.63, ICD-10 code Q65.89, Q65.9)	0.2%	0.2%	0.2%	0.2%
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16, ICD-10 code M12.551, M12.552, M12.559)	0.4%	0.4%	0.4%	0.2%

4.3 Attributed Clinicians and Clinician Groups

Table 2 shows the distribution of admissions assigned across ECs and EC groups.

Table 2. Distribution of volumes for MIPS ECs and EC groups with more than 25 cases (dataset: Medicare full sample [April 2013 – March 2016])

Characteristic	Eligible clinicians	Eligible clinician groups
Number of entities	7,928	3,572
Number of admissions per entity		
Mean (SD)	108 (110)	253 (406)
Median (IQR)	69 (41 – 129)	109 (53 – 282)
Range (min. – max.)	25 – 1,504	25 – 5,463

4.4 Model Parameter Estimates

Table 3 and Table 4 show model parameter estimates, adjusted odds ratios (ORs), and 95% CIs for MIPS ECs and EC groups, respectively.

Table 3. MIPS eligible clinicians: model estimates, adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for the THA/TKA hierarchical logistic regression model (dataset: Medicare full sample [April 2013 – March 2016])

Variable	Estimate	Standard error	Odds ratio (95% confidence interval)
Mean age minus 65 (SD)	0.03	0.00	1.03 (1.03 - 1.03)
Male	0.11	0.01	1.12 (1.08 - 1.15)
Index admissions with an elective THA procedure	0.35	0.01	1.41 (1.38 - 1.45)
Number of procedures (2 vs. 1)	0.55	0.04	1.73 (1.6 - 1.87)
Metastatic cancer or acute leukemia (CC 8)	0.07	0.08	1.07 (0.92 - 1.26)
Other major cancers (CC 9-12)	0.00	0.02	1 (0.96 - 1.03)
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	-0.06	0.02	0.94 (0.91 - 0.98)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	0.11	0.01	1.11 (1.08 - 1.14)
Protein-calorie malnutrition (CC 21)	0.86	0.04	2.36 (2.16 - 2.58)
Bone/joint/muscle infections/necrosis (CC 39)	0.14	0.03	1.15 (1.08 - 1.22)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	0.13	0.02	1.14 (1.1 - 1.19)
Osteoarthritis of hip or knee (CC 42)	-0.06	0.03	0.94 (0.88 - 1.01)
Osteoporosis and other bone/cartilage disorders (CC 43)	-0.01	0.02	0.99 (0.96 - 1.02)
Dementia or other specified brain disorders (CC 51-53)	0.20	0.03	1.23 (1.16 - 1.29)
Major psychiatric disorders (CC 57-59)	0.30	0.03	1.36 (1.29 - 1.42)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	0.16	0.05	1.18 (1.07 - 1.3)
Cardio-respiratory failure and shock (CC 84), plus ICD-10-CM codes R09.01 and R09.02	0.14	0.03	1.15 (1.08 - 1.23)
Coronary atherosclerosis or angina (CC 88-89)	0.26	0.01	1.3 (1.26 - 1.33)
Stroke (CC 99-100)	0.03	0.04	1.03 (0.96 - 1.11)
Vascular or circulatory disease (CC 106-109)	0.12	0.02	1.13 (1.1 - 1.16)

Variable	Estimate	Standard error	Odds ratio (95% confidence interval)
Chronic obstructive pulmonary disease (COPD) (CC 111)	0.43	0.02	1.54 (1.49 - 1.59)
Pneumonia (CC 114-116)	0.21	0.03	1.24 (1.17 - 1.31)
Pleural effusion/pneumothorax (CC 117)	-0.07	0.04	0.94 (0.86 - 1.02)
Dialysis status (CC 134)	0.25	0.10	1.29 (1.07 - 1.56)
Renal failure (CC 135-140)	0.29	0.02	1.33 (1.29 - 1.38)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	0.29	0.03	1.33 (1.25 - 1.42)
Trauma (CC 166-168, 170-173)	0.17	0.03	1.18 (1.12 - 1.25)
Vertebral fractures without spinal cord injury (CC 169)	0.10	0.05	1.11 (1.01 - 1.22)
Other injuries (CC 174)	0.06	0.01	1.06 (1.03 - 1.09)
Major complications of medical care and trauma (CC 176-177)	0.16	0.02	1.17 (1.12 - 1.23)
Morbid obesity (CC 22)	0.47	0.02	1.61 (1.54 - 1.67)
Other congenital deformity of hip (joint) (ICD-9 code 755.63, ICD-10 code Q65.89, Q65.9)	0.18	0.12	1.2 (0.94 - 1.52)
Post-traumatic osteoarthritis (ICD-9 codes 716.15, 716.16, ICD-10 code M12.551, M12.552, M12.559)	0.27	0.09	1.3 (1.09 - 1.56)

Table 4. MIPS eligible clinician groups: model estimates, adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for the THA/TKA hierarchical logistic regression model (dataset: Medicare Full Sample [April 2013 – March 2016])

Variable	Estimate	Standard Error	Odds ratio (95% confidence interval)
Mean age minus 65 (SD)	0.03	0.00	1.03 (1.03 - 1.03)
Male	0.11	0.01	1.12 (1.08 - 1.15)
Index admissions with an elective THA procedure	0.34	0.01	1.41 (1.37 - 1.44)
Number of procedures (2 vs. 1)	0.56	0.04	1.75 (1.62 - 1.89)
Metastatic cancer or acute leukemia (CC 8)	0.08	0.08	1.08 (0.92 - 1.27)
Other major cancers (CC 9-12)	-0.01	0.02	0.99 (0.96 - 1.03)
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	-0.06	0.02	0.94 (0.91 - 0.97)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	0.11	0.01	1.11 (1.08 - 1.14)
Protein-calorie malnutrition (CC 21)	0.86	0.04	2.36 (2.17 - 2.57)
Bone/joint/muscle infections/necrosis (CC 39)	0.14	0.03	1.15 (1.08 - 1.23)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	0.13	0.02	1.14 (1.1 - 1.19)
Osteoarthritis of hip or knee (CC 42)	-0.07	0.03	0.94 (0.88 - 1)
Osteoporosis and other bone/cartilage disorders (CC 43)	-0.01	0.02	0.99 (0.96 - 1.02)
Dementia or other specified brain disorders (CC 51-53)	0.20	0.03	1.22 (1.16 - 1.29)
Major psychiatric disorders (CC 57-59)	0.30	0.03	1.36 (1.29 - 1.43)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	0.17	0.05	1.18 (1.07 - 1.3)
Cardio-respiratory failure and shock (CC 84), plus ICD-10-CM codes R09.01 and R09.02	0.14	0.03	1.15 (1.08 - 1.23)
Coronary atherosclerosis or angina (CC 88-89)	0.26	0.01	1.29 (1.26 - 1.33)
Stroke (CC 99-100)	0.03	0.04	1.03 (0.96 - 1.11)
Vascular or circulatory disease (CC 106-109)	0.12	0.02	1.13 (1.09 - 1.16)
Chronic obstructive pulmonary disease (COPD) (CC 111)	0.43	0.02	1.54 (1.49 - 1.59)
Pneumonia (CC 114-116)	0.21	0.03	1.24 (1.17 - 1.3)
Pleural effusion/pneumothorax (CC 117)	-0.07	0.04	0.93 (0.86 - 1.02)

Variable	Estimate	Standard Error	Odds ratio (95% confidence interval)
Dialysis status (CC 134)	0.24	0.10	1.28 (1.06 - 1.54)
Renal failure (CC 135-140)	0.29	0.02	1.34 (1.29 - 1.38)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	0.29	0.03	1.33 (1.25 - 1.42)
Trauma (CC 166-168, 170-173)	0.17	0.03	1.19 (1.13 - 1.25)
Vertebral fractures without spinal cord injury (CC 169)	0.10	0.05	1.11 (1.01 - 1.22)
Other injuries (CC 174)	0.06	0.01	1.06 (1.03 - 1.09)
Major complications of medical care and trauma (CC 176-177)	0.16	0.02	1.17 (1.12 - 1.23)
Morbid obesity (CC 22)	0.48	0.02	1.61 (1.55 - 1.68)
Other congenital deformity of hip (joint) (ICD-9 code 755.63, ICD-10 code Q65.89, Q65.9)	0.18	0.12	1.2 (0.95 - 1.52)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16, ICD-10 code M12.551, M12.552, M12.559)	0.27	0.09	1.31 (1.1 - 1.56)

4.5 Point Estimates of Risk-standardized Complication Rates

The national observed complication rate in the Medicare Full Sample was 3.0% for ECs and 3.2% for EC groups.

Table 5 shows the distribution of MIPS EC and EC group RSCRs. The mean RSCR was 2.8% for ECs and EC groups between April 2013 and March 2016. The median RSCR was 2.7% for ECs (IQR: 2.4% - 3.2%) and 2.8% for EC groups (IQR: 2.5% - 3.1%). Table 6 shows the between-entity variance for the combined three-year dataset; it was 0.183 for ECs (standard error [SE]: 0.01) and 0.114 for EC groups (SE: 0.01). All results are restricted to ECs or EC groups with at least 25 discharges in the 3-year measurement period. These data provide supportive evidence of performance variation.

Table 5. Distribution of unadjusted and risk-standardized complication rates for MIPS eligible clinicians (ECs) and eligible clinician groups (EC groups) with more than 25 cases (dataset: Medicare Full Sample [April 2013 – March 2016])

Characteristic	Eligible clinicians	Eligible clinician groups
Number of entities	7,928	3,572
Unadjusted outcome rate (%)		
Mean (SD)	3.0% (2.7%)	3.2% (2.4%)
Median (IQR)	2.6% (1.3% - 4.2%)	2.8% (1.8% - 4.1%)
Range (min. – max.)	0.0% - 22.2%	0.0% - 22.22%
Risk-standardized complication rate (%)		
Mean (SD)	2.8% (0.65%)	2.8% (0.51%)

Characteristic	Eligible clinicians	Eligible clinician groups
Median (IQR)	2.7% (2.4% - 3.2%)	2.8% (2.5% - 3.1%)
Range (min. – max.)	1.2% - 7.2%	1.4% - 5.7%

Table 6. Between-entity variance for MIPS eligible clinicians and eligible clinician groups with more than 25 cases (dataset: Medicare Full Sample [April 2013 – March 2016])

Characteristic	Eligible clinicians	Eligible clinician groups
Between-entity variance (SE)	0.183 (0.009)	0.114 (0.008)

Figure 3 shows the range of RSCR distribution for ECs with more than 25 cases. While the median RSCR for ECs is relatively low, there is a wide range of distribution (range = 1.2% – 7.2%) with several clinicians receiving higher-than-expected complication rates. This variation in RSCRs suggests opportunity for improvement.

Figure 3. Distribution of risk-standardized complication rates (dataset: Medicare full sample [April 2013 – March 2016]) - hierarchical logistic regression model – MIPS ECs with more than 25 cases

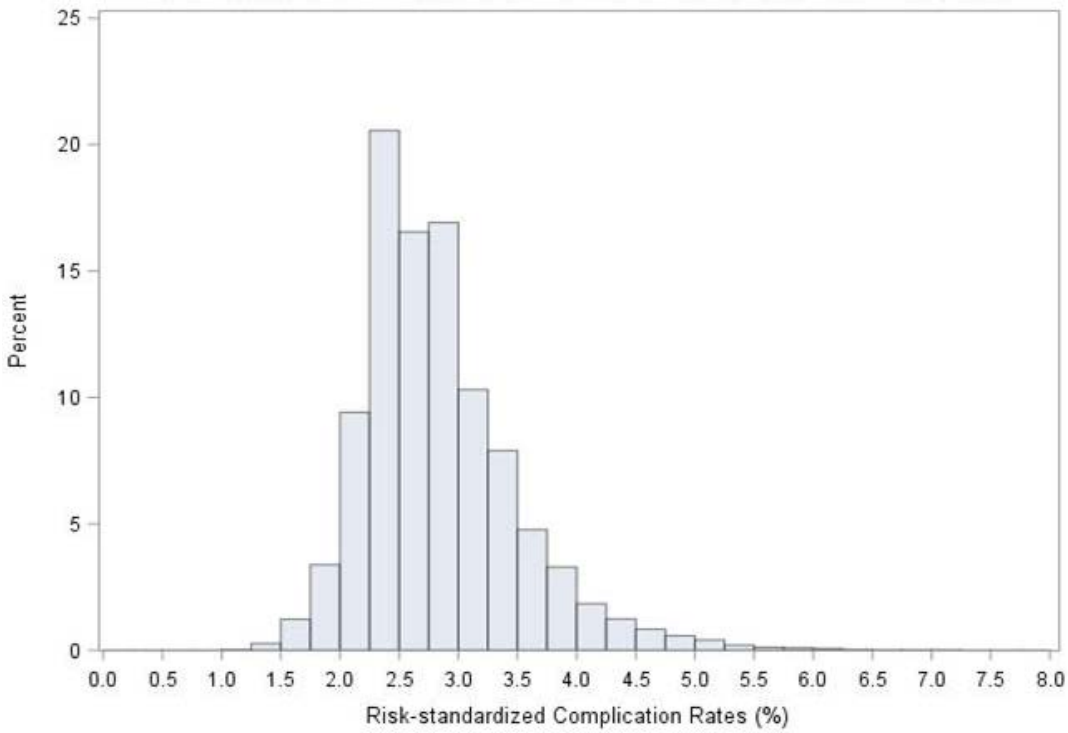
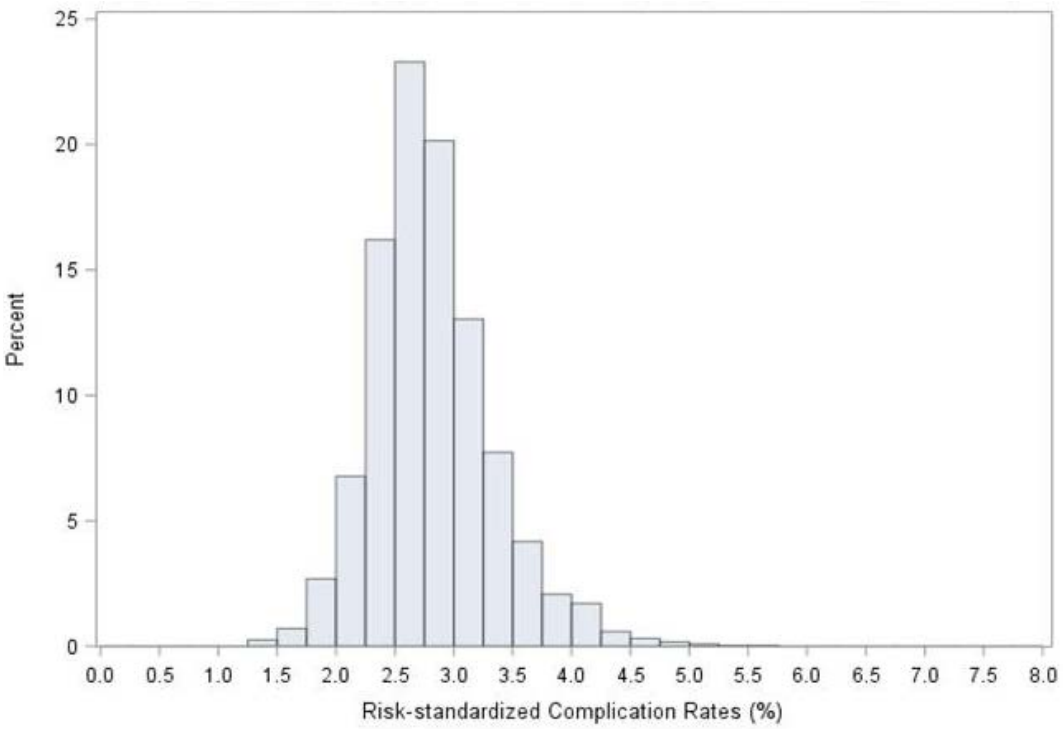


Figure 4 shows the range of RSCR distribution for MIPS ECs with more than 25 cases. Similar to the rates for ECs, the median RSCR for clinicians is relatively low. EC groups, however, have a narrower range of distribution (range = 1.4% – 5.7%) as compared to ECs. Several EC groups have higher-than-expected complication rates; this indicates opportunity for improvement.

Figure 4. Distribution of risk-standardized complication rates (dataset: Medicare full sample [April 2013 – March 2016]) - hierarchical logistic regression model – MIPS EC groups with more than 25 cases



4.6 Distributions of ECs and EC Groups by Performance Category in the 3-Year Dataset

Of 7,928 ECs the 3-year study cohort (Medicare Full Sample: April 1, 2013 – March 31, 2016):

- 51 performed "Better than the National Rate."
- 7,809 performed "No Different than the National Rate."
- 68 performed "Worse than the National Rate."
- 8,827 were classified as "Number of Cases Too Small" (fewer than 25) to reliably conclude how the EC is performing.

Of 3,572 EC groups the 3-year study cohort (Medicare Full Sample: April 2013 – March 2016):

- 67 performed "Better than the National Rate."
- 3,461 performed "No Different than the National Rate."
- 44 performed "Worse than the National Rate."
- 2,546 were classified as "Number of Cases Too Small" (fewer than 25) to reliably conclude how the EC is performing.

4.7 Reliability

4.7.1 Data Element Reliability

Because these measures are calculated from claims submitted by hospitals and clinicians, adjudicated by CMS, and stored electronically, the reliability of the data is extremely high. When the measures are computed on the same set of admissions, for the same providers, using the same time period, precisely the same results are obtained.

Table 7 and Table 8 show hierarchical logistic regression model variable coefficients for MIPS ECs and EC groups, respectively. The model estimates and adjusted odds ratios were similar for variables in the Development Sample, Validation Sample, and Temporal Validation Sample for ECs and EC groups.

The calibration statistics results across the Development Sample, Validation Sample, and Temporal Validation Sample in Table 9 show the values in close proximity to 0-1, indicating good model reliability. The calculated c-statistic (ROC curve) of 0.65 tested using the three samples indicate acceptable discrimination across the cohort models, consistent with the hospital-level measure. Additionally, the risk decile plots showed good calibration; the model performed well in each of the risk deciles in both the Development Sample (Figure 5) and the Validation Sample (Figure 6).

Table 7. MIPS eligible clinicians: hierarchical regression model variable coefficients

Variable	**Development sample** (April 2013 – March 2015)	**Validation sample** (April 2013 – March 2015)	**Temporal validation sample** (April 2015 – March 2017)
Mean age minus 65 (SD)	0.03 (0)	0.03 (0)	0.03 (0)
Male	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)
Index admissions with an elective THA procedure	0.38 (0.02)	0.33 (0.02)	0.25 (0.02)
Number of procedures (2 vs. 1)	0.55 (0.07)	0.53 (0.07)	0.55 (0.05)
Metastatic cancer or acute leukemia (CC 8)	0.17 (0.13)	0.12 (0.14)	-0.09 (0.1)
Other major cancers (CC 9-12)	0.01 (0.03)	-0.04 (0.03)	-0.01 (0.02)
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	-0.04 (0.03)	-0.06 (0.03)	-0.05 (0.02)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	0.11 (0.02)	0.09 (0.02)	0.12 (0.02)
Protein-calorie malnutrition (CC 21)	0.79 (0.08)	0.92 (0.08)	0.9 (0.06)
Bone/joint/muscle infections/necrosis (CC 39)	0.11 (0.06)	0.14 (0.06)	0.13 (0.04)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	0.19 (0.03)	0.07 (0.04)	0.14 (0.03)
Osteoarthritis of hip or knee (CC 42)	0.01 (0.06)	-0.09 (0.06)	-0.1 (0.04)
Osteoporosis and other bone/cartilage disorders (CC 43)	-0.03 (0.03)	0.03 (0.03)	-0.01 (0.02)
Dementia or other specified brain disorders (CC 51-53)	0.19 (0.05)	0.22 (0.05)	0.19 (0.03)
Major psychiatric disorders (CC 57-59)	0.34 (0.04)	0.22 (0.05)	0.34 (0.03)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	0.21 (0.09)	0.29 (0.09)	0.11 (0.06)

Variable	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Cardio-respiratory failure and shock (CC 84), plus ICD-10-CM codes R09.01 and R09.02	0.18 (0.06)	0.14 (0.06)	0.16 (0.04)
Coronary atherosclerosis or angina (CC 88-89)	0.24 (0.02)	0.29 (0.02)	0.26 (0.02)
Stroke (CC 99-100)	0 (0.07)	-0.06 (0.07)	0.09 (0.05)
Vascular or circulatory disease (CC 106-109)	0.15 (0.03)	0.09 (0.03)	0.12 (0.02)
Chronic obstructive pulmonary disease (COPD) (CC 111)	0.44 (0.03)	0.4 (0.03)	0.47 (0.02)
Pneumonia (CC 114-116)	0.19 (0.05)	0.27 (0.05)	0.18 (0.03)
Pleural effusion/pneumothorax (CC 117)	-0.1 (0.07)	-0.07 (0.07)	-0.04 (0.05)
Dialysis status (CC 134)	0.36 (0.17)	0.27 (0.16)	0.1 (0.12)
Renal failure (CC 135-140)	0.28 (0.03)	0.34 (0.03)	0.28 (0.02)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	0.3 (0.05)	0.27 (0.06)	0.28 (0.04)
Trauma (CC 166-168, 170-173)	0.17 (0.05)	0.22 (0.04)	0.15 (0.03)
Vertebral fractures without spinal cord injury (CC 169)	0.15 (0.08)	0.07 (0.09)	0.05 (0.07)
Other injuries (CC 174)	0.05 (0.03)	0.04 (0.03)	0.1 (0.02)
Major complications of medical care and trauma (CC 176-177)	0.15 (0.04)	0.15 (0.04)	0.21 (0.03)
Morbid obesity (CC 22)	0.43 (0.04)	0.48 (0.04)	0.51 (0.02)
Other congenital deformity of hip (joint) (ICD-9 code 755.63, ICD-10 code Q65.89, Q65.9)	-0.2 (0.25)	0.21 (0.21)	0.53 (0.13)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16, ICD-10 code M12.551, M12.552, M12.559)	0.13 (0.16)	0.27 (0.15)	0.38 (0.15)

Table 8. MIPS eligible clinician groups: hierarchical regression model variable coefficients for THA/TKA

Variable	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Mean age minus 65 (SD)	0.03 (0)	0.03 (0)	0.03 (0)
Male	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)
Index admissions with an elective THA procedure	0.38 (0.02)	0.33 (0.02)	0.25 (0.02)
Number of procedures (2 vs. 1)	0.56 (0.07)	0.53 (0.07)	0.55 (0.05)

Variable	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Metastatic cancer or acute leukemia (CC 8)	0.17 (0.13)	0.13 (0.14)	-0.09 (0.1)
Other major cancers (CC 9-12)	0.01 (0.03)	-0.04 (0.03)	-0.01 (0.02)
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	-0.04 (0.03)	-0.06 (0.03)	-0.05 (0.02)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	0.11 (0.02)	0.09 (0.02)	0.12 (0.02)
Protein-calorie malnutrition (CC 21)	0.79 (0.08)	0.92 (0.07)	0.9 (0.06)
Bone/joint/muscle infections/necrosis (CC 39)	0.11 (0.06)	0.14 (0.06)	0.14 (0.04)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	0.19 (0.03)	0.07 (0.04)	0.15 (0.03)
Osteoarthritis of hip or knee (CC 42)	0.01 (0.06)	-0.09 (0.06)	-0.12 (0.04)
Osteoporosis and other bone/cartilage disorders (CC 43)	-0.03 (0.03)	0.03 (0.03)	-0.01 (0.02)
Dementia or other specified brain disorders (CC 51-53)	0.19 (0.05)	0.22 (0.05)	0.19 (0.03)
Major psychiatric disorders (CC 57-59)	0.34 (0.04)	0.22 (0.05)	0.34 (0.03)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	0.21 (0.09)	0.29 (0.08)	0.11 (0.06)
Cardio-respiratory failure and shock (CC 84), plus ICD-10-CM codes R09.01 and R09.02	0.18 (0.06)	0.14 (0.06)	0.16 (0.04)
Coronary atherosclerosis or angina (CC 88-89)	0.24 (0.02)	0.28 (0.02)	0.26 (0.02)
Stroke (CC 99-100)	0.01 (0.07)	-0.06 (0.07)	0.09 (0.05)
Vascular or circulatory disease (CC 106-109)	0.14 (0.03)	0.09 (0.03)	0.12 (0.02)
Chronic obstructive pulmonary disease (COPD) (CC 111)	0.44 (0.03)	0.4 (0.03)	0.47 (0.02)
Pneumonia (CC 114-116)	0.19 (0.05)	0.27 (0.05)	0.18 (0.03)
Pleural effusion/pneumothorax (CC 117)	-0.1 (0.07)	-0.07 (0.07)	-0.03 (0.05)
Dialysis status (CC 134)	0.36 (0.17)	0.28 (0.16)	0.09 (0.12)
Renal failure (CC 135-140)	0.28 (0.03)	0.34 (0.03)	0.28 (0.02)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	0.29 (0.05)	0.26 (0.05)	0.28 (0.04)
Trauma (CC 166-168, 170-173)	0.17 (0.05)	0.22 (0.04)	0.15 (0.03)
Vertebral fractures without spinal cord injury (CC 169)	0.15 (0.08)	0.06 (0.08)	0.05 (0.07)
Other injuries (CC 174)	0.05 (0.03)	0.04 (0.03)	0.1 (0.02)

Variable	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Major complications of medical care and trauma (CC 176-177)	0.16 (0.04)	0.15 (0.04)	0.22 (0.03)
Morbid obesity (CC 22)	0.43 (0.04)	0.48 (0.04)	0.51 (0.02)
Other congenital deformity of hip (joint) (ICD-9 code 755.63, ICD-10 code Q65.89, Q65.9)	-0.19 (0.25)	0.19 (0.21)	0.53 (0.13)
Post traumatic osteoarthritis (ICD-9 codes 716.15, 716.16, IC D-10 code M12.551, M12.552, M12.559)	0.14 (0.16)	0.28 (0.15)	0.37 (0.14)

Table 9. Risk-adjustment model performance summaries (dataset: Medicare full sample [April 2013 – March 2016])

Characteristic	Development sample (April 2013 – March 2015)	Validation sample (April 2013 – March 2015)	Temporal validation sample (April 2015 – March 2017)
Number of admissions	302,561	302,519	653,598
Unadjusted complication rate calibration (r0, r1)	(0, 1)	(-0.02, 0.99)	(-0.12, 1.01)
Discrimination - predictive ability (lowest decile %, highest decile %)	1.36 – 6.77	1.41 – 6.84	1.19 – 6.01
c-statistic	0.65	0.65	0.65
Residuals lack of fit (Pearson residual %)			
< -2	0.00%	0.00%	0.00%
[-2, 0)	97.16%	97.12%	97.55%
[0, 2)	0.01%	0.01%	0.01%
[2 +)	2.83%	2.86%	2.44%
Model Wald x2 [number of covariates] (p-value)	2,811.5 [33] (<.0001)	2,855.6 [33] (<.0001)	5,684.6 [33] (<.0001)

Figure 5. Calibration plot (dataset: Development Sample)

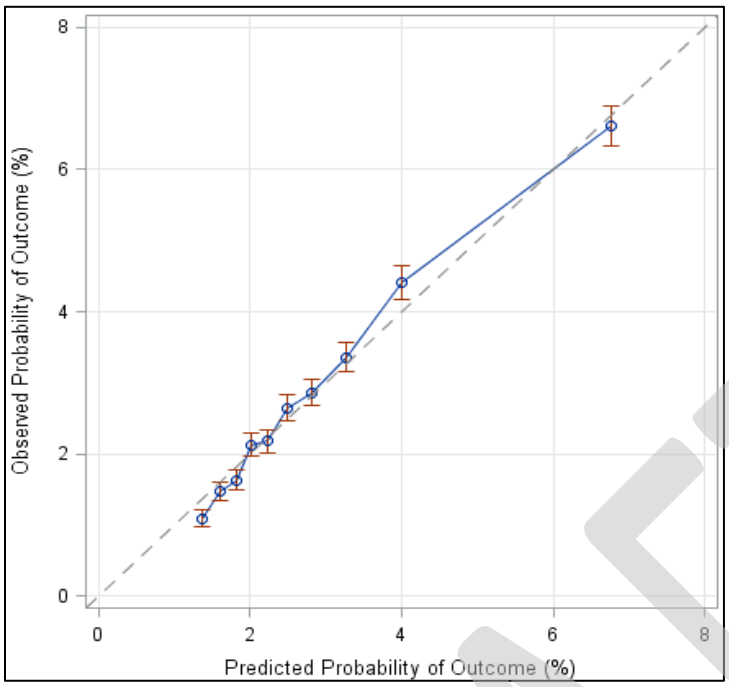
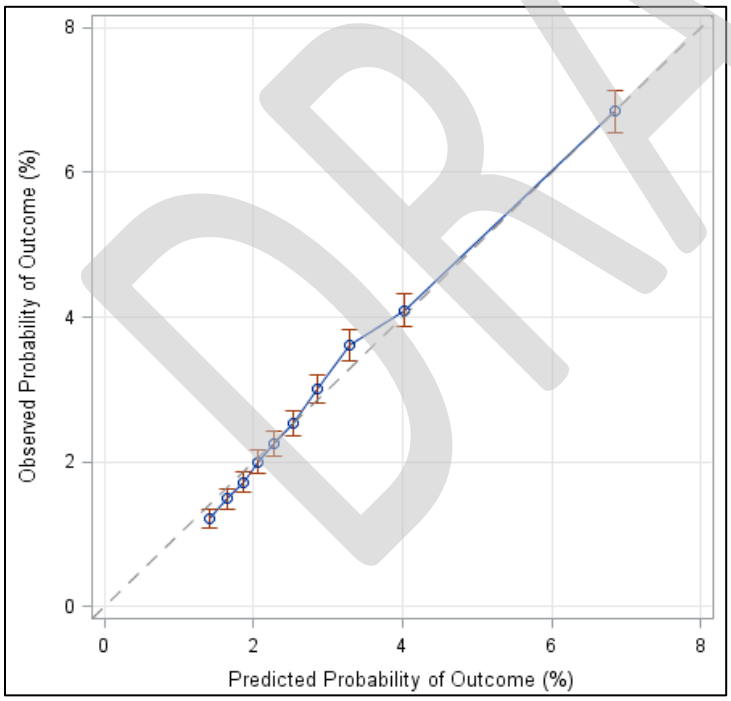


Figure 6. Calibration plot (dataset: Validation Sample)



4.7.2 Measure Score Reliability

The test-retest analysis used data from April 2013 to March 2015 and data from April 2015 to March 2017. The random split test-retest reliability using two randomly split data from

four years (April 2013 to March 2017) generated similar results/frequency of providers in comparison between the datasets with a good degree of overlap at both the EC (Table 10) and EC group-levels (Table 11). The intraclass correlation efficient (ICC) of 0.35 for ECs, calculated with two years of data, is considered "fair".[43] The ICC of 0.47 for EC groups, also calculated with two years of data, is considered "moderate".[43] This split-sample reliability represents the lower bound of estimate of the true orthopedic measure reliability.

The signal-to-noise reliability used data from the Medicare Full Sample. Using the approach used by Adams et. al.,[44] we obtained median reliability scores of 0.793 for ECs and 0.790 for EC groups.

Taken together, these results indicate acceptable reliability of the measure score for both ECs and EC groups.

Table 10. Score reliability for MIPS eligible clinicians and eligible clinician groups with more than 25 admissions over 4 years - hierarchical logistic regression model (datasets: Reliability Sample 1 versus Reliability Sample 2)

Characteristic	Eligible clinicians	Eligible clinician groups
Number of providers in Reliability Sample 1	9,232	3,920
Number of providers in Reliability Sample 2	9,232	3,920
Number of providers overlapping in Reliability Sample 1 and Reliability Sample 2	9,232	3,920
Unadjusted ICC (2,1)	0.22	0.31
Adjusted ICC (2,1) (*adjusted using Spearman Brown formula)*	0.35	0.47

Table 11. Signal-to-noise reliability results for MIPS eligible clinicians and clinician groups (ECs and EC groups) with more than 25 admissions - hierarchical logistic regression model (dataset: Medicare Full Sample [April 2013 – March 2016])

Reliability score	Eligible clinicians	Eligible clinician groups
Mean (SD)	0.786 (0.109)	0.769 (0.153)
Median (IQR)	0.793 (0.695 - 0.878)	0.790 (0.647 - 0.907)
Range (min. – max.)	0.582 - 0.988	0.463 - 0.996

4.8 Validity

4.8.1 Face Validity

Face Validity of Final Attribution Rule

The TEP strongly supported attribution to the Billing Surgeon.

Face Validity of MIPS Eligible Clinician and Eligible Clinician Group Measure Scores

Of 19 TEP members asked to complete a survey regarding validity and usability of the MIPS HKC measure, 16 responded; their responses are reported in Table 12.

Table 12. TEP reports of agreements

…measure scores are valid and useful	1	0	2	3	9	1
…measure will provide info to be used for quality improvement	1	1	2	4	5	3

As shown in Table 12, the majority of the respondents, 13/16 or 81%, agreed that the HKC measure scores were valid and useful, and 12/16 or 75% agreed that the measure would provide information that could be used to improve the quality of care.

Among those who disagreed, the primary concern was that lowest volume eligible clinicians would not be measured, rather than concern with the measure itself. Though this is a challenge with all quality measures, it may be of particular concern when there may be an inverse relationship between volume and quality. It is notable that even with the 25 volume threshold, over 96% of patients are retained; it is also important to note that the measure counts only Medicare Fee-For-Service patients, so the total case volume of those eligible clinicians excluded by the volume threshold is unknown, and could be quite high.

Overall, the survey indicates support of the validity and usability of the measure.

5. SUMMARY

Medicare beneficiaries commonly undergo inpatient THA and TKA procedures. Based on our empiric analyses of Medicare FFS patients aged 65 years and older, from April 2013 through March 2016, 924,786 THA/TKA procedures were performed by 7,928 eligible clinicians or 3,572 EC groups. The median (IQR) RSCR for ECs was 2.7% (2.4% - 3.2%) and for EC groups it was 2.8% (2.5% - 3.1%). The measure shows acceptable variation across providers and measure reliability, especially for EC groups, using a minimum case volume threshold of 25 cases. 81% of the Technical Expert Panel agreed the measure scores were valid and useful for measuring quality of care. The MIPS THA/TKA complication measure, as specified, has the potential to illuminate these differences in quality, inform patient choice, drive quality improvement, and enhance care coordination. We look forward to your input on any and all aspects of the measure specifications during public comment.

6. GLOSSARY

Attribution: Assignment of the outcome of a patient episode of care to one or more healthcare provider entities for the purpose of assessing performance quality.

Bootstrapping: The bootstrap is a computer-based method for estimating the standard error of an estimate when the estimate is based on a sample with an unknown probability distribution. Bootstrap methods depend on the bootstrap sample, which is a random sample of size *n* drawn with replacement from the population of *n* objects. The bootstrap algorithm works by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications, and estimating the standard error of the statistic by the empirical standard deviation of the replications.

C-statistic: An indicator of the model's discriminant ability or ability to correctly classify those who have and have not had a complication following a THA/TKA procedure. Potential values range from 0.5, meaning no better than chance, to 1.0, an indication of perfect prediction. Perfect prediction implies that patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play no role in their patients' outcomes.

Case mix: The particular illness severity, age, and, for some measures, gender characteristics of patients with index admissions at a given hospital.

Cohort: The index admissions included in the measure after inclusion and exclusion criteria have been applied.

Comorbidities: Medical conditions that the patient had in addition to his/her primary reason for admission to the hospital.

Complications: Medical conditions that may have occurred as a consequence of care rendered during hospitalization.

Condition Categories (CCs): Groupings of ICD-9-CM/ICD-10-CM diagnosis codes in clinically relevant categories, from the HCCs system.[45,46] CMS uses the grouping, but not the hierarchical logic of the system, to create risk factor variables. Mappings which show the assignment of ICD-9 and ICD-10-CM codes to the CCs are available on the ***QualityNet*** website.

Confidence interval (CI): A CI is a range of values that describes the uncertainty surrounding an estimate. It is indicated by its endpoints; for example, a 95% CI for the OR associated with protein-calorie malnutrition noted as "1.09 – 1.15" would indicate that there is 95% confidence that the OR lies between 1.09 and 1.15.

Eligible clinician (EC): An individual MIPS EC, is identified through their unique National Provider Identifier (NPI) and Taxpayer Identification Number (TIN) combination, listed on patient claims. Most NPIs are associated with only one TIN.

Eligible clinician group (EC group): An EC group is the aggregate of clinicians within a TIN.

EC-specific or EC group-specific effect: A measure of the EC or EC group quality of care that is calculated through hierarchical logistic regression, taking into consideration how many patients were eligible for the cohort, these patients' risk factors, and how many had THA/TKA complications. The EC- or EC group-specific effect is the calculated random effect intercept for each EC or EC group. The EC- or EC group-

specific intercept will be negative for a better-than-average EC or EC group, positive for a worse-than-average EC or EC group, and close to zero for an average EC or EC group. The EC or EC group-specific intercept is used in the numerator to calculate "predicted" complications.

Expected number of admissions with a complication: The number of admissions with a complication expected based on average EC or EC group's performance with a given EC or EC group's case mix.

Hierarchical model: A widely accepted statistical method that enables evaluation of relative hospital performance by accounting for patient risk factors and the number of patients that a hospital treats. This statistical model accounts for the hierarchical structure of the data (patients clustered within hospitals are assumed to be correlated) and accommodates modeling of the association between outcomes and patient characteristics. Based on the hierarchical model, we can evaluate (1) how much variation in hospital complication rates overall is accounted for by patients' individual risk factors (such as age and other medical conditions), and (2) how much variation is accounted for by hospital contribution to complication risk.

Index admission: Any admission included in the measure calculation as the initial admission for a qualifying elective THA/TKA procedure and evaluated for the outcome.

Interval estimate: Similar to a CI. The interval estimate is a range of probable values for the estimate that characterizes the amount of uncertainty. For example, a 95% interval estimate for a complication rate indicates there is 95% confidence that the true value of the rate lies between the lower and the upper limit of the interval.

Medicare fee-for-service (FFS): Original Medicare plan in which providers receive a fee or payment for each individual service provided directly from Medicare. Only beneficiaries in Medicare FFS, not in managed care (Medicare Advantage), are included in the measure.

National observed complication rate: All included hospitalizations with the outcome divided by all included hospitalizations.

Odds ratio (OR): The ORs express the relative odds of the outcome for each of the predictor variables. For example, the OR for Protein-calorie malnutrition (CC 21) represents the odds of the outcome for patients with that risk variable present relative to those without the risk variable present. The model coefficient for each risk variable is the log (odds) for that variable.

Outcome: The result of a broad set of healthcare activities that affect patients' well-being. For the complication measure, the outcome is any one of the specified complications occurring during the index admission or during a readmission, except for death, which can occur anywhere as long as it is within 30 days of the start of the index admission.

Predicted number of admissions with a complication: The number of admissions with a complication predicted based on the EC's or EC group's performance with its observed case mix.

Predictive ability: An indicator of the model's discriminant ability or ability to distinguish high-risk subjects from low-risk subjects. A wide range between the lowest decile and highest decile suggests better discrimination.

Risk-adjustment variables: Patient demographics and comorbidities used to standardize rates for differences in case mix across ECs or EC groups.

7. REFERENCES

1. Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation. Procedure-Specific Measure Updates and Specifications Report Hospital-Level Risk-Standardized Complication Measure Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty. 2017; https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228774719413. Accessed January 10, 2018.
2. Nwachukwu BU, McCormick F, Provencher MT, Roche M, Rubash HE. A comprehensive analysis of Medicare trends in utilization and hospital economics for total knee and hip arthroplasty from 2005 to 2011. *J Arthroplasty.* 2015;30(1):15-18.
3. Cram P, Lu X, Kates SL, Singh JA, Li Y, Wolf BR. Total knee arthroplasty volume, utilization, and outcomes among Medicare beneficiaries, 1991-2010. *JAMA.* 2012;308(12):1227-1236.
4. Courtney PM, Edmiston T, Batko B, Levine BR. Can Bundled Payments Be Successful in the Medicaid Population for Primary Joint Arthroplasty? *J Arthroplasty.* 2017;32(11):3263-3267.
5. Kurtz SM, Ong KL, Schmier J, et al. Future clinical and economic impact of revision total hip and knee arthroplasty. *J Bone Joint Surg Am.* 2007;89 Suppl 3:144-151.
6. Ong KL, Mowat FS, Chan N, Lau E, Halpern MT, Kurtz SM. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. *Clin Orthop Relat Res.* 2006;446:22-28.
7. Bozic KJ, Rubash HE, Sculco TP, Berry DJ. An analysis of medicare payment policy for total joint arthroplasty. *J Arthroplasty.* 2008;23(6 Suppl 1):133-138.
8. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am.* 2007;89(4):780-785.
9. Bongartz T, Halligan CS, Osmon DR, et al. Incidence and risk factors of prosthetic joint infection after total hip or knee replacement in patients with rheumatoid arthritis. *Arthritis Rheum.* 2008;59(12):1713-1720.
10. Kurtz S, Ong K, Lau E, Bozic K, Berry D, Parvizi J. Prosthetic joint infection risk after TKA in the Medicare population. *Clin Orthop Relat Res.* 2010;468:5.
11. Soohoo NF, Farng E, Lieberman JR, Chambers L, Zingmond DS. Factors that predict short-term complication rates after total hip arthroplasty. *Clin Orthop Relat Res.* 2010;468(9):2363-2371.
12. Cram P, Vaughan-Sarrazin MS, Wolf B, Katz JN, Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. *J Bone Joint Surg Am.* 2007;89(8):1675-1684.
13. Mahomed NN, Barrett JA, Katz JN, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. *J Bone Joint Surg Am.* 2003;85-A(1):27-32.
14. Khatod M, Inacio M, Paxton EW, et al. Knee replacement: epidemiology, outcomes, and trends in Southern California: 17,080 replacements from 1995 through 2004. *Acta Orthop.* 2008;79(6):812-819.
15. Solomon DH, Chibnik LB, Losina E, et al. Development of a preliminary index that predicts adverse events after total knee replacement. *Arthritis & Rheumatism.* 2006;54(5):1536-1542.
16. Browne J, Cook C, Hofmann A, Bolognesi M. Postoperative morbidity and mortality following total knee arthroplasty with computer navigation. *Knee.* 2010;17(2):152-156.
17. Huddleston JI, Maloney WJ, Wang Y, Verzier N, Hunt DR, Herndon JH. Adverse Events After Total Knee Arthroplasty: A National Medicare Study. *The Journal of Arthroplasty.* 2009;24(6, Supplement 1):95-100.

18. Centers for Medicare and Medicaid Services. Overview: THA/TKA Complication Measure. https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228772780555. Accessed January 10, 2018.
19. Centers for Medicare and Medicaid Services. Hospital Value-Based Purchasing: Fiscal Years 2018-2023 Measures. https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228775522697. Accessed January 10, 2018.
20. Centers for Medicare and Medicaid Services, Health Services Advisory Group. CMS Quality Measure Development Plan Environmental Scan and Gap Analysis Report (MACRA, Section 102). 2017; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MDP_EScan_GapAnalysis_Report.pdf. Accessed January 10, 2018.
21. Centers for Medicare and Medicaid Services, Health Services Advisory Group. CMS Quality Measure Development Plan: Supporting the Transition to The Quality Payment Program 2017 Annual Report. 2017; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2017-CMS-MDP-Annual-Report.pdf. Accessed January 10, 2018.
22. Centers for Medicare and Medicaid Services. Core Measures Overview. 2017; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures.html. Accessed January 10, 2018.
23. Medicare Program; Medicare Incentive Payment System/Alternative Payment Model (MIPS/APM) Final Rule. Vol 81 Fed. Reg. 77036 (to be codified at 42 C.F.R. Parts 414 and 495)2016.
24. Medicare Program; Calendar Year 2018 Updates to the Quality Payment Program Proposed Rule. Vol 82 Fed. Reg. 30028 (to be codified at 42 C.F.R. Part 414)2017.
25. Medicare Program; Medicare Incentive Payment System/Alternative Payment Model (MIPS/APM) Final Rule. Vol 81 Fed. Reg. 77101 (to be codified at 42 C.F.R. Parts 414 and 495)2016.
26. Center for Medicare and Medicaid Services. CY 2018 Updates to the Quality Payment Program Proposed Rule, 82 Fed. Reg. 30097 (to be codified at 42 C.F.R. Parts 414). June 30, 2017.
27. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician Focused Payment Models Final Rule with Comment Period. Vol 81 Fed. Reg. 77282 through 77284 (to be codified at 42 C.F.R. Parts 414 and 495)2016.
28. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician Focused Payment Models Final Rule with Comment Period. Vol 81 Fed. Reg. 77131 (to be codified at 42 C.F.R. Parts 414 and 495)2016.
29. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician Focused Payment Models Final Rule with Comment Period. Vol 81 Fed. Reg. 77127 (to be codified at 42 C.F.R. Parts 414 and 495)2016.
30. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models Final Rule with Comment Period. Vol 82 Fed. Reg. 53753 (to be codified at 42 C.F.R. Part 414)2017.

31. Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY 2019; Medicare Shared Savings Program Requirements; Quality Payment Program; and Medicaid Promoting Interoperability Program Proposed Rule. Vol 83 Fed. Reg. 35960 (to be codified at 42 C.F.R. Parts 405, 410, 411, 414, 415, and 495) 2018.
32. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models Final Rule with Comment Period. Vol 82 Fed. Reg. 53698 (to be codified at 42 C.F.R. Part 414)2017.
33. Centers for Medicaid and Medicare Services. Blueprint for the CMS Measures Management System Version 13.0. 2017; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Blueprint-130.pdf. Accessed June 5, 2018.
34. National Quality Forum. National Voluntary Consensus Standards for Patient Outcomes 2009: A Consensus Report. 2011; http://www.qualityforum.org/Publications/2011/07/National_Voluntary_Consensus_Standards_for_Patient_Outcomes_2009.aspx. Accessed August 15, 2018.
35. Krumholz HM, Brindis RG, Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* 2006;113(3):456-462.
36. National Quality Forum. Attribution: Principles and Approaches. 2016; https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=84236. Accessed August 10, 2018.
37. Pope G, Ellis R, Ash A, et al. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. *Health Care Financing Review.* 2000;21(3):26.
38. Centers for Medicaid and Medicare Services. March 31, 2016, HHS-Operated Risk Adjustment Methodology Meeting. 2016; https://www.cms.gov/CCIIO/Resources/Forms-Reports-and-Other-Resources/Downloads/RA-March-31-White-Paper-032416.pdf. Accessed August 15, 2018.
39. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med.* 2002;21(22):3431-3446.
40. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420-428.
41. Spearman C. Correlation Calculated from Faulty Data. *British Journal of Psychology.* 1910;3(3):271-295.
42. Brown W. Some Experimental Results in the Correlation of Mental Abilities. *British Journal of Psychology.* 1910;3:296-322.
43. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
44. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling--reliability and risk of misclassification. *N Engl J Med.* 2010;362(11):1014-1021.
45. Pope GC, Ellis RP, Ash AS, et al. Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment Final Report. 2000; https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/downloads/pope_2000_2.pdf. Accessed February 9, 2018.
46. Pope GC, Kautter J, Ingber MJ, Freeman S, Sekar R, Newhart C. Evaluation of the CMS-HCC Risk Adjustment Model. 2011; https://www.cms.gov/Medicare/Health-

Plans/MedicareAdvtgSpecRateStats/downloads/evaluation_risk_adj_model_2011.pdf. Accessed February 9, 2018.
47. Centers for Medicaid and Medicare Services. Medicare FFS Physician Feedback Program/Value-Based Payment Modifier. 2016; https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/index.html. Accessed August 5, 2018.
48. Centers for Medicaid and Medicare Innovation. Comprehensive Primary Care Plus. CPC+ Payment Methodologies: Beneficiary Attribution, Care Managment Fee, Performance-Based Incentive Payment, an Payment Under the Medicare Physician Fee Schedule. 2017; https://www.pcpcc.org/sites/default/files/resources/cpcplus-methodology.pdf. Accessed May 9, 2018.
49. Centers for Medicaid and Medicare Services. Multi-Payer Advanced Primary Care Practice. 2017; https://innovation.cms.gov/initiatives/Multi-Payer-Advanced-Primary-Care-Practice/. Accessed August 10, 2018.
50. Daniels MJ, Gatsonis C. Hierarchical generalized linear models in the analysis of variations in health care utilization. *J Am Stat Assoc.* 1999;94(445):29-42.
51. Normand SL, Wang Y, Krumholz HM. Assessing surrogacy of data sources for institutional comparisons. *Health Services and Outcomes Research Methodology.* 2007;7(1-2):79-96.

8. APPENDICES

Appendix A. Code Lists for Cohort, Outcome, and Risk Adjustment

A.1 Cohort

Table A1 below outlines the ICD-10-PCS codes used to identify THA/TKA procedures in claims for discharges on or after October 1, 2015. ICD-9 code lists for discharges prior to October 1, 2015 can be found in the 2016 hospital procedure-specific complication measure updates and specifications report posted on *QualityNet*.

Table A1. ICD-10-PCS codes used to identify eligible THA/TKA procedures

ICD-10-PCS Codes	Description
0SR9019	Replacement of Right Hip Joint with Metal Synthetic Substitute, Cemented, Open Approach
0SR901A	Replacement of Right Hip Joint with Metal Synthetic Substitute, Uncemented, Open Approach
0SR901Z	Replacement of Right Hip Joint with Metal Synthetic Substitute, Open Approach
0SR9029	Replacement of Right Hip Joint with Metal on Polyethylene Synthetic Substitute, Cemented, Open Approach
0SR902A	Replacement of Right Hip Joint with Metal on Polyethylene Synthetic Substitute, Uncemented, Open Approach
0SR902Z	Replacement of Right Hip Joint with Metal on Polyethylene Synthetic Substitute, Open Approach
0SR9039	Replacement of Right Hip Joint with Ceramic Synthetic Substitute, Cemented, Open Approach
0SR903A	Replacement of Right Hip Joint with Ceramic Synthetic Substitute, Uncemented, Open Approach
0SR903Z	Replacement of Right Hip Joint with Ceramic Synthetic Substitute, Open Approach
0SR9049	Replacement of Right Hip Joint with Ceramic on Polyethylene Synthetic Substitute, Cemented, Open Approach
0SR904A	Replacement of Right Hip Joint with Ceramic on Polyethylene Synthetic Substitute, Uncemented, Open Approach
0SR904Z	Replacement of Right Hip Joint with Ceramic on Polyethylene Synthetic Substitute, Open Approach
0SR90J9	Replacement of Right Hip Joint with Synthetic Substitute, Cemented, Open Approach
0SR90JA	Replacement of Right Hip Joint with Synthetic Substitute, Uncemented, Open Approach
0SR90JZ	Replacement of Right Hip Joint with Synthetic Substitute, Open Approach
0SRB019	Replacement of Left Hip Joint with Metal Synthetic Substitute, Cemented, Open Approach
0SRB01A	Replacement of Left Hip Joint with Metal Synthetic Substitute, Uncemented, Open Approach

ICD-10-PCS Codes	Description
0SRB01Z	Replacement of Left Hip Joint with Metal Synthetic Substitute, Open Approach
0SRB029	Replacement of Left Hip Joint with Metal on Polyethylene Synthetic Substitute, Cemented, Open Approach
0SRB02A	Replacement of Left Hip Joint with Metal on Polyethylene Synthetic Substitute, Uncemented, Open Approach
0SRB02Z	Replacement of Left Hip Joint with Metal on Polyethylene Synthetic Substitute, Open Approach
0SRB039	Replacement of Left Hip Joint with Ceramic Synthetic Substitute, Cemented, Open Approach
0SRB03A	Replacement of Left Hip Joint with Ceramic Synthetic Substitute, Uncemented, Open Approach
0SRB03Z	Replacement of Left Hip Joint with Ceramic Synthetic Substitute, Open Approach
0SRB049	Replacement of Left Hip Joint with Ceramic on Polyethylene Synthetic Substitute, Cemented, Open Approach
0SRB04A	Replacement of Left Hip Joint with Ceramic on Polyethylene Synthetic Substitute, Uncemented, Open Approach
0SRB04Z	Replacement of Left Hip Joint with Ceramic on Polyethylene Synthetic Substitute, Open Approach
0SRB0J9	Replacement of Left Hip Joint with Synthetic Substitute, Cemented, Open Approach
0SRB0JA	Replacement of Left Hip Joint with Synthetic Substitute, Uncemented, Open Approach
0SRB0JZ	Replacement of Left Hip Joint with Synthetic Substitute, Open Approach
0SRC0J9	Replacement of Right Knee Joint with Synthetic Substitute, Cemented, Open Approach
0SRC0JA	Replacement of Right Knee Joint with Synthetic Substitute, Uncemented, Open Approach
0SRC0JZ	Replacement of Right Knee Joint with Synthetic Substitute, Open Approach
0SRD0J9	Replacement of Left Knee Joint with Synthetic Substitute, Cemented, Open Approach
0SRD0JA	Replacement of Left Knee Joint with Synthetic Substitute, Uncemented, Open Approach
0SRD0JZ	Replacement of Left Knee Joint with Synthetic Substitute, Open Approach

A.2 Outcome

Table A2 provides hyperlinks to the ICD-10 code lists used to define the complications captured in the measure outcome. The ICD-10 codes used to define the complications in claims for discharges on or after October 1, 2015 are posted on *QualityNet* due to volume. ICD-9 code lists for discharges prior to October 1, 2015 can be found in the 2016 procedure-specific complication measure updates and specifications report posted on *QualityNet*.

Table A2. Identification of complications following THA/TKA

Complication	Follow-up period in days	ICD-10 codes defining complication	Required coding placement
Acute myocardial infarction	During index admission or a subsequent inpatient admission that occurs within 7 days of the start of the index admission	ICD-10-CM code list	• Index admission – secondary discharge diagnosis fields only AND not coded as POA • Readmissions – principal discharge diagnosis field only
Pneumonia	During index admission or a subsequent inpatient admission that occurs within 7 days of the start of the index admission	ICD-10-CM code list	• Index admission – secondary discharge diagnosis fields only AND not coded as POA • Readmissions – principal discharge diagnosis field only
Sepsis/ septicemia/ shock	During index admission or a subsequent inpatient admission that occurs within 7 days of the start of the index admission	ICD-10-CM code list	• Index admission – secondary discharge diagnosis fields only AND not coded as POA • Readmissions – principal or secondary discharge diagnosis fields

Complication	Follow-up period in days	ICD-10 codes defining complication	Required coding placement
Surgical site bleeding	During index admission or a subsequent inpatient admission that occurs within 30 days of the start of the index admission	One of the diagnosis codes in ICD-10-CM code list AND One of the procedure codes in ICD-10-PCS code list	• Index admission o Diagnosis code in secondary discharge diagnosis fields only AND not coded as POA o Procedure code in secondary procedure fields only • Readmissions o Diagnosis code in principal or secondary discharge diagnosis fields o Procedure code in principal or secondary procedure fields
Pulmonary embolism	During index admission or a subsequent inpatient admission that occurs within 30 days of the start of the index admission	ICD-10-CM code list	• Index admission – secondary discharge diagnosis fields only AND not coded as POA • Readmissions – principal or secondary discharge diagnosis fields
Death	During index admission or within 30 days of the start of the index admission	N/A	N/A
Mechanical complications	During index admission or a subsequent inpatient admission that occurs within 90 days of the start of the index admission	ICD-10-CM code list	• Index admission – secondary discharge diagnosis fields only AND not coded as POA • Readmissions – principal or secondary discharge diagnosis fields

Complication	Follow-up period in days	ICD-10 codes defining complication	Required coding placement
Periprosthetic Joint Infection / Wound Infection	During index admission or a subsequent inpatient admission that occurs within 90 days of the start of the index admission	One of the diagnosis codes in ICD-10-CM code list AND One of the procedure codes in ICD-10-PCS code list	• Index admission o Diagnosis code in secondary discharge diagnosis fields only AND not coded as POA o Procedure code in secondary procedure fields only • Readmissions o Diagnosis code in principal or secondary discharge diagnosis fields o Procedure code in principal or secondary procedure fields

A.3 Risk Adjustment

The CCs outlined in Table A3 below are used to identify risk variables in claims for discharges on or after October 1, 2015 as well as discharges prior to October 1, 2015.

The ICD-10 codes used to identify certain risk variables (e.g., post-traumatic osteoarthritis) in discharges on or after October 1, 2015 are posted on *QualityNet*; hyperlinks to these lists are provided in the table. For a list of ICD-9 codes used to identify these variables in discharges prior to October 1, 2015, please refer to the 2016 hospital procedure-specific complication measure updates and specifications report posted on *QualityNet*.

Table A3. Risk variables for MIPS THA/TKA Measure

Description of risk variable	CCs and/or ICD codes included	Variables not used in risk adjustment if occurred only during index admission (indicated by "X")
Age minus 65 (years above 65, continuous)	n/a	
Male	n/a	
Index admissions with an elective THA procedure	ICD-10-PCS codes 0SR9019, 0SR901A, 0SR901Z, 0SR9029, 0SR902A, 0SR902Z, 0SR9039, 0SR903A, 0SR903Z, 0SR9049, 0SR904A, 0SR904Z, 0SR90J9, 0SR90JA, 0SR90JZ, 0SRB019, 0SRB01A, 0SRB01Z, 0SRB029, 0SRB02A, 0SRB02Z, 0SRB039, 0SRB03A, 0SRB03Z, 0SRB049, 0SRB04A, 0SRB04Z, 0SRB0J9, 0SRB0JA, 0SRB0JZ	

Description of risk variable	CCs and/or ICD codes included	Variables not used in risk adjustment if occurred only during index admission (indicated by "X")
Number of procedures (two vs. one)	n/a	
Other congenital deformity of hip (joint)	ICD-10-CM code list	
Post traumatic osteoarthritis	ICD-10-CM code list	
Metastatic cancer and acute leukemia (CC 8)	Metastatic cancer and acute leukemia (CC 8)	
Other major cancers (CC 9-12)	Lung and other severe cancers (CC 9)	
	Lymphoma and other cancers (CC 10)	
	Colorectal, bladder, and other cancers (CC 11)	
	Breast, prostate, and other cancers and tumors (CC 12)	
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	Other respiratory and heart neoplasms (CC 13)	
	Other digestive and urinary neoplasms (CC 14)	
	Other neoplasms (CC 15)	
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	Diabetes with acute complications (CC 17)	X
	Diabetes with chronic complications (CC 18)	
	Diabetes without complications (CC 19)	
	Proliferative diabetic retinopathy and vitreous hemorrhage (CC 122)	
	Diabetic and other vascular retinopathies (CC 123)	
Protein-calorie malnutrition (CC 21)	Protein-calorie malnutrition (CC 21)	
Morbid obesity (CC 22)	Morbid obesity (CC 22)	
Bone/joint/muscle infections/necrosis (CC 39)	Bone/joint/muscle infections/necrosis (CC 39)	X
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	
Osteoarthritis of hip or knee (CC 42)	Osteoarthritis of hip or knee (CC 42)	
Osteoporosis and other bone/cartilage disorders (CC 43)	Osteoporosis and other bone/cartilage disorders (CC 43)	
Dementia or other specified brain disorders (CC 51-53)	Dementia with complications (CC 51)	
	Dementia without complications (CC 52)	
	Nonpsychotic organic brain syndromes/conditions (CC 53)	
Major psychiatric disorders (CC 57-59)	Schizophrenia (CC 57)	

Description of risk variable	CCs and/or ICD codes included	Variables not used in risk adjustment if occurred only during index admission (indicated by "X")
	Major depressive, bipolar, and paranoid disorders (CC 58)	
	Reactive and unspecified psychosis (CC 59)	
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	Quadriplegia (CC 70)	
	Paraplegia (CC 71)	
	Spinal cord disorders/injuries (CC 72)	
	Amyotrophic lateral sclerosis and other motor neuron disease (CC 73)	
	Cerebral palsy (CC 74)	
	Hemiplegia/hemiparesis (CC 103)	X
	Monoplegia, other paralytic syndromes (CC 104)	X
	Amputation status, lower limb/amputation complications (CC 189)	X
	Amputation status, upper limb (CC 190)	X
Cardio-respiratory failure and shock	Cardio-respiratory failure and shock (CC 84 plus ICD-10-CM codes R09.01 and R09.02, for discharges on or after October 1, 2015; CC 84 plus ICD-9-CM codes 799.01 and 799.02, for discharges prior to October 1, 2015)	X
Coronary atherosclerosis or angina (CC 88-89)	Angina pectoris (CC 88)	
	Coronary atherosclerosis/other chronic ischemic heart disease (CC 89)	
Stroke (CC 99-100)	Cerebral hemorrhage (CC 99)	X
	Ischemic or unspecified stroke (CC 100)	X
Vascular or circulatory disease (CC 106-109)	Atherosclerosis of the extremities with ulceration or gangrene (CC 106)	X
	Vascular disease with complications (CC 107)	X
	Vascular disease (CC 108)	X
	Other circulatory disease (CC 109)	X
Chronic obstructive pulmonary disease (COPD) (CC 111)	Chronic obstructive pulmonary disease (COPD) (CC 111)	
Pneumonia (CC 114-116)	Aspiration and specified bacterial pneumonias (CC 114)	X
	Pneumococcal pneumonia, empyema, lung abscess (CC 115)	X
	Viral and unspecified pneumonia, pleurisy (CC 116)	X

Description of risk variable	CCs and/or ICD codes included	Variables not used in risk adjustment if occurred only during index admission (indicated by "X")
Pleural effusion/pneumothorax (CC 117)	Pleural effusion/pneumothorax (CC 117)	X
Dialysis status (CC 134)	Dialysis status (CC 134)	X
Renal failure (CC 135-140)	Acute renal failure (CC 135)	X
	Chronic kidney disease, stage 5 (CC 136)	
	Chronic kidney disease, severe (stage 4) (CC 137)	
	Chronic kidney disease, moderate (stage 3) (CC 138)	
	Chronic kidney disease, mild or unspecified (stages 1-2 or unspecified) (CC 139)	
	Unspecified renal failure (CC 140)	X
Decubitus ulcer or chronic skin ulcer (CC 157-161)	Pressure ulcer of skin with necrosis through to muscle, tendon, or bone (CC 157)	X
	Pressure ulcer of skin with full thickness skin loss (CC 158)	X
	Pressure ulcer of skin with partial thickness skin loss (CC 159)	X
	Pressure pre-ulcer skin changes or unspecified stage (CC 160)	X
	Chronic ulcer of skin, except pressure (CC 161)	
Trauma (CC 166-168, 170-173)	Severe head injury (CC 166)	X
	Major head injury (CC 167)	X
	Concussion or unspecified head injury (CC 168)	X
	Hip fracture/dislocation (CC 170)	X
	Major fracture, except of skull, vertebrae, or hip (CC 171)	X
	Internal injuries (CC 172)	X
	Traumatic amputations and complications (CC 173)	X
Vertebral fractures without spinal cord injury (CC 169)	Vertebral fractures without spinal cord injury (CC 169)	
Other injuries (CC 174)	Other injuries (CC 174)	X
Major complications of medical care and trauma (CC 176-177)	Complications of specified implanted device or graft (CC 176)	X
	Other complications of medical care (CC 177)	X

Appendix B. Attribution

B.1 Candidate Attribution Rules

Our approach to identifying attribution rules was guided by historical, analytic, policy, and clinical considerations. This includes review of prior work by the NQF, a scan of methods used in existing CMS programs, a literature review, descriptive analyses of claims patterns, and consultation with clinicians and a national TEP. This appendix describes the attribution rules considered for the MIPS THA/TKA complication measure and rationale for those not adopted for the measure.

National Quality Forum Recommendations

We reviewed work completed by the NQF under contract to the Department of Health and Human Services in 2016. As part of its work, the NQF convened a researcher and clinician-based team to conduct a comprehensive literature review and environmental scan to identify attribution rules proposed for use in or implemented in healthcare delivery models. The NQF also convened a multi-stakeholder committee that reviewed the research team's findings, developed principles of fair attribution models, and developed a guide to assist measure developers and those designing payment models in selecting attribution rules.[36] The NQF Attribution Committee recognized that there are both program-level and measure-level attribution methods.

Consistent with the NQF Attribution Committee's recommendations, we considered multiple approaches determined by measure cohort and outcome. We also were attentive to the minimum standards for any attribution rule proposed by the NQF Attribution Committee:

- *Use transparent, clearly articulated methods that produce consistent and reproducible results*. Consistent with this standard, we developed attribution rules that were reproducible and straightforward to implement.
- *Ensure that accountable units can meaningfully influence measured outcomes.* We met this standard by obtaining clinical and other stakeholder input on all candidate attribution rules.

Environmental Scan

In 2017, we conducted an environmental scan to understand approaches that had been used or were currently in use for attributed hospital outcomes to individual clinicians or their practice groups. These included:

- *Value-based Payment Modifier:* two-step attribution methodology based on plurality of primary care service delivery, first assigning to primary care provider and secondly to a specialist who provides primary care service.[47]
- *Medicare Accountable Care Organizations* (*Medicare Shared Savings Program, Pioneer ACO Model, Next Generation ACOs):* two-step attribution method for beneficiaries who receive at least one primary care service from a physician

within an ACO, first assigning them to the primary care physician who provides the plurality of services and secondly to an ACO professional who provides primary care services.[47]

- *Comprehensive Primary Care Plus (CPC+):* attribution primarily based on billings for complex care management services and secondarily based on plurality of primary care visits, if not assigned in first step.[48]

- *Medicare Multi-Payer Advanced Primary Care Practice (MAPCP) Demonstration:* attribution to provider with most primary care visits and break tie with most recent visit.[49]

Literature Review

We updated the findings of the NQF Attribution Committee's literature review,[36] which evaluated medical literature through October 2016. We searched PubMed (January 1, 2016 to January 4, 2017) and EMBASE (January 1, 2016 to January 4, 2017) to identify any new attribution methods not captured in the NQF's 2016 report. We adopted the NQF's search strategy and supplemented their search strategy by consulting content experts to include additional studies focused on assigning beneficiaries to clinicians.

Our literature review identified several attribution approaches that were used in high-impact or multiple studies; we considered these as candidates for the current assessment. These included:

- Plurality of charges or claims during a fixed time frame.
- Most recent charges/claims/visits prior to an event.
- Procedure claim for patients undergoing a procedure.

Descriptive Analysis of Claim Patterns

In order to better understand patterns of care that could help identify or exclude from consideration different attribution rules, we examined the patterns of claims around each inpatient stay for THA/TKA (data not shown). This included both institutional and outpatient claims. We also examined the distribution in numbers and types of ECs seen by patients during their hospitalization and the completeness of institutional claims with respect to unique National Provider Identifiers (NPI) and Taxpayer Identification Number (TIN) combinations. These kinds of data, while not used for evaluation of the attribution approaches, provided a profile of the kinds of clinician contact patients in a given measure cohort had prior to and during their hospitalization to help identify feasible attribution rules.

Clinical and Expert Input

We organized a group of clinician researchers at CORE and convened a national TEP. We gave them background information on the objectives of the measure and our initial list of candidate attribution approaches. We then solicited their thoughts or concerns about

potential attribution rules and their input on any additional attribution rules we should consider. Table B1 lists the attribution rules we considered.

The TEP favored attributing the complication outcome to a single EC or EC group using the Billing Surgeon attribution.

Table B1. Attribution rules considered for THA/TKA complication measure

Attribution rule	Definition	Justification for inclusion as candidate attribution rule
1. Attending	Identified as the "attending provider" on the inpatient claim	Logically responsible for patient care and discharge transition
2. Operator	Identified as the "operating provider" on the inpatient claim	Logically responsible for operation and discharge transition
3. Billing Surgeon	Identified using an algorithm shown in Figure B1, using fields on institutional claim and Part B claim lines	Commonly used definition for identifying clinician responsible for patient care

Empiric Analysis

Finally, we empirically evaluated the candidate attribution rules to understand the implications of each approach with regards to feasibility, validity, reliability, and sample size. Our analytic evaluation was attentive to the minimum standards for any attribution rule proposed by the NQF Attribution Committee:

- *Use adequate sample sizes, outlier exclusion, and/or risk adjustment to fairly compare the performance of attributed units.* We examined sample size distribution and outlier patterns and used the same risk-adjustment model as CMS's hospital-level THA/TKA complication measure.

- *Conduct sufficient testing with scientific rigor at the level of accountability being measured*. Though additional testing would be necessary before adoption, we undertook implementation that was consistent with the CMS's hospital-level THA/TKA measure, which has been rigorously tested.

The analytic evaluation of each attribution rule focused on the following aspects of each:

- *Face validity*: For each approach, we assessed face validity by summarizing the number and percent of unattributed patients as well as rates of missing clinician or TIN information. The distribution also provided face validity in that an attribution rule which led to unexpected or senseless results would unlikely be accepted by stakeholders. Implementation also provided a measure of feasibility; if an approach led to a high proportion of unattributed patients, then it was considered less valid. Thus, we examined the patterns of volume for ECs and EC groups overall and by specialty.

- *Differentiation among providers*: The greater the variation in entity performance, the more evidence that the attribution is aligned with some underlying true quality signal. Therefore, for each attribution method, we examined: the distribution of unadjusted outcome rates across physicians and TINs; the between-clinician and between-TIN variance estimated from an HGLM for different volume cut-offs; distribution of risk-standardized rates; and the impact of risk adjustment on these variances.

- *Reliability and sample size*: Reliability relates the accuracy of measurement to the sample size of the measured entities. For each approach, we calculated the estimated average unit (EC and EC group) reliability for a volume cut-off of 25 cases.

- *Overlap with other attribution rules*: As recommended by the NQF Attribution Committee, we examined the overlap between the different candidate attribution rules. If several different attribution rules were consistent (had high overlap), then the overlap would suggest little practical difference in choosing among them. For all attribution rules that assigned to a single entity, we summarized how much pairwise overlap there was in their assignments.

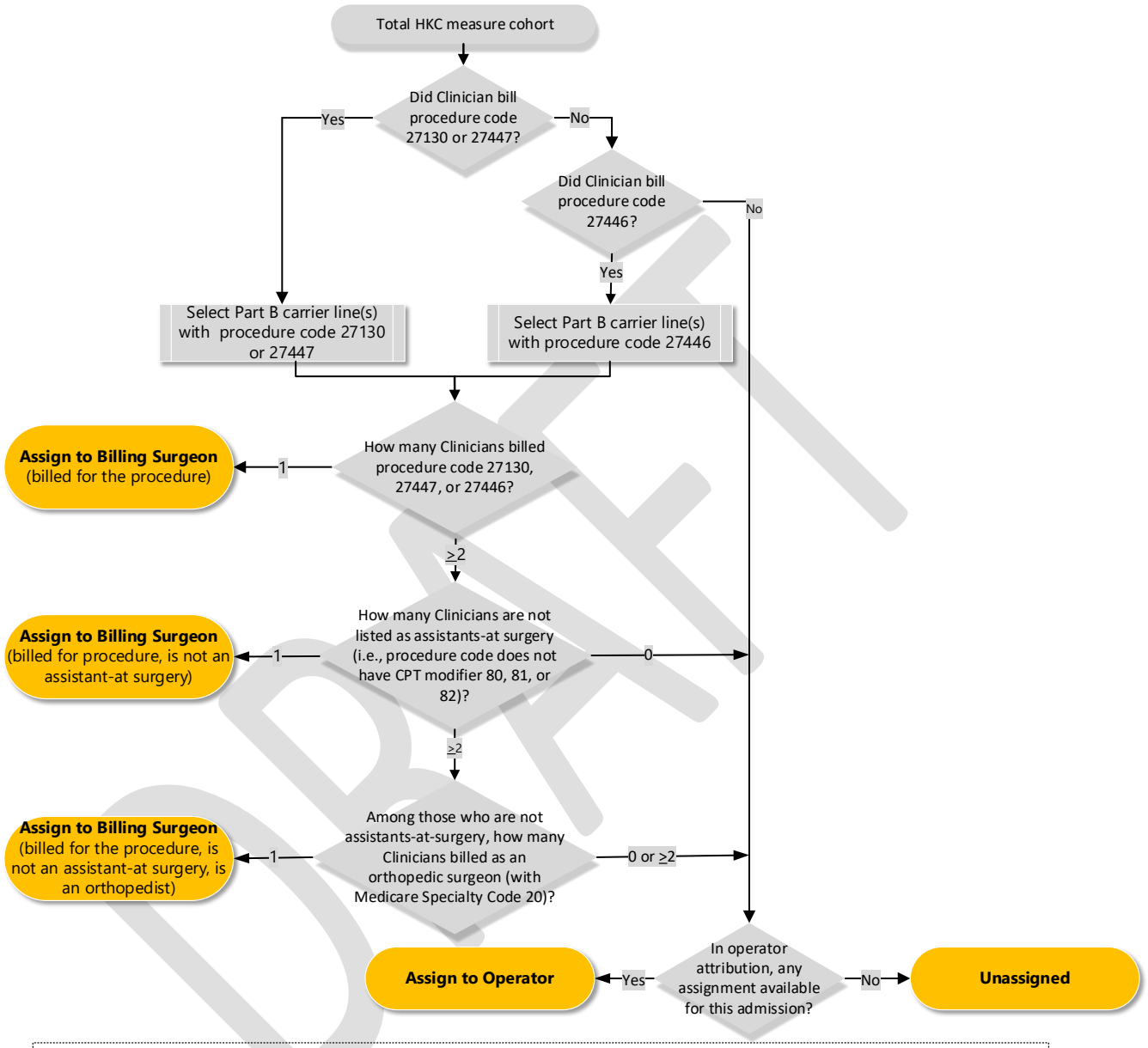
For all attribution rules, we evaluated implementation at the EC and EC group-levels (data not shown).

B.2 Final Attribution

CORE sought consensus from a national TEP around which of the three rules should be used for the MIPS THA/TKA complication measure. The TEP strongly supported attributing the outcome to the Billing Surgeon. Conceptually, the TEP supported that the clinician being compensated for performing the procedure should have primary responsibility for patient outcomes. The alternative approaches all use information from the institutional claim submitted by the hospital. The TEP felt these options were less under a clinician's control than the Billing Surgeon that is defined by clinician billing claims and therefore less valid for a MIPS measure.

The MIPS THA/TKA complication measure attributes the outcome for each patient in the cohort to a single clinician or clinician group. Figure B1 shows the approach to attribution to the Billing Surgeon at the MIPS EC level. An EC is identified through his/her unique NPI and TIN combination. The MIPS EC group assigned is identified as the TIN on the Part B carrier claim line item for the procedure identified by the attribution in Figure B1.

Figure B1. Approach to identifying Billing Surgeon in Medicare claims data at MIPS EC level; EC group is the TIN of the attributed EC



KEY

Clinician: A unique National Provider Identifier (NPI) and Taxpayer Identification Number (TIN) combination

Current Procedural Terminology (CPT®) codes

- **27130:** Arthroplasty, acetabular and proximal femoral prosthetic replacement (total hip arthroplasty), with or without autograft or allograft
- **27446:** Arthroplasty, knee, condyle and plateau; medial OR lateral compartment
- **27447:** Arthroplasty, knee, condyle and plateau; medial AND lateral compartments with or without patella resurfacing (total knee arthroplasty)

CPT® modifiers

- **80:** Assistant Surgeon
- **81:** Minimum Assistant Surgeon
- **82:** Assistant Surgeon (when qualified Resident Surgeon not available)

Appendix C. Statistical Approach

The MIPS THA/TKA re-specification measure uses an HGLM to estimate RSCRs for ECs and EC groups (providers). This modeling approach accounts for the within-provider correlation of the observed outcome and accommodates the assumption that underlying differences in quality across ECs or EC groups lead to systematic differences in outcomes.

In the MIPS THA/TKA measure, an HGLM model is estimated. Then for each EC or EC group, a standardized risk ratio (SRR) is calculated. The RSCR is calculated by multiplying the SRR for each ECs or EC groups by the national observed complication rate.

C.1 Hierarchical Generalized Linear Model

We fit an HGLM, which accounts for clustering of observations within ECs or EC groups (providers). We assume the outcome has a known exponential family distribution and relates linearly to the covariates via a known link function, *h*. Specifically, we assume a binomial distribution and a logit link function. Further, we account for the clustering within providers by estimating a provider-specific effect, α_i, which we assume follows a normal distribution with a mean μ and variance τ^2, the between-provider variance component. The following equation defines the HGLM:

$$h\left(\Pr\left(Y_{ij}=1\middle|\boldsymbol{Z}_{ij},\omega_i\right)\right)=log\left(\frac{\Pr(Y_{ij}=1|\boldsymbol{Z}_{ij},\omega_i)}{1-\Pr(Y_{ij}=1|\boldsymbol{Z}_{ij},\omega_i)}\right)=\alpha_i+\boldsymbol{\beta}\boldsymbol{Z}_{ij} \quad (1)$$

$$\text{where } \alpha_i=\mu+\omega_i;\ \omega_i\sim N(0,\tau^2)$$

$$i=1,\ldots,\ I;\ j=1,\ldots,n_i$$

where Y_{ij} denotes the outcome (equal to 1 if the patient has a complication, 0 otherwise) for the *j-th* patient at the *i-th* provider; $Z_{ij}=\left(Z_{ij1},Z_{ij2},\ldots,Z_{ijp}\right)^T$ is a set of *p* patient-specific covariates derived from the data; and *I* denotes the total number of providers and n_i denotes the number of index admissions at provider *i*. The provider-specific intercept of the *i-th* provider, α_i, defined above, comprises μ, the adjusted average intercept over all providers in the sample, and ω_i, the provider-specific intercept deviation from Daniels MJ and Gatsonis C, 1999.[50]

We estimate the HGLM using the SAS software system (GLIMMIX procedure).

C.2 Risk-standardized Measure Score Calculation

Using the HGLM defined by Equation (1), to obtain the parameter estimates $\hat{\mu},\{\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_I\},\widehat{\boldsymbol{\beta}}$, and $\hat{\tau}^2$, we calculate an SRR, $\hat{s}_i$, for each EC or EC group by computing the number of the predicted complications to the number of expected complications. Specifically, we calculate:

$$\text{Predicted Value: } \hat{p}_{ij}=h^{-1}\left(\hat{\alpha}_i+\widehat{\boldsymbol{\beta}}\boldsymbol{Z}_{ij}\right)=\frac{\exp(\hat{\alpha}_i+\widehat{\boldsymbol{\beta}}\boldsymbol{Z}_{ij})}{\exp(\hat{\alpha}_i+\widehat{\boldsymbol{\beta}}\boldsymbol{Z}_{ij})+1} \quad (2)$$

$$\text{Expected Value: } \hat{e}_{ij}=h^{-1}\left(\hat{\mu}+\widehat{\boldsymbol{\beta}}\boldsymbol{Z}_{ij}\right)=\frac{\exp(\hat{\mu}+\widehat{\boldsymbol{\beta}}\boldsymbol{Z}_{ij})}{\exp(\hat{\mu}+\widehat{\boldsymbol{\beta}}\boldsymbol{Z}_{ij})+1} \quad (3)$$

Standardized Risk Ratio: $\hat{s}_i = \frac{\sum_{j=1}^{n_i} \hat{p}_{ij}}{\sum_{j=1}^{n_i} \hat{e}_{ij}}$ (4)

We calculate an RSCR, $\widehat{RSCR}_i$, for each EC or EC group by using the estimate from Equation (4) and multiplying by the national observed complication rate, denoted by $\bar{y}$. Specifically, we calculate:

Risk-Standardized Complication Rate: $\widehat{RSCR}_i = \hat{s}_i \times \bar{y}$ (5)

C.3 Creating Interval Estimates

The measure score is a complex function of parameter estimates; therefore, we use re-sampling and simulation techniques to derive an interval estimate to determine if an EC or EC group is performing better than, worse than, or no different than expected. An EC or EC group is considered better than expected if the upper bound of their CI falls below the national observed complication rate, $\bar{y}$, and considered worse if the lower bound of their CI falls above $\bar{y}$. An EC or EC group is considered no different than expected if the CI overlaps $\bar{y}$.

More specifically, we use bootstrapping procedures to compute CIs. Because the theoretical-based standard errors are not easily derived, and to avoid making unnecessary assumptions, we use the bootstrap to empirically construct the sampling distribution for each EC or EC group risk-standardized ratio. The bootstrapping algorithm is described below.

C.4 Bootstrapping Algorithm

Let *I* denote the total number of ECs or EC groups in the sample. We repeat steps 1 – 4 below for b = 1,2, …B times:

1. Sample *I* ECs or EC groups with replacement.
2. Fit the hierarchical logistic regression model defined by Equation (1) using all patients within each sampled EC or EC group. The starting values are the parameter estimates obtained by fitting the model to all EC or EC groups. If some ECs or EC groups are selected more than once in a bootstrapped sample, we treat them as distinct so that we have random effects to estimate the variance components. After Step 2, we have:
   a. The estimated regression coefficients of the risk factors, $\widehat{\boldsymbol{\beta}}^{(b)}$.
   b. The parameters governing the random effects, provider adjusted outcomes, distribution $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.
   c. The set of provider-specific intercepts and corresponding variances, $\left\{\hat{\alpha}_i^{(b)} - \hat{var}\left(\alpha_i^{(b)}\right); i = 1,2, \ldots, I\right\}$
3. We generate a provider level random effect by sampling from the distribution of the provider-specific distribution obtained in Step 2c. We approximate the distribution for

each random effect by a normal distribution. Thus, we draw $\alpha_i^{(b*)} \sim N(\hat{\alpha}_i^{(b)}, \hat{var}\left(\alpha_i^{(b)}\right))$ for the unique set of providers sampled in Step 1.

4. Within each unique EC or EC group *i* sampled in Step 1, and for each case *j* in that EC or EC group, we calculate $\hat{p}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i^{(b)}$ where $\widehat{\boldsymbol{\beta}}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\alpha_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the provider-standardized outcome can be computed by identifying the 2.5th and 97.5th percentiles of a large selected number of estimates for all providers (or the percentiles corresponding to the alternative desired intervals).[51]