NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3732

Corresponding Measures:

Measure Title: Hospital-Level 90-Day Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for a Combined Inpatient (IP) and Outpatient (OP) Setting (IP/OP 90-Day THA/TKA Complication Measure)

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: The measure estimates a hospital-level risk-standardized complication rate (RSCR) associated with elective primary THA and TKA procedures performed in hospital inpatient and outpatient departments on Medicare fee-for-service (FFS) beneficiaries who are 65 years of age or older. The outcome (complication) is defined as any one of the specified complications occurring within 90 days of the index encounter. More details on the outcome are provided in sp.13.

Note: An index encounter is defined as either an inpatient admission for a THA/TKA procedure or a hospital outpatient department THA/TKA procedure to which the complication outcome is attributed.

1b.01. Developer Rationale:

The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level risk standardized complication rates (RSCRs) following elective primary THA and TKA occurring in hospital-based inpatient and outpatient settings. This measure is intended for use in a To-Be-Determined CMMI Payment Model that includes THA and TKA procedures performed in both the inpatient and outpatient settings. This will allow more complete and uniform performance measurement of hospitals performing THA and TKA procedures and the patients they serve.

Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process of care measures. 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, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk adjust for patients' conditions at the time of hospital admission or outpatient procedure and then evaluate patient outcomes. This measure is being developed to identify institutions whose performance is better or worse than would be expected based on each institution's patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Measuring and reporting RSCRs that include procedures performed in both the inpatient and outpatient hospital setting will help inform health care providers about opportunities to improve care, strengthen incentives for quality improvement, and promote improvements in the quality of care received by patients and the outcomes they experience. The measure will also increase transparency by providing patients with information to help guide their choices about where they seek care for these elective procedures.

Complications in the outpatient hospital setting are more likely to be influenced by care coordination, a focus of potential episode payment programs. This measure will serve as a critical surveillance tool to help ensure cost reductions (e.g., reducing use of costly inpatient rehabilitation and/or shifting patients to the outpatient setting) do not adversely impact care.

sp.12. Numerator Statement: The outcome for this measure is a dichotomous assessment of whether a complication occurred ("yes" for any complication[s]; "no" for no complications) within 90 days of the index encounter (see sp.14 for Numerator Details). Complications are counted in the measure outcome only if they occur during the index inpatient or outpatient encounter or are captured in claims associated with an inpatient readmission, observation stay, emergency department visit, or ambulatory surgical center (ASC) encounter.

sp.14. Denominator Statement:

The target population for the measure includes Medicare fee-for-service (FFS) beneficiaries who are 65 years of age and older undergoing elective primary THA and/or TKA procedures in the hospital inpatient and outpatient settings.

Additional details are provided in sp.16 Denominator Details.

sp.16. Denominator Exclusions:

This measure excludes index encounters for the following patients:

1. Without at least 90 days post-discharge enrollment in Medicare FFS.

   Rationale: The 90-day complication outcome cannot be assessed in this group since claims data are used to determine whether a complication of care occurred.
2. Who were discharged against medical advice (AMA)

   Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.
3. Who had more than two THA/TKA procedure codes during the index encounter.

   Rationale: Although clinically possible, it is highly unlikely patients would receive more than two elective THA/TKA procedures in one hospitalization, which may reflect a coding error.

After applying these exclusion criteria, we randomly select one index encounter for patients who have multiple index encounters in a calendar year. We therefore exclude the eligible index encounters that are not randomly selected in that calendar year.

Measure Type: Outcome

sp.28. Data Source:

Claims

Other (specify)

Medicare Enrollment Database

sp.07. Level of Analysis:

Facility

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a ***health outcome*** measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a new outcome measure at the facility level that estimates a hospital-level risk standardized complication rate associated with Total Knee Arthroplasty (TKA)/Total Hip Arthroplasty (THA) for a combined inpatient and outpatient setting.
- The developer provides a logic model that depicts that hospital practices, such as guideline based protocols, patient education, discharge planning, and follow-up care can lead to improved patient health status which leads to reduced risk of complication.

Summary:

- The developer referenced several studies that describe care processes and clinical interventions that are associated with reduced risk of complications:
    - A 2010 study found that patients operated on in hospitals that used critical pathways had a lower risk of postoperative complications, compared to patients operated on in hospitals without critical pathways.
        - The study also found that patients at critical pathway hospitals had lower odds of any adverse event and an average length of stay that was 0.5 days shorter after controlling for patient and hospital level factors.
    - A 2017 literature review summarized methods for establishing a successful outpatient Total Joint Arthroplasty (TJA) program. The study outlined a number of patient and provider level strategies for minimizing postsurgical complications, such as patient risk assessment, care coordination, staff training, and the use of proper sterilization and pain management techniques
    - A 2019 study stressed prevention methods that are within a hospital's control for TJA such as proper patient/family education, opioid-sparing analgesia, and prompt care coordination post-discharge.
- The developer also presented studies that stated that complication rates vary across both inpatient and outpatient settings:
    - A 2021 study found outpatient procedures were associated with lower rates of adverse events with no increase in the 30-day readmissions when compared to risk-matched inpatient procedures.

- Other studies have found outpatient procedures to have comparable complication rates to inpatient procedures.

Question for the Committee:

- *Is there at least one action that the facility can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm

Health Outcome (Box 1) -> Empirical data demonstrates the relationship between the outcome and at least one healthcare action (Box 2) -> Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer presented performance data using Medicare Fee-for-service (FFS) claims data from THA and TKA procedures performed between April 1, 2018, and March 31, 2021 (n=989,587 procedures from 3,452 hospitals).
- The Mean RSCR was 2.91 percent with a standard deviation of .47, ranging from a min of 1.53 percent to a max of 5.86 percent. The 25th percentile performance was 2.65 percent and the 75th percentile performance was 3.12 percent.
- They also presented RSCR for hospitals with at least 25 procedures and results were similar.

Disparities

- The developer presented RSCRs for TJA between hospitals with a lower proportion (bottom quartile) of Black/non-White/Dual eligible/AHRQ SES Index beneficiaries and hospitals with a higher proportion (top quartile) of Black/non-White/Dual eligible/AHRQ SES Index beneficiaries.
- Among the 2,733 hospitals with at least 25 index procedures, the mean RSCR of hospitals with a lower proportion of Black beneficiaries is slightly lower than the mean of hospitals with a higher proportion of Black beneficiaries (2.99percent vs 3.02percent).
- Among the 2,733 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer Other non-White beneficiaries combined is lower than the mean of hospitals with more Other non-White beneficiaries (2.95percent vs 2.99percent).
- Among all 2,747 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer dually eligible beneficiaries is lower (2.86percent vs 3.02percent).
- Among the 2,740 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer beneficiaries with a lower ASI is lower compared to hospitals with a higher proportion of beneficiaries with a higher ASI (2.85percent vs 3.00percent).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are the disparities in performance score by subgroups across hospitals meaningful?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluators: Staff

2a. Reliability: Specifications and Testing

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- Measure specifications are clear and precise.

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
    - Reliability testing was conducted using signal to noise method. Facility-to-Facility variance is estimated from a hierarchical logistic regression model.
    - The median signal-to-noise reliability across all hospitals with at least one case of eligible THA/TKA procedure is 0.78 (IQR: 0.48 – 0.91).
    - The median signal-to-noise reliability across hospitals with at least 25 cases, is 0.85 (IQR: 0.69 – 0.93).
    - The developer states the signal-to-noise reliability results for facilities with at least 25 procedures indicates that this measure is sufficiently reliable.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity: Validity testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Validity Testing

- Validity testing conducted at the Accountable Entity Level:

- The developer assessed the measure's correlation with other measures that target the same domain of quality to determine if better performance on the RSCR measure was related to better performance on other relevant structural or outcome measures. The following measures were tested.
    - *Overall Hospital Star Rating Summary Score:* The correlation between THA/TKA complications and Star-Rating summary score is -0.101 (p-value <.0001) for all hospitals and -0.121 (p-value <.0001) for hospitals with 25+ cases, which suggests that hospitals with lower THA/TKA RSCRs are more likely to have higher Star-Rating summary scores, especially at the extremes.
    - *Hospital THA/TKA Surgical Volume:* There is a general trend that high volume hospitals (those in the upper quartiles) have lower RSCRs than hospitals in other volume quartiles. Hospitals with the highest volumes show a relatively smaller range of RSCRs compared to the other three quartiles (Figure 3). Statistical tests were not provided.
- The developer also assessed the face validity of the measure by surveying a multi-stakeholder TEP with 13 members that included physicians, consumers, purchasers and individuals with experience in quality improvement. The developer found the following:
    - 100 percent agreement among the 12 respondents on whether the measure will provide a valid assessment of complications following elective THA/TKA
    - 100 percent agreement among the 12 respondents on whether the measure can be used to distinguish between better and worse quality care among facilities
    - 5 members provided additional feedback encouraging CMMI to account for patient social risk and to monitor racial and other disparities.
    - One commenter recommended considering opiate use in future measure iterations.

Exclusions

- The measure excludes index encounters for the following patients:
    - Patients without at least 90 days post-discharge enrollment in Medicare FFS to ensure the ability to determine whether a complication in care occurred (n=7,532 (0.72 percent)).
    - Patients who were discharged against medical advice (AMA) because the provider did not have the full opportunity to care for the patient (n=192 (0.02 percent)).
    - Patients who had more than two THA/TKA procedure codes during the index encounter. This likely reflects a coding error (n=0).
- The developer states that the exclusions accounted for a very small percentage of the index encounters and were unlikely to impact the measure score.

Risk-Adjustment

- The measure uses a logistical regression model on the combined IP/OP data. The model includes indicators for age, gender, number of procedures, a variety of comorbidities, and setting (IP/OP).
- They tested three different models: 1. IP Only, OP only, Combined IP/OP with setting indicator and chose the final model based on C-statistic and predictive ability.
- The C-statistic for the final risk model is 0.66. They evaluated predictive ability, a comparison of highest to lowest deciles to evaluate the ability of the model to distinguish high-risk subjects from low-risk subjects; the predictive ability from the lowest to highest decile is 1.18 percent-7.10 percent.

- The developer extensively tested the inclusion of social risk factors. They found that the overall prevalence of these factors was low (Black: 4.4percent, Dual Eligible: 3.5percent, AHRQ SES Index (ASI): 10.9percent). Odds of complications were lower or equivalent for Black or other non-White patients compared to White patients and higher for dual-eligible and low ASI patients. The developer did not include these factors in the model to align with the original inpatient-only measure (NQF #1550).
- With regard to model performance in subgroups of patients with social risk factors, the risk-decile plot for patients with the low ASI and exposure to race variables were similar to that for all patients, suggesting the base model is well calibrated for these subgroups. However, risk-decile plots show that the base model underpredicts risk for patients with dual eligibility.

Meaningful Differences

- Out of all 3,452 hospitals, the mean risk-standardized complication rate (RSCR) is 2.91percent, and the median RSCR is 2.86percent. The maximum and minimum are 5.86percent and 1.53percent, respectively, with IQR being 2.65-3.12percent.
- Out of the 2,747 hospitals that have at least 25 index procedures, the mean RSCR is still 2.91percent, and the median RSCR is 2.85percent. The maximum and minimum are the same as those in all hospitals, with IQR being 2.59-3.18percent.
- The median odds ratio was 1.33. This represents the median increase in odds of a complication within 90 days of a THA/TKA admission date on a single patient if the admission occurred at a higher risk hospital compared to a lower risk hospital. This indicates that a patient has a 33percent increase in the odds of a complication at a higher risk hospital.

Missing Data

- There is no missing data in the development and testing data.

Comparability

- The measure only uses one set of specifications.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *The developer found that complication rates were significantly higher for both dually eligible patients and those in a low ASI area. Is the developer rationale sufficient for not including these factors in the final risk model specifications?*
- *Are the other validity tests provided sufficient to demonstrate validity of the measure?*

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The data source for the measure is administrative claims and eligibility data from CMS.
- All data elements are in defined fields in a combination of electronic sources.
- The information is coded by someone other than the person obtaining original information

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 4: Use and Usability

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported?	☐	Yes	☒	No		
Current use in an accountability program?	☐	Yes	☒	No	☐	UNCLEAR
Planned use in an accountability program?	☒	Yes	☐	No	☐	NA

Accountability program details

- The developer states the measure is not currently being used in any accountability program and intends for the measure to be used in a cross-setting payment model, such as CMMI's existing Comprehensive Joint Replacement (CJR) model, in the future.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- To date, feedback was obtained via five teleconference meetings with the Clinical Working Group, two teleconference meetings with the TEP, and a 30-day Public Comment period where feedback was

received from three medical associations and societies (American Medical Association [AMA], American Association of Hip and Knee Surgeons [AAHKS], American Society of Anesthesiologists [ASA]). The TEP, Clinical Working Group, and public comments indicated strong support for the measure specifications and provided recommendations for ongoing evaluation, such as consideration of shifts in outpatient procedure volume, the impact of social determinants of health, and disparities in access.

- Measured entities (acute care hospitals) and other stakeholders or interested parties may submit questions or comments about the measure/measure development through the QualityNet Q&A tool.
- The feedback received from the TEP patient representatives showed their interest in the measure and support for the measure rationale and decision to expand the measure to the hospital outpatient setting. Patients also provided feedback about how closely the measure mirrored the actual experience of having a total hip or knee replacement and of experiencing a complication. Patient representatives also shared their thoughts about the value of the measure for patients, stating that it would give them the ability to be better advocates for their own health.
- The TEP recommended ongoing evaluation of the risk model and analyses on the social determinants of health.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer states that this information is not yet available as 9his is a new measure.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- No unexpected findings were noted during measure development or testing.

Potential harms

- No unexpected findings were noted during measure development or testing.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 5: Related and Competing Measures

Related measures

- 1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
- 1551: Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
- 3474: Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)
- 3493: Risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) for Merit-based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups
- 3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
- 3639: Clinician-Level and Clinician Group-Level Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA and TKA) Patient-Reported Outcome-Based Performance Measure (PRO-PM)

Harmonization

- The developer states that the measures have been harmonized with existing related measures to the extent feasible, specifically with CMS' existing hospital-level THA/TKA complication and readmission measures.
- The developer also notes that while this measure represents the same outcome and a similar patient population (patients undergoing elective primary THA/TKA procedures) as Measure NQF#1550: *Hospital-Level Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA),* the goal of this measure is to serve as a more accurate quality assessment tool for payment models that cross care settings. Therefore, this measure will never directly compete with NQF#1550 in a CMS program because it is intended for use only in applications that include inpatient and outpatient hospital settings.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a.01. Provide a logic model.

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

[Response Begins]

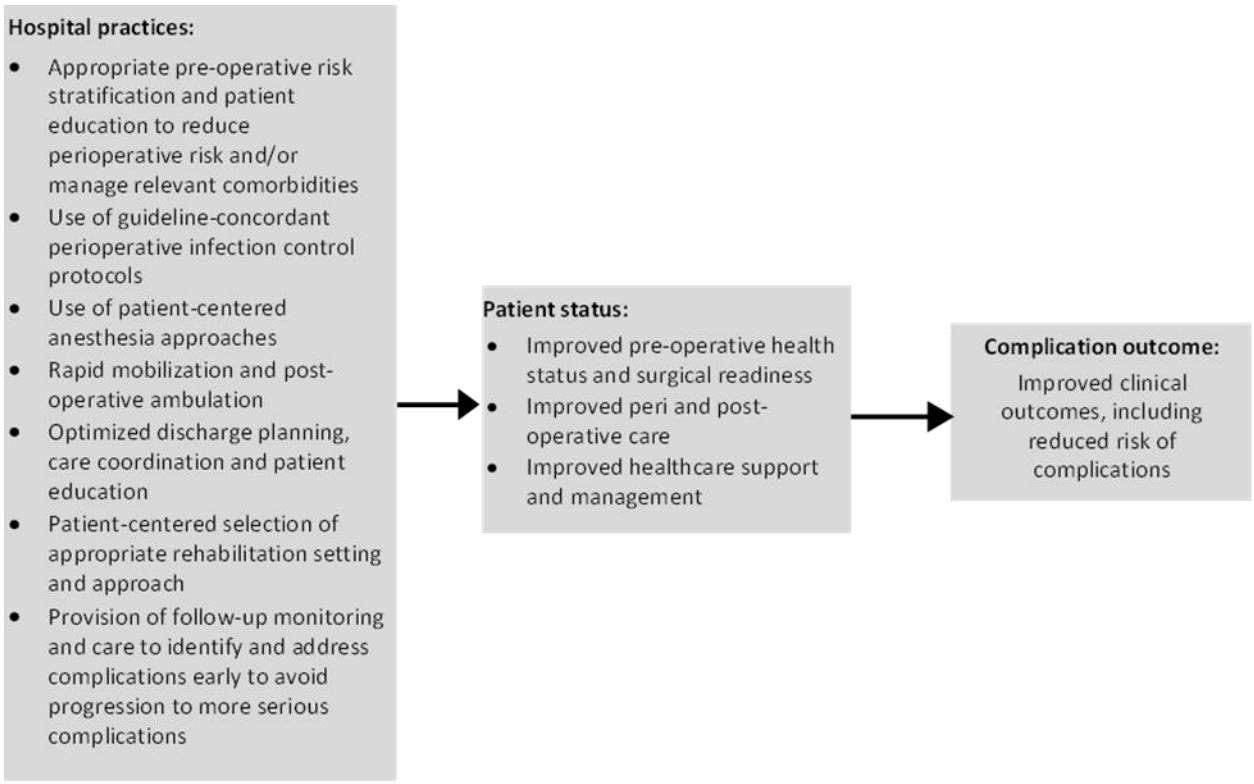

Figure 1. Logic Model.

Total hip arthroplasty (THA) and total knee arthroplasty (TKA) complications are a priority area for outcome measure development, as they are outcomes that are likely attributable to care processes. Elective primary THA and TKA procedures have been in increasing demand recently. Since the Centers for Medicare and Medicaid Services (CMS) began allowing Medicare providers to bill for TKAs and THAs in the outpatient setting (Centers for Medicare and Medicaid Services 2018, Centers for Medicare and Medicaid Services 2019, Centers for Medicare and Medicaid Services 2020), the proportion of THA and TKA procedures performed in the outpatient hospital setting has exceeded that of the inpatient setting. The lower cost and greater convenience of the outpatient setting (hospital outpatient departments [HOPDs] and Ambulatory Surgical Centers [ASCs]) make outpatient procedures an attractive option for patients, surgeons, and hospitals alike. This has led to an increase in the number of procedures being performed in the outpatient setting (Kurtz et al, 2007; Lopez et al, 2020; Kurtz et al, 2014). The goal of this measure is to improve patient outcomes by giving patients, providers, and hospitals information on complications that occur after THA and TKA procedures performed on Medicare patients in both hospital inpatient and outpatient settings. Measurement of THA/TKA outcomes provides a broader view of a hospital's quality of care, encompassing more than what can be captured by individual process of care

measures. Findings from several studies suggest that some complications after THA/TKA procedures are preventable. For example, studies show that critical aspects of care, such as communication between providers; patient education and safety practices; and coordinated care transitions in the hospital inpatient and outpatient environment (Browne et al, 2010; Kurtz et al, 2010; Mariorenzi et al, 2020) can help improve patient outcomes, but would be difficult to capture using individual process measures.

Generally speaking, a critical aspect of outcomes measurement is to risk adjust for patients' conditions at the time of hospital admission. This helps account for the influence of sick patients who may be disproportionately distributed across providers. The goal of this respecified version of a preexisting inpatient-only THA/TKA complication measure is to expand the cohort (denominator) to include outpatient THA/TKA procedures and also expand the outcome (numerator) to include serious complication events occurring in outpatient settings; these updates allow for better capture of procedures and complications regardless of setting. As the number of TKA and THA procedures being performed in the outpatient setting continues to rise, so does the need to monitor the quality of care being delivered across the hospital inpatient and outpatient settings. Measuring and reporting THA/TKA complication rates will serve as a useful tool to help inform healthcare providers and facilities about opportunities to improve the quality of the care they deliver to Medicare patients, as well as provide much needed information to patients to help guide their choices about where to receive care.

References

1. Browne, J.A., et al., *Postoperative morbidity and mortality following total knee arthroplasty with computer navigation.* Knee, 2010. **17**(2): p. 152-6.
2. Centers for Medicare & Medicaid Services (CMS), HHS. "Medicare Program: Changes to Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems and Quality Reporting Programs. Final Rule With Comment Period." *Federal Register,* vol. 83, no. 225, 2018, pp. 58818-9179.
3. Centers for Medicare and Medicaid Services (CMS), HHS. "Medicare Program: Changes to Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems and Quality Reporting Programs. Final Rule With Comment Period." *Federal Register,* vol. 84, no. 228, 2019, pp. 61142-61492.
4. Centers for Medicare and Medicaid Services (CMS), HHS. "Medicare Program: Changes to Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems and Quality Reporting Programs. Proposed Rule." *Federal Register*, vol 85, no. 159, 2020, pp. 50074-50665.
5. Kurtz, S.M., et al., *Impact of the economic downturn on total joint replacement demand in the United States: updated projections to 2021.* J Bone Joint Surg Am, 2014. **96**(8): p. 624-30.
6. Kurtz, S., et al., *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): p. 780-5.
7. Kurtz, S.M., et al., *Prosthetic joint infection risk after TKA in the Medicare population.* Clin Orthop Relat Res, 2010. **468**(1): p. 52-6.
8. Lopez, C.D., et al., *Hospital and Surgeon Medicare Reimbursement Trends for Total Joint Arthroplasty.* Arthroplasty Today, 2020. **6**(3): p. 437-444.
9. Mariorenzi, M., Levins, J., Marcaccio, S., Orfanos, A., & Cohen, E. (2020). Outpatient Total Joint Arthroplasty: A Review of the Current Stance and Future Direction. R I Med J (2013), 103(3), 63-67.

[Response Ends]

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

Describe how and from whom input was obtained.

[Response Begins]

Patients who have undergone a THA or TKA have been engaged for input on measure development through participation on the Technical Expert Panel (TEP). In alignment with the Centers for Medicare & Medicaid Services (CMS) Measures Management System (MMS), the Center for Outcomes Research & Evaluation (CORE) convened a TEP to provide feedback and recommendations on key methodological and clinical decisions in measure development. Four female patients provided input through participation in the TEP meetings in August 2020 and June 2022. Feedback from patients indicate interest in, and support for, the measure outcome and rationale for the expanded focus on inpatient and outpatient hip/knee procedures. Patients also expressed appreciation for how closely the measure related to the actual experience of having a hip and knee replacement and of experiencing a complication. Patient representatives on the TEP also indicated that the measure was very meaningful, stating that it would have value for patients; it would give them the ability to make better decisions and be better advocates for their own health before undergoing total hip/knee surgery.

[Response Ends]

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

[Response Begins]

This measure is intended for use in a To-Be-Determined CMMI Payment Model that includes THA and TKA procedures performed in both the inpatient and outpatient settings.

THA and TKA (collectively, Total Joint Arthroplasty (TJA)) are common and effective surgical procedures that can significantly improve a patient's quality of life through pain reduction and improved function. These procedures are being increasingly performed in more diverse surgical locations; as such, quality assurance across surgical settings is essential.

Multiple Care Processes and Clinical Interventions are Associated with Complication Risk

Findings from several studies suggest complications after THA/TKA procedures are preventable. Husni et al. examined the influence of critical pathways (care plans meant to achieve optimal procedure efficiency by delineating the sequence of actions) on the postoperative outcomes after TKA procedures. The authors found that patients operated upon in hospitals that used critical pathways had a lower risk of postoperative complications, including death, compared to patients operated on in hospitals without critical pathways (Husni et al, 2010). Researchers also found that patients receiving care under critical pathways had significantly lower odds of experiencing any adverse event compared to patients without a critical pathway (adjusted odds ratio [OR] = 0.68, 95% CI 0.50, 0.92), and had an average length of stay that was 0.5 days shorter after controlling for patient- and hospital-level factors (Husni et al, 2010). Similarly, in their literature review, Bert et al. summarize methods for establishing a successful outpatient TJA program within the framework of bundled payment reimbursement (Bert et al, 2017). They outlined a number of patient- and provider-level strategies for minimizing postsurgical complications and improving patient recovery including patient risk assessment, care coordination, staff training, and the use of proper sterilization and pain management techniques. Li et al. also stressed proper patient selection, proper patient/family education, opioid-sparing analgesia, and prompt care coordination post-discharge. This would suggest that prevention of postsurgical complications is possible and that many prevention methods are within hospitals' control (Li et al, 2019).

Complication Rates Vary across Both Inpatient and Outpatient Settings

Studies examining differences in complication rates in the inpatient versus the outpatient setting have arrived at differing conclusions. Some studies have shown lower complication rates after outpatient procedures compared to inpatient procedures, suggesting that the outpatient setting represents a productive alternative in select patient populations (Kimball et al, 2020; Greenky et al, 2019; Lan et al, 2021). For example, Lan et al. found outpatient procedures were associated with lower rates of adverse events, with no increase in the rate of 30-day readmissions when compared to risk-matched inpatient procedures (Lan et al, 2021). Studies have also found outpatient procedures to have comparable complication rates to inpatient procedures (Arshi et al, 2019; Xu et al, 2019; Aynardi et al, 2014; Bert et al, 2017; Goyal et al, 2017; Darrith et al, 2019; Migliorini et al, 2021; Mariorenzi et al, 2013). For example, Aynardi et al. conducted an observational, case-control study to compare the outcomes of patients undergoing THA in an inpatient and outpatient setting and found no differences in complications between groups (Aynardi et al, 2014).

Other studies still report differing findings (Bordoni et al, 2020). The lack of consensus about outcomes associated with TJA procedures performed in hospital inpatient versus outpatient settings suggests the need for further analyses; indeed, many of these studies have stressed the need for future analyses to confirm past findings (Xu et al, 2019; Lopez et al, 2020; Bemelmans et al, 2021), providing further support for the respecified IP/OP THA/TKA Complication measure, intended for use in a To-Be-Determined CMMI Payment Model that includes THA and TKA procedures performed in both the inpatient and outpatient settings. Re-specifying the existing Hospital-level THA/TKA Complication measure for use in a To-Be-Determined CMMI combined inpatient and outpatient payment model has many benefits, including incentivizing coordination of care across care settings and clinicians (especially for longer-term outcomes after elective procedures); incentivizing quality improvement among providers; and improving health for patients served by inpatient and outpatient facilities.

Importantly, the re-specified Hospital-level THA/TKA Complication measure will serve as a critical surveillance tool to help ensure that the cost reductions associated with outpatient TJA procedures do not adversely impact care.

References

1. Arshi, A., et al., Outpatient total hip arthroplasty in the United States: A population-based comparative analysis of complication rates. J Am Acad Orthop Surg, 2019. 27(2): p. 61-67.
2. Aynardi, M., et al., Outpatient surgery as a means of cost reduction in total hip arthroplasty: A case-control study. HSS J, 2014. 10(3): p. 252-5.
3. Bemelmans, Y.F.L., et al., Safety and efficacy of outpatient hip and knee arthroplasty: a systematic review with meta-analysis. Arch Orthop Trauma Surg, 2021.
4. Bert, J.M., J. Hooper, and S. Moen, Outpatient Total Joint Arthroplasty. Curr Rev Musculoskelet Med, 2017. 10(4): p. 567-574.
5. Bordoni, V., et al., Outpatient total knee arthroplasty leads to a higher number of complications: a meta-analysis. J Orthop Surg Res, 2020. 15(1): p. 408.
6. Darrith, B., et al., Inpatient Versus Outpatient Arthroplasty: A Single-Surgeon, Matched Cohort Analysis of 90-Day Complications. J Arthroplasty, 2019. 34(2): p. 221-227.
7. Goyal, N., et al., Otto Aufranc Award: A Multicenter, Randomized Study of Outpatient versus Inpatient Total Hip Arthroplasty. Clin Orthop Relat Res, 2017. 475(2): p. 364-372.
8. Greenky, M.R., et al., Total Hip Arthroplasty and the Medicare Inpatient-Only List: An Analysis of Complications in Medicare-Aged Patients Undergoing Outpatient Surgery. J Arthroplasty, 2019. 34(6): p. 1250-1254.
9. Husni, M.E., et al., Decreasing medical complications for total knee arthroplasty: Effect of critical pathways on outcomes. BMC Musculoskelet Disord, 2010. 11: p. 160.
10. Kimball, C.C., C.I. Nichols, and J.G. Vose, Outpatient Versus Rapid Recovery Inpatient Knee Arthroplasty: Comparison of Matched Cohorts. Orthopedics, 2020. 43(1): p. 36-41.
11. Lan, R.H., et al., Contemporary Outpatient Arthroplasty Is Safe Compared with Inpatient Surgery: A Propensity Score-Matched Analysis of 574,375 Procedures. J Bone Joint Surg Am, 2021. 103(7): p. 593-600.
12. Li, J., L.E. Rubin, and E.R. Mariano, Essential elements of an outpatient total joint replacement programme. Current Opinion in Anesthesiology, 2019. 32(5): p. 643-648.
13. Lopez, C.D., et al., Hospital and Surgeon Medicare Reimbursement Trends for Total Joint Arthroplasty. Arthroplasty Today, 2020. 6(3): p. 437-444.
14. Mariorenzi, M., et al., Outpatient Total Joint Arthroplasty: A Review of the Current Stance and Future Direction. R I Med J (2013), 2020. 103(3): p. 63-67.
15. Migliorini F, C.L., Cuozzo F, Oliva F, Valerio Marino A, Maffulli N, Outpatient Total Hip Arthroplasty: A Meta-Analysis. Applied Sciences, 2021. 11(15): p. 6853-6864.
16. Xu, J., et al., Comparison of outpatient versus inpatient total hip and knee arthroplasty: A systematic review and meta-analysis of complications. Journal of Orthopaedics, 2019. 17: p. 38-43.

[Response Ends]

1b. Gap in Care/Opportunity for Improvement and Disparities

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

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

[Response Begins]

The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level risk standardized complication rates (RSCRs) following elective primary THA and TKA occurring in hospital-based inpatient and outpatient settings. This measure is intended for use in a To-Be-Determined CMMI Payment Model that includes THA and TKA procedures performed in both the inpatient and outpatient settings. This will allow more complete and uniform performance measurement of hospitals performing THA and TKA procedures and the patients they serve.

Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process of care measures. 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, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk adjust for patients' conditions at the time of hospital admission or outpatient procedure and then evaluate patient outcomes. This measure is being developed to identify institutions whose performance is better or worse than would be expected based on each institution's patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Measuring and reporting RSCRs that include procedures performed in both the in patient and outpatient hospital setting will help inform health care providers about opportunities to improve care, strengthen incentives for quality improvement, and promote improvements in the quality of care received by patients and the outcomes they experience. The measure will also increase transparency by providing patients with information to help guide their choices about where they seek care for these elective procedures.

Complications in the outpatient hospital setting are more likely to be influenced by care coordination, a focus of potential episode payment programs. This measure will serve as a critical surveillance tool to help ensure cost reductions (e.g., reducing use of costly inpatient rehabilitation and/or shifting patients to the outpatie nt setting) do not adversely impact care.

[Response Ends]

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

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

[Response Begins]

Variation in complication rates indicates opportunity for improvement. We conducted analyses using Medicare Fee -for-service (FFS) claims data from THA and TKA procedures performed between April 1, 2018, and March 31, 2021

(n=989,587 procedures from 3,452 hospitals). These data included: 36 months of inpatient THA/TKA procedures (Apr 2018-Mar 2021); 36 months of outpatient TKA procedures (Apr 2018-Mar 2021); and 15 months of outpatient THA procedures (Jan 2020-Mar 2021).

*	Volume	Volume	Volume	Volume	RSCR (%)	RSCR (%)
*	**All hospitals**	**All Hospitals**	**Hospitals >= 25**	**Hospitals >= 25**	**All hospitals**	**Hospitals >= 25**
*	No. of procedures	% OP	No. of procedures	% OP	**All Hospitals**	**Hospitals >=25**
No. of hospitals	3,452	3,452	2,747	2,747	3,452	2,747
No. of index procedures	989,587	331,184	982,384	329,338	989,587	982,384
Mean	286.67	30.29	357.62	31.28	2.91	2.91
SD	429.21	26.34	454.80	23.64	0.47	0.52
Max	7,990	100.00	7,990	100.00	5.86	5.86
99%	2,015	100.00	2,158	86.05	4.37	4.41
95%	1,064	76.83	1,180	72.37	3.76	3.83
90%	741	67.31	855	64.96	3.49	3.56
75%	373	48.83	462	49.32	3.12	3.18
Median	128	26.09	205	28.88	2.86	2.85
25%	34	5.64	79	9.93	2.65	2.59
10%	9	0.00	43	0.99	2.40	2.34
5%	4	0.00	32	0.00	2.22	2.16
1%	1	0.00	25	0.00	1.87	1.80
Min	1	0.00	25	0.00	1.53	1.53

Table 1. Distribution of Hospital THA/TKA Volumes and RSCRs

*Cells intentionally left blank

[Response Ends]

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

[Response Begins]

This is a new measure that is being submitted for initial endorsement and has not yet been implemented; thus, limited performance data exists.

However, results from the existing Hospital THA/TKA Complication measure—Hospital-level Risk-standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (NQF #1550)—show variation in complication rates based on inpatient procedures. These results indicate an opportunity for improvement. Analyses conducted using data from April 1, 2018, to October 2, 2019 and July 1, 2020 to March 31,

2021 Medicare administrative claims data (n= 542,093 admissions from 3,445 hospitals) demonstrated a median risk-standardized complication rate of 2.4%, and range from 1.2%-5.6% in the measure cohort (Bernheim et al, 2022).

A review of the literature on THA/TKA complication rates also suggests that there is an opportunity for improvement. Complications following elective TJA procedures are rare, but the results can be devastating. Evidence shows that periprosthetic joint infection rates following THA and TKA range from 0.7% to 1.6% depending upon the population (Kurtz et al, 2010; Bozic et al, 2014). Reported 30- and 90-day death rates following THA range from 0.4% to 0.7% (Bozic et al, 2014; Soohoo et al, 2010). Reported 30-day death rates following TKA range from 0.1% to 0.3% in a Medicare population (Courtney et al, 2018). Rates for pulmonary embolism following THA range from 0.5% to 1.22% (Arshi et al, 2019) and range from 0.5% to 0.9% (Bozic et al, 2014; Khatod et al, 2008) following TKA. Rates for wound infection in Medicare population-based studies vary between 0.21% and 1.0% (Bozic et al, 2014; Soohoo et al, 2010; Browne et al, 2010). Rates for sepsis/septicemia range from 0.09%, during the index admission to 0.3% 90 days following discharge for primary TKA (Bozic et al, 2014; Browne et al, 2010). Rates for bleeding and hematoma following TKA range from 0.94% to 1.7% (Huddleston et al, 2009).

Studies have arrived at differing conclusions regarding complication rates of TJA procedures performed in the inpatient versus outpatient setting. Some studies have shown lower complication rates after outpatient procedures compared to inpatient procedures, suggesting that the outpatient setting represents a productive alternative in select patient populations (Kimball et al, 2020; Greenky et al, 2019; Lan et al, 2021). For example, Lan et al. found outpatient procedures were associated with lower rates of adverse events, with no increase in the rate of 30-day readmissions when compared to risk-matched inpatient procedures (Lan et al, 2021).

Other studies have found outpatient procedures to have comparable complication rates to inpatient procedures (Arshi et al, 2019; Xu et al, 2019; Aynardi et al, 2014; Bert et al, 2017; Goyal et al, 2017; Darrith et al, 2019; Migliorini et al, 2021; Mariorenzi et al, 2020). For example, Aynardi, et. al. conducted an observational, case-control study to compare the outcomes of patients undergoing THA in an inpatient and outpatient setting and found no differences in complications between groups (Aynardi et al, 2014).

Finally, findings from several studies suggest that complications after THA/TKA procedures are preventable. Husni et al. examined the influence of critical pathways on the postoperative outcomes after total knee replacement and found that patients operated upon in hospitals that used critical pathways had a lower risk of postoperative complications including death compared to patients operated upon in hospitals without pathways (Husni et al, 2010). Researchers also found that patients on critical pathways had significantly lower odds of experiencing any adverse event (adjusted OR = 0.68, 95% CI 0.50, 0.92) compared to patients without a critical pathway, and had an average length of stay that was 0.5 days shorter, after controlling for patient- and hospital-level factors (Husni et al, 2010). Similarly, in their literature review, Bert, et. al. summarized methods for establishing a successful outpatient TJA program within the framework of bundled payment reimbursement (Bert et al, 2017). Researchers outline several patient- and provider-level strategies for minimizing postsurgical complications and improving patient recovery, including patient risk assessment, care coordination, staff training, and the use of proper sterilization and pain management techniques. This would suggest that prevention of postsurgical complications is possible and that some prevention methods are within the providers control (Suter et al, 2020).

References

1. Arshi, A., et al., *Outpatient total hip arthroplasty in the United States: A population-based comparative analysis of complication rates.* J Am Acad Orthop Surg, 2019. **27**(2): p. 61-67.
2. Aynardi, M., et al., *Outpatient surgery as a means of cost reduction in total hip arthroplasty: A case-control study.* HSS J, 2014. **10**(3): p. 252-5.
3. Bernheim, S., J.N. Grady, and J.e.a. Debuhr, *Procedure Specific Complication Measure Updates and Specifications Report: Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) Risk-Standardized Complication Measure (Version 11.0).* 2022.
4. Bert, J.M., J. Hooper, and S. Moen, *Outpatient Total Joint Arthroplasty.* Curr Rev Musculoskelet Med, 2017. **10**(4): p. 567-574.

5. Bozic, K.J., et al., *Variation on hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty.* J Bone Joint Surg Am, 2014. **96**(8): p. 640-647.
6. Browne, J.A., et al., *Postoperative morbidity and mortality following total knee arthroplasty with computer navigation.* Knee, 2010. **17**(2): p. 152-6.
7. Courtney, P.M., et al., *Can Total Knee Arthroplasty Be Performed Safely as an Outpatient in the Medicare Population?* J Arthroplasty, 2018. **33**(7S): p. S28-S31.
8. Darrith, B., et al., *Inpatient Versus Outpatient Arthroplasty: A Single-Surgeon, Matched Cohort Analysis of 90-Day Complications.* J Arthroplasty, 2019. **34**(2): p. 221-227.
9. Goyal, N., et al., *Otto Aufranc Award: A Multicenter, Randomized Study of Outpatient versus Inpatient Total Hip Arthroplasty.* Clin Orthop Relat Res, 2017. **475**(2): p. 364-372.
10. Greenky, M.R., et al., *Total Hip Arthroplasty and the Medicare Inpatient-Only List: An Analysis of Complications in Medicare-Aged Patients Undergoing Outpatient Surgery.* J Arthroplasty, 2019. **34**(6): p. 1250-1254.
11. Huddleston, J.I., et al., *Adverse events after total knee arthroplasty: A national Medicare study.* J Arthroplasty, 2009. **24**(6 Suppl): p. 95-100.
12. Husni, M.E., et al., *Decreasing medical complications for total knee arthroplasty: Effect of critical pathways on outcomes.* BMC Musculoskelet Disord, 2010. **11**: p. 160.
13. Khatod, M., et al., *Knee replacement: Epidemiology, outcomes, and trends in Southern California: 17,080 replacements from 1995 through 2004.* Acta Orthop, 2008. **79**(6): p. 812-9.
14. Kimball, C.C., C.I. Nichols, and J.G. Vose, *Outpatient Versus Rapid Recovery Inpatient Knee Arthroplasty: Comparison of Matched Cohorts.* Orthopedics, 2020. **43**(1): p. 36-41.
15. Kurtz, S.M., et al., *Prosthetic joint infection risk after TKA in the Medicare population.* Clin Orthop Relat Res, 2010. **468**(1): p. 52-6.
16. Lan, R.H., et al., *Contemporary Outpatient Arthroplasty Is Safe Compared with Inpatient Surgery: A Propensity Score-Matched Analysis of 574,375 Procedures.* J Bone Joint Surg Am, 2021. **103**(7): p. 593-600.
17. Mariorenzi, M., et al., *Outpatient Total Joint Arthroplasty: A Review of the Current Stance and Future Direction.* R I Med J (2013), 2020. **103**(3): p. 63-67.
18. Migliorini F, C.L., Cuozzo F, Oliva F, Valerio Marino A, Maffulli N, *Outpatient Total Hip Arthroplasty: A Meta-Analysis.* Applied Sciences, 2021. **11**(15): p. 6853-6864.
19. Soohoo, N.F., et al., *Factors that predict short-term complication rates after total hip arthroplasty.* Clin Orthop Relat Res, 2010. **468**(9): p. 2363-71.
20. Suter, L.G., et al., *90-Day Risk-Standardized Complication Rates Following Elective Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for a Potential Combined Inpatient and Outpatient Episode Payment Model (EPM) Business Case*. 2020.
21. Xu, J., et al., *Comparison of outpatient versus inpatient total hip and knee arthroplasty: A systematic review and meta-analysis of complications.* Journal of Orthopaedics, 2019. **17**: p. 38-43.

[Response Ends]

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

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for

improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

Distribution of THA/TKA RSCRs by Proportion of Non-White Race, Dual Eligible Patients, and low ASI Score

Data Source: Medicare FFS claims and enrollment data

Dates of Data: April 1, 2018 to March 31, 2021

Table 2 shows a comparison of the distribution of hospital-level RSCRs between hospitals with a lower proportion of Black beneficiaries (first quartile of ascending-ordered proportions) and hospitals with a higher proportion of Black beneficiaries (fourth quartile of ascending-ordered proportions) along with the distribution of the overall proportion of black patients at the hospital level. Among the 2,733 hospitals with at least 25 index procedures, the mean RSCR of hospitals with a lower proportion of Black beneficiaries is slightly lower than the mean of hospitals with a higher proportion of Black beneficiaries (2.99% vs 3.02%).

*	*	RSCR by quartile of proportion of black patients	RSCR by quartile of proportion of black patients
*	% Patients who are Black	Q1 range of %Black (0-0):	Q4 range of %Black (5.92-90.32):
No. of hospitals N	2,733	704	683
Mean	4.94	2.99	3.02
SD	8.69	0.44	0.52
Max	90.32	5.35	5.07
99%	42.10	4.37	4.64
95%	20.29	3.76	3.95
90%	12.73	3.54	3.68
75%	5.92	3.20	3.29
Median	1.90	2.92	2.96
25%	0	2.69	2.67
10%	0	2.52	2.48
5%	0	2.41	2.27
1%	0	2.19	1.87
Min	0	2.02	1.64

Table 2. Risk-Standardized Complication Rate (RSCR) by Black race only including hospitals with at least 25 cases

*Cells intentionally left blank

Table 3 shows a comparison of the distribution of hospital-level RSCRs between hospitals with a lower proportion of Other non-White (Asian, Hispanic, North American Native, and Other) beneficiaries (first quartile of ascending-ordered proportions) and hospitals with a higher proportion of other non-White beneficiaries (fourth quartile of ascending-

ordered proportions) along with the distribution of the overall proportion of patients of other non-white races at the hospital level. Among the 2,733 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer Other non-White beneficiaries combined is lower than the mean of hospitals with more Other non-White beneficiaries (2.95% vs 2.99%).

*	*	**RSCR by quartile of proportion of other non-White combined**	**RSCR by quartile of proportion of other non-White combined**
*	**% Patients who are Other non-White races**	**Q1 range of %other (0-0.54):**	**Q4 range of %other (3.36-100):**
No. of hospitals N	2,733	683	684
Mean	3.56	2.95	2.99
SD	7.52	0.45	0.53
Max	100	5.35	6.00
99%	41.67	4.33	4.64
95%	12.5	3.67	3.95
90%	8	3.52	3.69
75%	3.36	3.19	3.25
Median	1.52	2.90	2.92
25%	0.54	2.65	2.66
10%	0	2.50	2.44
5%	0	2.33	2.23
1%	0	2.02	1.98
Min	0	1.64	1.58

Table 3. Risk-Standardized Complication Rate (RSCR) by Other non-White race combined only including hospitals with at least 25 cases

*Cells intentionally left blank

Table 4 shows the difference of hospital-level RSCR distribution between hospitals with a lower proportion of dually eligible beneficiaries (first quartile of ascending-ordered proportions) and hospitals with a higher proportion of dually eligible beneficiaries (fourth quartile of ascending-ordered proportions) along with the distribution of the overall proportion of dually eligible patients at the hospital level. Among all 2,747 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer dually eligible beneficiaries is lower (2.86% vs 3.02%).

*	*	**RSCR by quartile of proportion of dual**	**RSCR by quartile of proportion of dual**
Quantile	**% Patients who are dually eligible**	**Q1 range of %dual (0-1.33):**	**Q4 range of %dual (5.79-94):**
No. of hospitals N	2,747	684	687

*	*	RSCR by quartile of proportion of dual	RSCR by quartile of proportion of dual
Mean	5.22	2.86	3.02
SD	8.62	0.50	0.53
Max	94.00	5.29	5.86
99%	47.83	4.23	4.81
95%	17.19	3.72	4.04
90%	11.11	3.46	3.70
75%	5.79	3.14	3.24
Median	2.82	2.83	2.92
25%	1.33	2.54	2.67
10%	0.44	2.28	2.50
5%	0	2.06	2.34
1%	0	1.75	2.07
Min	0	1.53	1.67

Table 4. Risk-Standardized Complication Rate (RSCR) by dual eligibility only including hospitals with at least 25 cases

*Cells intentionally left empty.

Table 5 shows the difference of RSCR distribution between hospitals with a lower proportion of beneficiaries with lower AHRQ SES Index (ASI; first quartile of ascending-ordered proportions) and hospitals with a higher proportion of beneficiaries with lower ASI (fourth quartile of ascending-ordered proportions) along with the distribution of the overall proportion of patients who have a low ASI at the hospital level. Among the 2,740 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer beneficiaries with a lower ASI is lower compared to hospitals with a higher proportion of beneficiaries with a higher ASI (2.85% vs 3.00%).

*	*	RSCR by quartile of proportion of low ASI	RSCR by quartile of proportion of low ASI
Quantile	**% Patient who have a low ASI**	**Q1 range of %dual (0-4.80):**	**Q4 range of %dual (20-81.08):**
No. of hospitals N	2,740	684	686
Mean	14.48	2.85	3.00
SD	13.53	0.52	0.48
Max	81.08	5.70	5.40
99%	61.70	4.41	4.40
95%	42.80	3.81	3.84
90%	33.94	3.50	3.61

*	*	RSCR by quartile of proportion of low ASI	RSCR by quartile of proportion of low ASI
75%	20	3.10	3.25
Median	10.21	2.79	2.94
25%	4.81	2.53	2.69
10%	2.17	2.28	2.49
5%	1.11	2.15	2.31
1%	0	1.70	1.99
Min	0	1.54	1.70

Table 5. Complication rate by ASI only including hospitals with at least 25 cases

*Cells intentionally left empty.

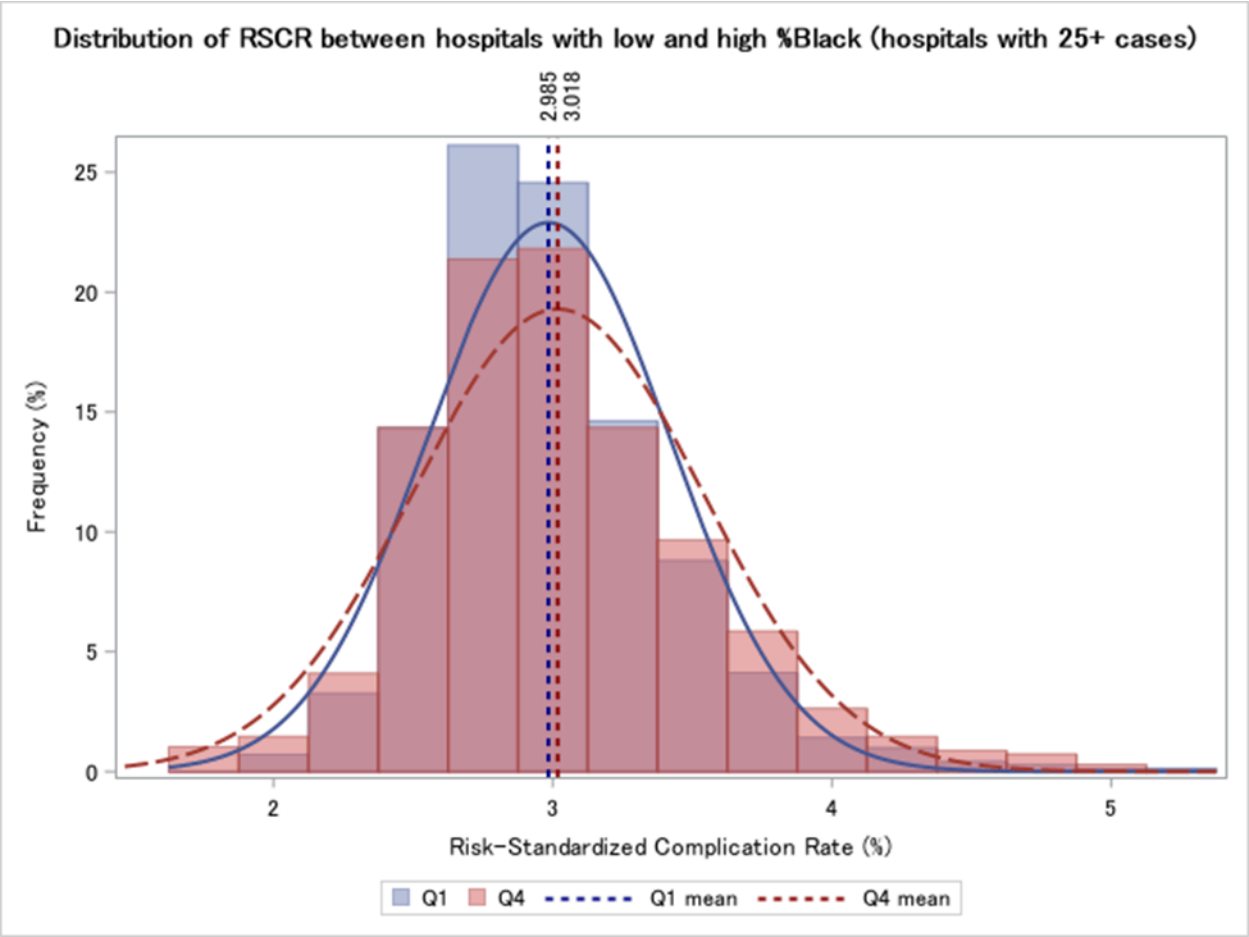

Figure 2. Distribution of RSCR between hospitals with low and high %Black (hospitals with 25+ cases)

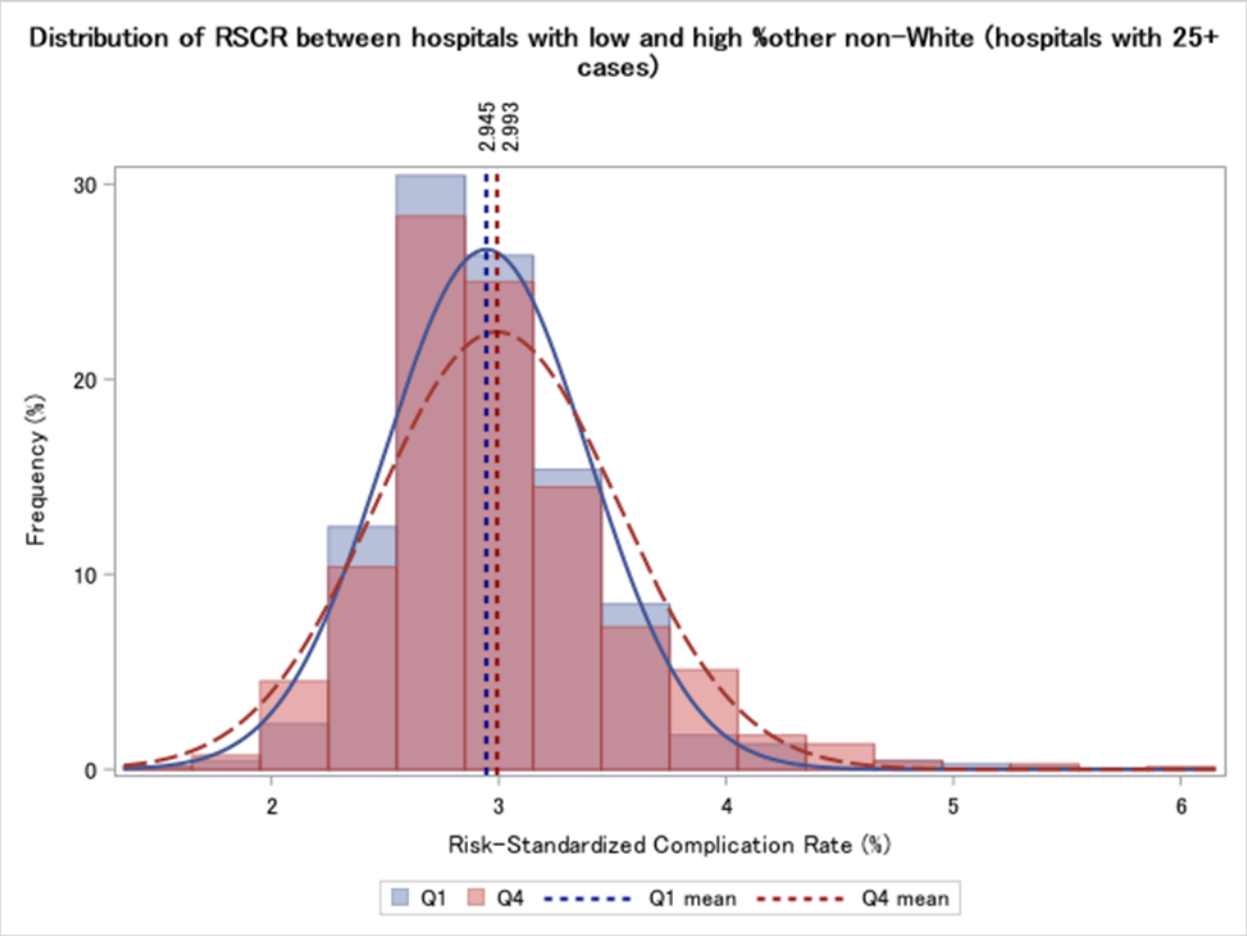

Figure 3. Distribution of RSCR between hospitals with low and high % other non-White (hospitals with 25+ cases)

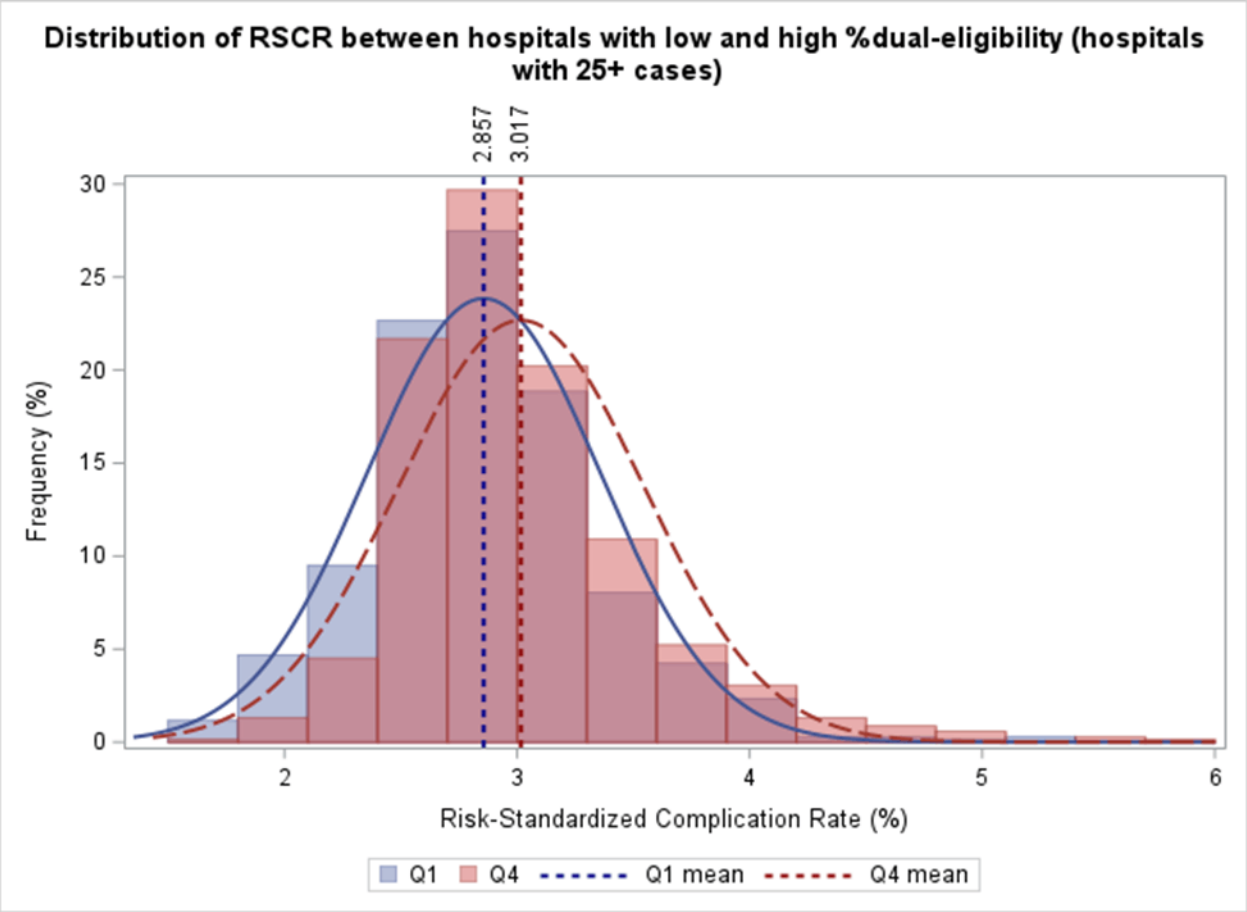

Figure 4. Distribution of RSCR between hospitals with low and high % dual eligibility (hospitals with 25+ cases)

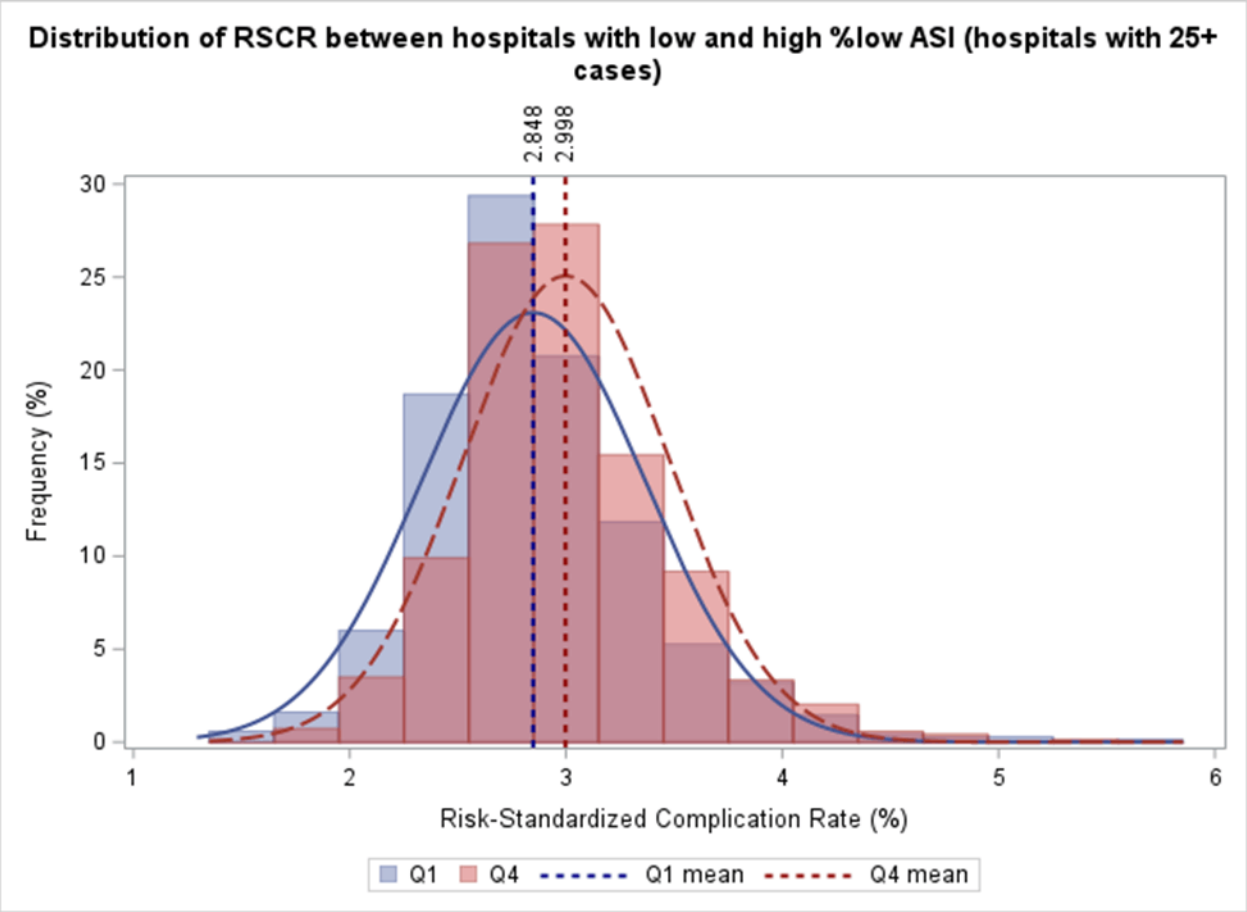

Figure 5. Distribution of RSCR between hospitals with low and high % ASI (hospitals with 25+ cases)

[Response Ends]

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

[Response Begins]

N/A.

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

sp.01. Provide the measure title.

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

[Response Begins]

Hospital-Level 90-Day Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for a Combined Inpatient (IP) and Outpatient (OP) Setting (IP/OP 90-Day THA/TKA Complication Measure)

[Response Ends]

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

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

[Response Begins]

The measure estimates a hospital-level risk-standardized complication rate (RSCR) associated with elective primary THA and TKA procedures performed in hospital inpatient and outpatient departments on Medicare fee-for-service (FFS) beneficiaries who are 65 years of age or older. The outcome (complication) is defined as any one of the specified complications occurring within 90 days of the index encounter. More details on the outcome are provided in sp.13.

Note: An index encounter is defined as either an inpatient admission for a THA/TKA procedure or a hospital outpatient department THA/TKA procedure to which the complication outcome is attributed.

[Response Ends]

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

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

Please do not select:

- *Surgery: General*

[Response Begins]

Musculoskeletal: Joint Surgery

Musculoskeletal: Osteoarthritis

Musculoskeletal: Rheumatoid Arthritis

Surgery: Orthopedic

[Response Ends]

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

[Response Begins]

Care Coordination

Care Coordination: Readmissions

Care Coordination: Transitions of Care

Disparities Sensitive

Safety: Complications

Safety: Healthcare Associated Infections

[Response Ends]

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

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

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

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Elderly (Age >= 65)

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Facility

[Response Ends]

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

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

[Response Begins]

Inpatient/Hospital

Outpatient Services

[Response Ends]

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

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

[Response Begins]

None available; the measure methodology report is attached to this submission.

[Response Ends]

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

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

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3732_3732_3732_IP.OP THA_TKA Complication_CodeSetFileforNQF.6.27.22_(1)-508.xlsx

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

sp.13. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]

The outcome for this measure is a dichotomous assessment of whether a complication occurred ("yes" for any complication[s]; "no" for no complications) within 90 days of the index encounter (see sp.14 for Numerator Details). Complications are counted in the measure outcome only if they occur during the index inpatient or outpatient encounter or are captured in claims associated with an inpatient readmission, observation stay, emergency department visit, or ambulatory surgical center (ASC) encounter.

[Response Ends]

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

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

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

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

[Response Begins]

The composite complication is a dichotomous outcome ("yes" for any complication[s]; "no" for no complications). If a patient experiences one or more of the following complications within the specified time frame for each complication (all within 90 days of the index encounter), the outcome variable will be coded as a "yes." Complications are counted in the measure only if they occur during the index encounter or are captured in claims associated with an inpatient readmission, observation stay, emergency department visit, or ambulatory surgical center (ASC) encounter, as specified below.

The measure outcome is defined as any one of the complications listed below occurring within the time-period specified for that complication:

- Acute myocardial infarction (AMI) during the index encounter or a subsequent inpatient admission (only) that occurs within 7 days from the start of the index encounter.
- Pneumonia or other acute respiratory complication during the index encounter or a subsequent inpatient admission (only) that occurs within 7 days from the start of the index encounter.
- Sepsis/septicemia/shock during the index encounter or a subsequent inpatient admission (only) that occurs within 7 days from the start of the index encounter.
- Surgical site bleeding or other surgical site complication during the index encounter or a subsequent inpatient admission or observation stay or ASC encounter that occurs within 30 days from the start of the index encounter.
- Pulmonary embolism during the index encounter or a subsequent inpatient admission or observation stay that occurs within 30 days from the start of the index encounter.
- Death during the index encounter or within 30 days from the start of the index encounter.
- Mechanical complication during the index encounter or a subsequent inpatient admission or observation stay or emergency department or ASC encounter that occurs within 90 days from the start of the index encounter.
- Periprosthetic joint infection/wound infection or other wound complication during the index encounter or a subsequent inpatient admission or observation stay or ASC encounter that occurs within 90 days from the start of the index encounter.

[Response Ends]

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

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

The target population for the measure includes Medicare fee-for-service (FFS) beneficiaries who are 65 years of age and older undergoing elective primary THA and/or TKA procedures in the hospital inpatient and outpatient settings.

Additional details are provided in sp.16 Denominator Details.

[Response Ends]

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

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

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

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

[Response Begins]

To be included in the measure cohort, patients must meet the following inclusion criteria:

1. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to and including the date of encounter;

2. Aged 65 years or older.

3. Having a qualifying elective primary THA/TKA procedure; elective primary THA/TKA procedures are defined as those procedures without any of the following:

- Fracture of the pelvis or lower limbs coded as present on admission [POA] in the principal or secondary discharge diagnosis fields on the index encounter (Note: This criterion is not applicable to periprosthetic fractures for procedure performed in the outpatient setting).
- A concurrent partial hip or knee arthroplasty procedure.
- A concurrent revision, resurfacing, or implanted device/prosthesis removal procedure.
- Mechanical complication coded in the principal discharge diagnosis field on the index encounter.
- 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 on the index encounter; or,
- Transfer from another acute care facility for the THA/TKA.

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

This measure excludes index encounters for the following patients:

1. Without at least 90 days post-discharge enrollment in Medicare FFS.

   Rationale: The 90-day complication outcome cannot be assessed in this group since claims data are used to determine whether a complication of care occurred.

2. Who were discharged against medical advice (AMA)

   Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.

3. Who had more than two THA/TKA procedure codes during the index encounter.

   Rationale: Although clinically possible, it is highly unlikely patients would receive more than two elective THA/TKA procedures in one hospitalization, which may reflect a coding error.

After applying these exclusion criteria, we randomly select one index encounter for patients who have multiple index encounters in a calendar year. We therefore exclude the eligible index encounters that are not randomly selected in that calendar year.

[Response Ends]

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

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

[Response Begins]

We identified enrollment in Medicare FFS as enrollment in Parts A & B for 3 months after index encounter. We identified AMA using Patient Discharge Status Code = 07 (Left against medical advice or discontinued care.) We identified THA/TKA procedures using cohort inclusion/exclusion codes set (see attached code set file).

[Response Ends]

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

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

[Response Begins]

Implementation planning, including potential stratification, has not been finalized.

[Response Ends]

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

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

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

[Response Begins]

Statistical risk model

[Response Ends]

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

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

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

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

[Response Begins]

Better quality = Lower score

[Response Ends]

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

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

[Response Begins]

The measure will estimate RSCRs following elective primary THA/TKA using hierarchical logistic regression models in the hospital-based inpatient and outpatient settings.

In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of a complication occurring within the specified time period (up to 90 days) after the index encounter using age, sex, selected clinical covariates, hospital setting, and a hospital-specific effect. At the hospital level, it models the hospital-specific effect as arising from a normal distribution. The hospital effect represents the underlying risk of a complication at the hospital, after accounting for patient risk. The hospital-specific effects are given a distribution to account for the clustering (non-independence) of patients within the same hospital (Normand and Shahian, 2007). If there were no differences among hospitals, then after adjusting for patient risk, the hospital effects should be identical across all hospitals.

The RSCR is calculated as the ratio of the number of "predicted" to the number of "expected" encounters with a complication at a given hospital, multiplied by the national observed complication rate. For each hospital, the numerator of the ratio is the number of complications within the specified time period (up to 90 days) predicted on the basis of the hospital's performance with its observed case mix, and the denominator is the number of complications expected based on the nation's performance with that hospital's case mix. 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 hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected complication rates or better quality, and a higher ratio indicates higher-than-expected complication rates or worse quality.

The "predicted" number of encounters with a complication (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific effect on the risk of having an encounter with a complication. The estimated hospital-specific effect is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are log transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of encounters with a complication (the denominator) is obtained in the same manner, but a common effect using all hospitals in our sample is added in place of the hospital-specific effect. The results are log transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we will re-estimate the model coefficients using the years of data in that period.

The ratio of predicted over expected is transformed into a rate by multiplying by the national observed complication rate. The hierarchical logistic regression models are described fully in the original measure methodology report for the inpatient-only version of this measure (Grosso et al, 2012).

References

1. Grosso L, Curtis J, Geary L, et al. Hospital-Level Risk-Standardized Complication Rate Following Elective Primary Total Hip Arthroplasty (THA) And/Or Total Knee Arthroplasty (TKA) Measure Methodology Report. 2012.
2. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

[Response Ends]

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

Examples of samples used for testing:

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

[Response Begins]

N/A. This measure is not based on a sample.

[Response Ends]

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

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Medicare Enrollment Database

[Response Ends]

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

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

[Response Begins]

The measure uses the following data sources:

1. Medicare inpatient, outpatient, and physician/professional claims: These include data for Medicare FFS inpatient and outpatient services such as Medicare inpatient hospital care, HOPD services, and physician claims for the 12 months prior to an index encounter and for the three months after (Triche et al, 2020; Fleming et al, 1992).
2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source is used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission, vital status at discharge, and death information post-discharge. These data have previously been shown to accurately reflect patient vital status (Fleming et al, 1992).

References

1. Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. Medical Care. 1992; 30(5): 377-91.

2. Triche E, Grady JN, Debuhr D et al. 2020 Procedure Specific Complication Measure Updates and Specifications Report: Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) Risk-Standardized Complication Measure (Version 9.0). 2020.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

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

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

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

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

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

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

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

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

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

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

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

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

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

Definitions

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

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

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

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

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

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e .g.,

$5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Medicare Enrollment Database (EDB); American Community Survey data; CMS Overall Hospital Star Ratings July 2021 Data

[Response Ends]

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

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

[Response Begins]

Medicare Fee-For-Service (FFS) inpatient, outpatient, and physician/professional claims, as well as enrollment data were used for testing. Dual eligibility was obtained through the state Medicare Modernization Act (MMA) files. The Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score was obtained from American Community Survey (ACS) data. We used the ACS (2013-2017) data to derive an updated AHRQ SES index score at the patient nine-digit zip code level for use in studying the association between our measure and social risk factors (SRFs). We used CMS Overall Hospital Star Rating Data (July 2021 Refresh) for empiric validity testing.

[Response Ends]

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

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

[Response Begins]

04-01-2018 – 03-31-2021. These data included: 36 months of inpatient THA/TKA procedures (04-01-2018 - 03-31-2021); 36 months of outpatient TKA procedures (04-01-2018 - 03-31-2021); and 15 months of outpatient THA procedures (01-01-2020 - 03-31-2021).

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Facility

[Response Ends]

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

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

[Response Begins]

For this measure, hospitals are the measured entities. All non-federal, acute care US hospitals (including territories) providing care to Medicare FFS beneficiaries aged 65 years and older are included. Our analyses included 3,452 hospitals, and 2,747 of those hospitals had at least 25 cases (the likely reporting threshold).

[Response Ends]

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

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

[Response Begins]

We included all clinically eligible procedures in the analysis. See Table 1 for details.

Demographics	**Respecification Dataset (April 1, 2018 to March 31, 2021**	*	*
*	**Inpatient THA/TKA n(%)**	**Outpatient THA/TKA n(%)**	**Both n(%)**
Total procedures[1]	658,403	331,184	989,587
Number of hospitals	3,363	2,912	3,452
Age distribution (years) - Mean (SD)	74 (6)	73 (5)	74 (6)

Demographics	Respecification Dataset (April 1, 2018 to March 31, 2021	*	*
Male	242,997 (36.91)	135,807 (41.01)	378,804 (38.28)
Race/Ethnicity - Non-White	48,456 (9.44)	20,813 (8.96)	69,269 (9.27)

Table 1. Demographic Information for Respecification Dataset (April 1, 2018 to March 31, 2021)

[1]The number of total procedures is smaller than reported elsewhere due to the random selection of one procedure per patient per year.

*cells intentionally left blank

[Response Ends]

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

[Response Begins]

There were 989,587 total THA/TKA procedures that met the measure inclusion/exclusion criteria. Of these, 67% (n=658,403) were inpatient, and 33% (n=331,184) were HOPD.

For this measure, hospitals are the measured entities. All non-federal, acute inpatient United States (US) hospitals (including territories) with Medicare FFS beneficiaries aged 65 years and older are included. All eligible procedures were included in testing.

Measure Respecification

We used Medicare administrative claims data and enrollment information for patients with qualifying procedures between April 1, 2018 and March 31, 2021. Specifically, we used the following data sources:

1. Medicare inpatient, outpatient, and physician/professional claims: these include data for Medicare FFS inpatient and outpatient services such as Medicare inpatient hospital care, HOPD services, and physician claims for the 12 months prior to an index encounter and for the three months after.
2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source is used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission, vital status at discharge and death information post-discharge. These data have previously been shown to accurately reflect patient vital status.

This measure does not have any specific exclusions for COVID, as both empiric analyses and clinical expert input supported that COVID infections would be unlikely to significantly impact this measure assessing elective procedures performed primarily to relieve pain and improve function.

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are in Table 2.

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
Respecification and Testing Dataset (Medicare Fee-For-Service Administrative Claims Data)	Section 2a.10 Reliability Testing Section 2b.02 Validity Testing Section 2b.16 Testing of Measure Exclusion Section 2b.23 Risk Adjustment/Stratification Section 2b.26 Statistical Risk Model Discrimination Statistics Section 2b.28 Statistical Risk Model Calibration Statistics Section 2b.31 Meaningful Differences	Entire Cohort: Dates of Data: April 1, 2018 to March 31, 2021 Number of admissions/encounters = 989,587 Patient Descriptive Characteristics: mean age = 74 years; % male = 38.28% Number of measured hospitals: 3,452
The American Community Survey (ACS)	Section 2a.08: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: 2013-2017 We used the AHRQ SES index score derived from the American Community Survey (2013-2017) to study the association between the 90-day complication outcome and social risk factors (SRFs). The AHRQ SES index score is based on beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
Master Beneficiary Summary File (MBSF)	Section 2a.08: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: April 2018 – March 2021 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the 90-day measure outcome and dual-eligible status.

Table 2. Dataset Descriptions

[Response Ends]

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

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

[Response Begins]

Social risk factors available and analyzed included exposure to racism (as represented by White, Black or other non-White race), dual eligibility (dual Medicare and Medicaid coverage), and the AHRQ SES index. Please note: We do not consider race a marker of socioeconomic status; we include it in our social risk factor analyses based upon literature specifically

documenting racial and ethnic disparities in THA/TKA offer and acceptance rates as well as outcomes (Irgit and Nelson, 2011; Kerman et al, 2018).

We selected social risk factor variables to analyze after reviewing the literature and examining available national data sources. We sought to find variables that are consistently captured in a reliable fashion for all patients in this measure. Therefore, robust, validated proxy measures of exposure to racism, income, education level and economic status were selected and are described below. The conceptual model for how social risk variables are related to the outcome is described in section 2b.23.

The social risk factor variables used for analysis were:

- **Race, as a proxy for exposure to racism**: Race was categorized as White, Black, or Other Non-White; this grouping (of categories "Asian," "Hispanic," "North American Native," and "Other" into "Other Non-White"), was chosen due to the very low numbers of people in these categories undergoing THA/TKA in national Medicare Fee For Service claims data.

As noted above, we do not consider race a marker of socioeconomic status, but we include it here due to known racial and ethnic disparities in THA/TKA offer and acceptance rates as well as outcomes (Irgit and Nelson, 2011; Kerman et al, 2018). Because this measure assesses outcomes after an elective surgery, race can impact whether or not a patient can access surgical care, whether or not a patient is offered surgery once they are in the care of a surgeon, whether or not a patient accepts surgery once offered, what setting the patient undergoes surgery, and ultimately a patient's outcomes. Racial disparities in all steps in this complex process have been documented (Irgit and Nelson, 2011; Kerman et al, 2018).

- **Dual eligible status**: Dual eligible status (in other words, being enrolled in both Medicare and Medicaid) patient-level data is obtained from the CMS Master Beneficiary Summary File (MBSF)

Following guidance from Department of Health and Human Services Assistant Secretary for Policy and Evaluation (ASPE) and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable (ASPE 2016; ASPE 2020). We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous outcome. However, the threshold for over 65-year-old Medicare patients is valuable, as it takes into account both income and assets and is consistently applied across states for the older population. We acknowledge that it is important to test a wider variety of SRFs including key variables such as education and poverty level; therefore, we also tested a validated composite based on census data linked to as small a geographic unit as possible.

- **AHRQ-validated SES index score (ASI)** (summarizing the information from the following 7 variables): percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room)

We selected the AHRQ SES index score (ASI) because it is a well-validated variable that describes the average socioeconomic status of people living in defined geographic areas (Bonito et al, 2008). Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of social risk factors. We ultimately did not include the ADI at the time, partly due to the fact that the coefficients used to derive ADI had not been updated for many years. Recently, the coefficients for ADI have been updated and therefore we compared the ADI with the ASI and found them to be highly correlated.

In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the ASI were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low socioeconomic status neighborhoods in high expense geographic areas. We then calculated an ASI score for census block groups that can be linked to 9-digit ZIP codes.

References

1. Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;2.
2. U.S. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2019.
3. Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.
4. Irgit, K., & Nelson, C. L. (2011). Defining Racial and Ethnic Disparities in THA and TKA. Clinical Orthopaedics and Related Research®, 469(7), 1817–1823.
5. Kerman, H. M., Smith, S. R., Smith, K. C., Collins, J. E., Suter, L. G., Katz, J. N., & Losina, E. (2018). Disparities in Total Knee Replacement: Population Losses in Quality-Adjusted Life-Years Due to Differential Offer, Acceptance, and Complication Rates for African Americans. Arthritis Care & Research, 70(9), 1326–1334

[Response Ends]

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

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

Choose one or both levels.

[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

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

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

[Response Begins]

Measure Score Reliability

We estimated the facility-level reliability (signal-to-noise reliability).

We provide facility-level measure score reliability using the signal-to-noise method, using the formula presented by Adams and colleagues (Adams et al, 2010; Yu et al, 2013). Facility-to-facility variance is estimated from the hierarchical logistic regression model, n is equal to each facility's observed case size, and the facility error variance is estimated using the variance of the logistic distribution (π^2/3). The facility-level reliability testing is limited to facilities with at least 25 admissions for public reporting.

Where facility-to-facility variance is estimated from the model, n is equal to each facility's observed case size, and the facility error variance is estimated using the variance of the logistic distribution.

Signal-to-noise reliability scores can range from 0 to 1. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real difference in performance.

We calculated the measure score reliability for all facilities, and for facilities with a volume cutoff of 25 procedures.

Additional Reliability and Validity Information

We provide additional reliability and validity information to provide a more complete picture of the overall measure score reliability and validity. However, as noted in NQF criteria for initial endorsement, only measure score reliability and validity testing are required.

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

In addition, CMS has in place several hospital auditing programs used to assess overall claims code 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 to payment.

Furthermore, we assessed the variation in the frequency of the variables over time: Detailed information is presented in the measure's 2020 Condition-Specific Measure Updates and Specifications Report cited below.

References

1. Adams J, Mehrotra, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.
2. DeBuhr J, McDowell, K, Grady J, et al., 2020 Condition-Specific Complication Measure Updates and Specifications Report - Available at: https://www.qualitynet.org/inpatient/measures/complication/methodology.
3. Yu, H, Mehrotra, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.

[Response Ends]

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

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

[Response Begins]

Measure Score-Level Signal-to-Noise Reliability Results

The median signal-to-noise reliability across all hospitals with at least one case of eligible THA/TKA procedure is 0.78 (IQR: 0.48 – 0.91). The median signal-to-noise reliability across hospitals with at least 25 cases, is 0.85 (IQR: 0.69 – 0.93). See Table 3 for details.

Distribution of signal-to-noise reliabilities	All hospitals	Hospitals volume >= 25
No. of hospitals (%)	3,452 (100.0%)	2,747 (79.6%)
No. of THA/TKA procedures captured (%)	989,587 (100.0%)	982,384 (99.3%)
No. of patients with complications captured (%)	28,493 (100.0%)	28,153 (98.8%)
Mean	0.674	0.795
Std Dev	0.283	0.159
Max	0.995	0.995
Q3	0.911	0.927
Median	0.779	0.849
Q1	0.480	0.685
Min	0.027	0.408

Table 3. Volume Thresholds Based on Signal-to-Noise Reliability and Distribution of Signal-to-Noise Reliabilities

[Response Ends]

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

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

[Response Begins]

Measure Score-Level Signal-to-Noise Reliability

Signal-to-noise reliability results for facilities with at least 25 procedures (median reliability of 0.849) show that this measure is sufficiently reliable for its intended use.

References

1. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling — reliability and risk of misclassification. *New England Journal of Medicine*. 2010;362(11):1014–1021.

[Response Ends]

2b. Validity

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

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[Response Ends]

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

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

[Response Begins]

Empiric Validity

We identified and assessed the measure's correlation with other measures that target the same domain of quality (e.g. complications, safety, or post-procedure utilization) for the same or similar populations. The goal was to identify whether better performance in this measure was related to better performance on other relevant structural or outcomes measures. After a literature review and consultations with measures experts in the field, there were very few measures identified that assess the same domains of quality. Given that challenge, we selected the following to use for validity testing:

1. **Overall Hospital Star Rating Summary Score**: CMS's Overall Hospital Star Rating assesses hospitals' overall performance based on a weighted average of "group scores" from different domains of quality (mortality, readmissions, safety, patient experience, timely and effective care). Each group is comprised of individual quality measures, which are publicly reported on *Care Compare*. Group scores for each individual group are calculated using a simple average of the (standardized) measure scores for the quality measures in each group. Group scores are combined into an overall hospital summary score using fixed weights. The full methodology for the Overall Hospital Star Rating can be found at: https://qualitynet.cms.gov/inpatient/public-reporting/overall-ratings/resources.

   The star rating summary score was chosen for validity testing because it measures the same population (Medicare FFS patients) and because the quality domains (groups) of the summary score, including the safety and readmissions domains, fall within the same causal pathway as the IP/OP 90-Day THA/TKA Complication measure (the measure currently under review). In fact, the inpatient-only THA/TKA Complications measure is part of the Safety of Care group score, but we have removed that measure from the validity analysis that we performed for this NQF submission.

   For this analysis, we examined hospital-performance on the IP/OP 90-Day THA/TKA Complication measure within quartiles of the star ratings summary score and calculated an overall Pearson's Correlation Coefficient. For this analysis we used Hospital Star Ratings summary score (July 2021) from 2,747 Medicare FFS hospitals with at least one eligible THA/TKA procedure and 2,358 Medicare FFS hospitals with at least 25 eligible procedures. We predicted the THA/TKA complication measure scores would have a weak association with the overall hospital star rating summary scores, with lower RSCRs associated with better summary scores.
2. **Hospital THA/TKA Surgical Volume**: There is evidence that surgical complication rates for providers (both surgeons and hospitals) decline with increasing volume (Sibley et al, 2017; Murphy et al, 2019; Courtney et al, 2018). Thus, we assessed validity of the measure by examining the relationship between volume and the measure score for hospitals. We predict that lower RSCRs will be moderately associated with higher volume hospitals.

Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input via two mechanisms: regular discussions with an advisory clinical working group and a nationally convened multidisciplinary Technical Expert Panel (TEP) to increase transparency and to gain broader input into the measure.

We assembled the clinical working group and held regular meetings throughout the development phase. The Clinical Working Group was tailored for development of this measure and consisted of clinicians nominated by the four professional societies addressing THA/TKA procedures and best practices in the US: American Academy of Orthopaedic

Surgeons (AAOS); American Association of Hip and Knee Surgeons (AAHKS); The Hip Society; and The Knee Society. All members were nominated by their respective societies based upon their clinical, health policy, and quality measurement/ methodological expertise. Working group meetings addressed key clinical and practice-related issues related to measure development, including weighing the pros and cons of and finalizing key decisions (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The Clinical Working Group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In addition to the working group, and in alignment with the CMS Measures Management System, we convened a multi-stakeholder TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives, including physicians, consumers, and purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held two structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members.

Face Validity

We assessed face validity by asking TEP members to rate the measure according to the following two statements using a six-point scale (1 = Strongly Agree, 2 = Moderately Agree, 3 = Somewhat Agree, 4 = Somewhat Disagree, 5 = Moderately Disagree, 6 = Strongly Disagree):

- Statement #1: IP/OP THA/TKA Complication measure as specified will provide a valid assessment of complications following elective THA/TKA
- Statement #2: IP/OP THA/TKA Complication measure as specified can be used to distinguish between better and worse quality care among hospitals performing THAs/TKAs

Additional Validity Information: Setting indicator

This measure includes patients undergoing procedures in both the inpatient and HOPD settings. Initial empiric analyses indicated that complication rates were lower for HOPD procedures compared to inpatient procedures despite only modest differences in clinical risk factor frequency across settings. Through discussions with internal clinical and health policy experts, the Clinical Working Group and the TEP, we concluded that there are multiple factors influencing the decision to perform elective THA/TKA procedures in the inpatient vs. HOPD setting. These factors include clinical risk assessment by the surgical team (including frailty), hospital policies and resources, patient preference, and social determinants of health such as transportation, access to care, housing situation, home support, health literacy, and income. In response to stakeholder questions regarding face validity of a clinical setting indicator based upon claims data, we assessed the validity of the clinical setting indicator by examining the correlation between setting and length of stay.

References

1. Courtney M, Frisch N, Bohl D, Della Valle C. Improving Value in Total Hip and Knee Arthroplasty: The Role of High Volume Hospitals. *The Journal of Arthroplasty.* 2018;33(1):1-5. https://doi.org/10.1016/j.arth.2017.07.040 .
2. Murphy WS, Cheng T, Lin B, Terry D, Murphy SB. Higher Volume Surgeons Have Lower Medicare Payments, Readmissions, and Mortality After THA. *Clin Orthop Relat Res*. 2019;477(2):334-341. doi:10.1097/CORR.0000000000000370.
3. Sibley R, Charumbhumi, V, Hutzler L, Paoli A, Bosco, J. Joint Replacement Volume Positively Correlates With Improved Hospital Performance on Centers for Medicare and Medicaid Services Quality Metrics. *The Journal of Arthroplasty*. 2017;32(5):1409-1413. https://doi.org/10.1016/j.arth.2016.12.010.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

Empiric Validity Results

Comparison to Star-Rating Summary Scores

Figures 1 and 2 show the box-whisker plots of the THA/TKA complications measure RSCRs within each quartile of Star-Rating summary scores calculated without the Inpatient only THA/TKA Complications measure for all hospitals (Figure 1) and hospitals with 25 or more cases (Figure 2). The diamonds represent the mean RSCRs of Star-Rating summary score quartiles. The correlation between THA/TKA complications and Star-Rating summary score is -0.101 (p-value <.0001) for all hospitals and -0.121 (p-value <.0001) for hospitals with 25+ cases, which suggests that hospitals with lower THA/TKA RSCRs are more likely to have higher Star-Rating summary scores, especially at the extremes.

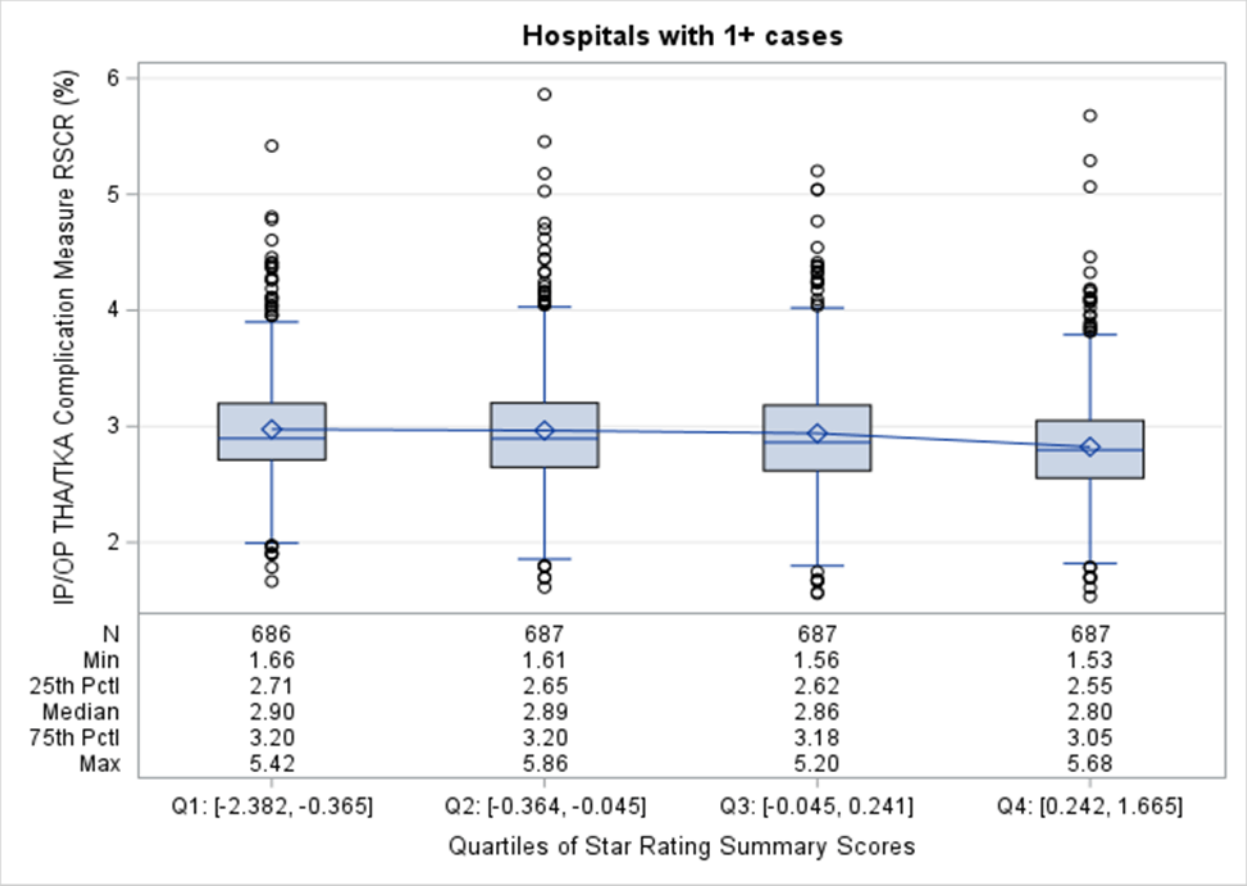

Figure 1. Distribution of RSCRs within quartiles of hospital star ratings summary scores (for hospitals with at least one case)

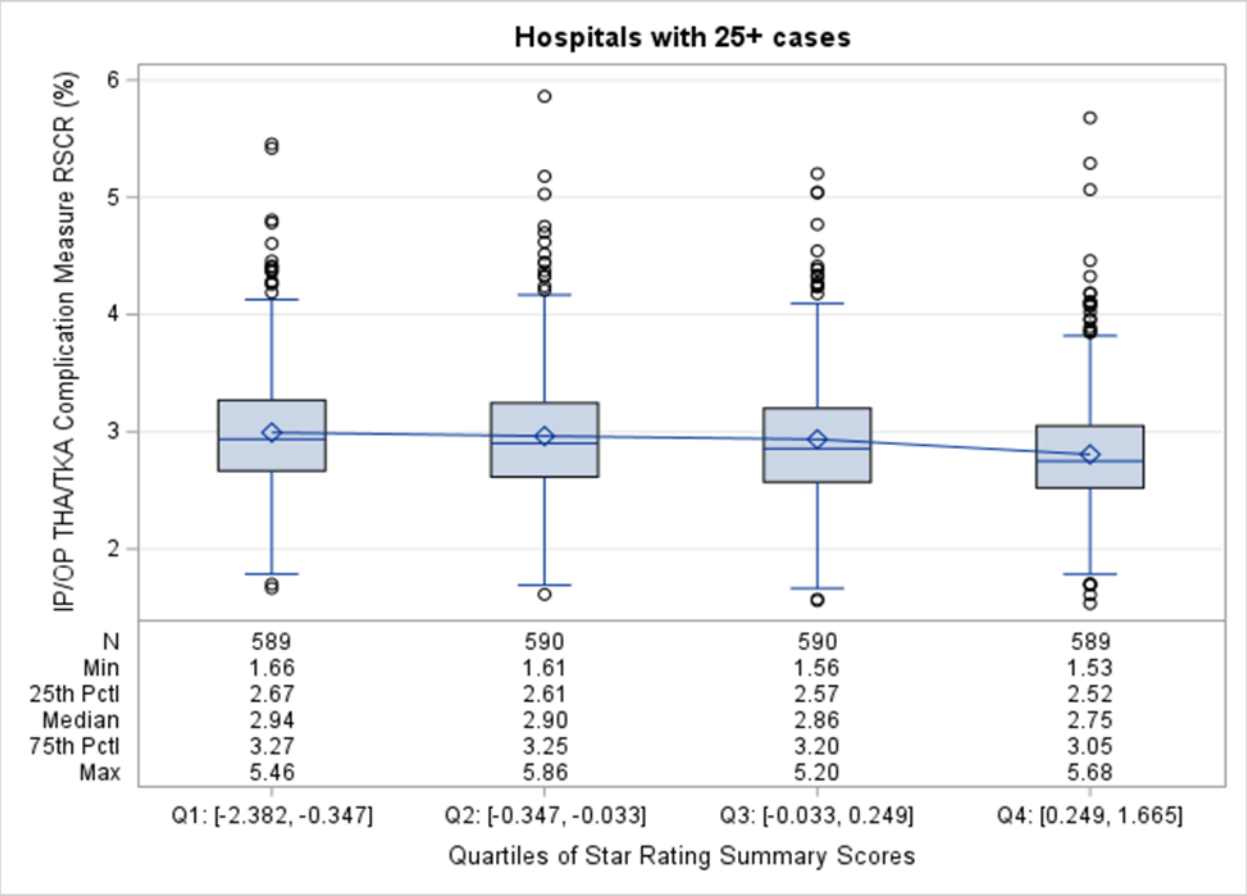

Figure 2. Distribution of RSCRs within quartiles of hospital star ratings summary scores (for hospitals with at least 25 cases)

Comparison to Hospital THA/TKA Surgical Admission Volume

Figures 3 and 4 illustrate the relationship between quartiles of admission volume and THA/TKA RSCRs for all hospitals (Figure 3) and for hospitals with 25+ cases (Figure 4). There is a general trend that high volume hospitals (those in the upper quartiles) have lower RSCRs than hospitals in other volume quartiles. Quarter 1 shows a relatively smaller range of RSCRS compared to the other three quartiles.

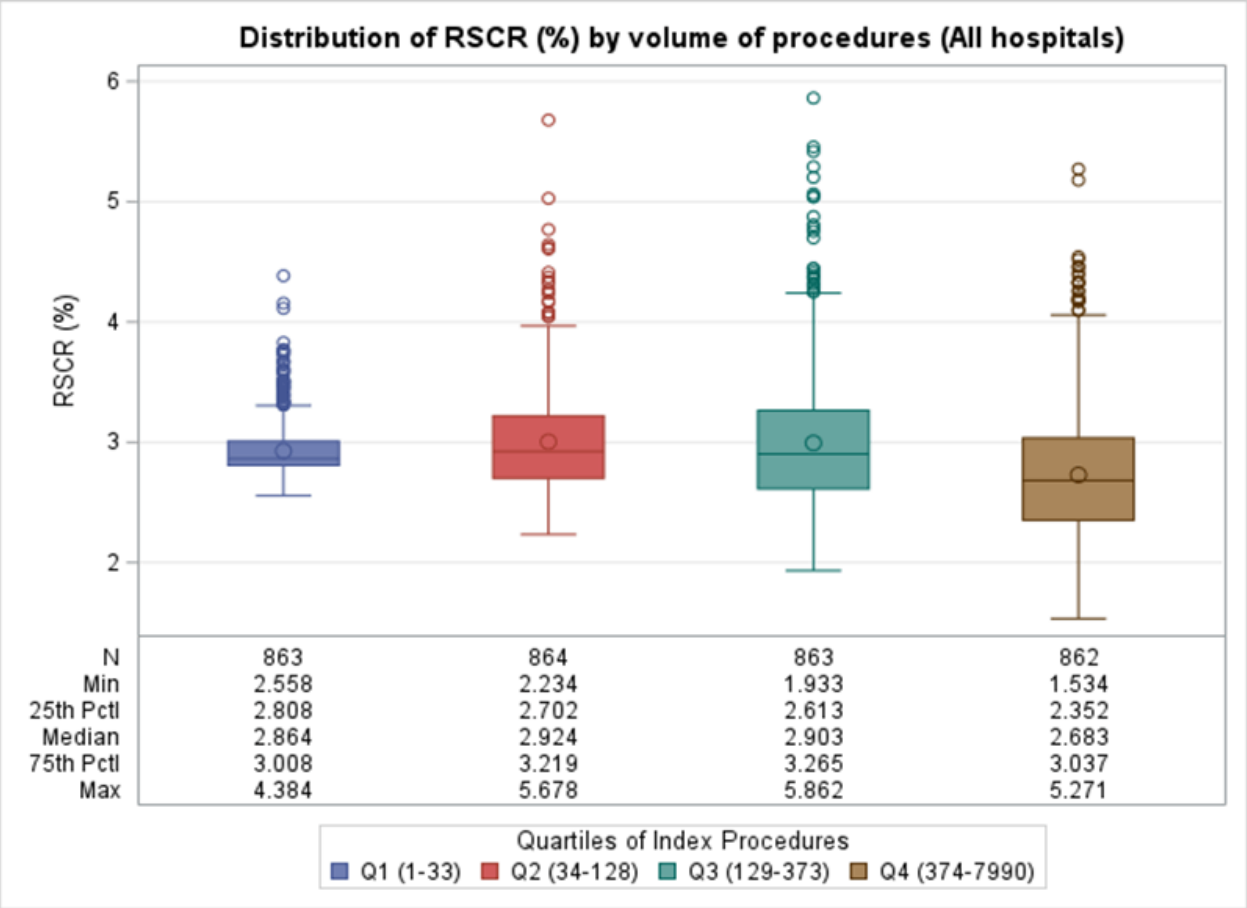

Figure 3. Distribution of RSCRs within quartiles of procedural volume (hospital with at least one case)

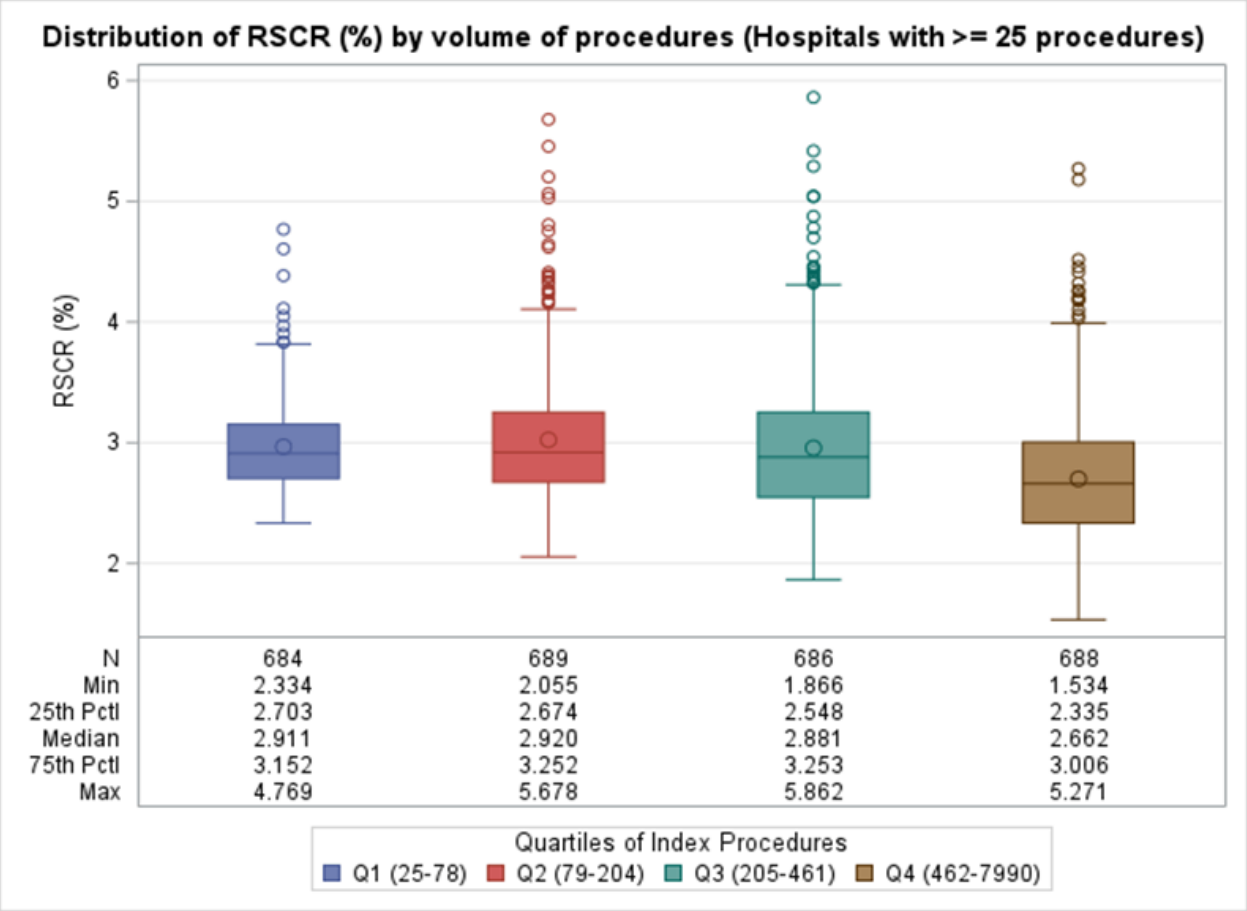

Figure 4. Distribution of RSCRs within quartiles of procedural volume (hospital with at least 25 cases)

Face Validity Results

Question #1

Among the 12 of 13 TEP Members who provided responses, 5 responded "Strongly Agree," 6 responded "Moderately Agree," and 1 responded "Somewhat Agree" to this question.

Question #2

Among the 12 of 13 TEP Members who provided responses, 3 responded "Strongly Agree," 6 responded "Moderately Agree," and 3 responded "Somewhat Agree" to this question.

We note that because this survey was anonymous, we do not know the identity of the one TEP member who did not complete the survey, nor do we know the reason why the TEP member did not complete the survey.

Additional Validity Results

We found 99.6% of all eligible outpatient procedures had a length of stay less than or equal to 3 days, which is consistent with the CMS definition of observation status.

[Response Ends]

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

[Response Begins]

Empiric Validity Results

This validation approach compares the IP/OP 90-Day THA/TKA Complication Measure results against the overall Hospital Star Rating summary scores. As expected, we found an observed trend of lower risk-standardized complications with higher star ratings, especially at the extremes, which supports measure score validity. We expected this relationship would be weak given the summary score includes 40 measures (41 measures minus the inpatient-only THA/TKA Complications Measure) from five different quality domains. Additionally, this validation approach compared various categories and quartiles of hospital THA/TKA admission volume with THA/TKA complication measure scores in Figure 3 and 4 – these results demonstrate an observed trend of higher hospital volume with lower complication measure scores, which is consistent with what has been published in the literature (Katz et al, 2004). Overall, the results above show that the trend and direction of this association is in line with what would be expected.

Face Validity Results

Among TEP members who provided responses, 100% of members agreed with the first statement that the THA/TKA complication measure as specified will provide a valid assessment of complications following elective THA/TKA, and 100% agreed with the second statement that the complication measure as specified can be used to distinguish between better and worse quality care among hospitals performing THAs/TKAs.

Among the 5 of 12 respondents who provided additional comments, 3 commenters encouraged CMMI to account for patient social risk, including adjusting for or stratifying by social risk if this measure is implemented in a payment program. Two commenters encouraged CMS/CMMI to continue to monitor racial and other disparities in THA/TKA outcomes. One commenter recommended considering opiate use in future measure reevaluation.

In summary, the TEP supported the face validity of this measure as demonstrated by universal agreement in responses to the two face validity statements.

Additional Validity Information

Our assessment of outpatient length of stay demonstrated rare cases beyond 3 days, supporting the claims-based setting indicator is consistent with policy definitions of setting.

Overall, we believe the combination of empiric and stakeholder expert face validity strongly support that this measure is a valid quality metric for hospitals for use in a hospital-level episode payment model (Comprehensive Care for Joint Replacement or CJR) that accounts for social risk, for which this measure was intended.

Reference

Katz JN, Barrett J, Mahomed NN, Baron JA, Wright RJ, Losina E. Association between hospital and surgeon procedure volume and the outcomes of total knee replacement. J Bone Joint Surg Am. 2004 Sep;86(9):1909-16. doi: 10.2106/00004623-200409000-00008. PMID: 15342752.

[Response Ends]

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

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

[Response Begins]

The measure score is a hospital-specific risk-standardized complication rate. These rates are obtained as the ratio of predicted to expected complications, multiplied by the national unadjusted rate. The "predicted" number of complications (the numerator) is calculated using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of complications. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are then transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of complications (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are then transformed and summed over all patients in the hospital to get an expected value.

We characterize the degree of variability by:

1. Reporting the distribution of RSCRs.
2. Providing the median odds ratio (MOR) (Merlo et al, 2006). The MOR represents the median increase in the odds of a complication within 90 days of a THA/TKA admission date on a single patient if the admission occurred at a higher risk hospital compared to a lower risk hospital. MOR quantifies the between-hospital variance in terms of odds ratio, it is comparable to the fixed effects odds ratio.

Reference

Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, Råstam L, Larsen K. (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health, 60(4):290-7.

[Response Ends]

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

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

[Response Begins]

Out of all 3,452 hospitals, the mean risk-standardized complication rate (RSCR) is 2.91%, and the median RSCR is 2.86% (Figure 5). The maximum and minimum are 5.86% and 1.53%, respectively, with IQR being 2.65-3.12%. Out of the 2,747 hospitals that have at least 25 index procedures, the mean RSCR is still 2.91%, and the median RSCR is 2.85% (Figure 6). The maximum and minimum are the same as those in all hospitals, with IQR being 2.59-3.18%.

The median odds ratio was 1.33.

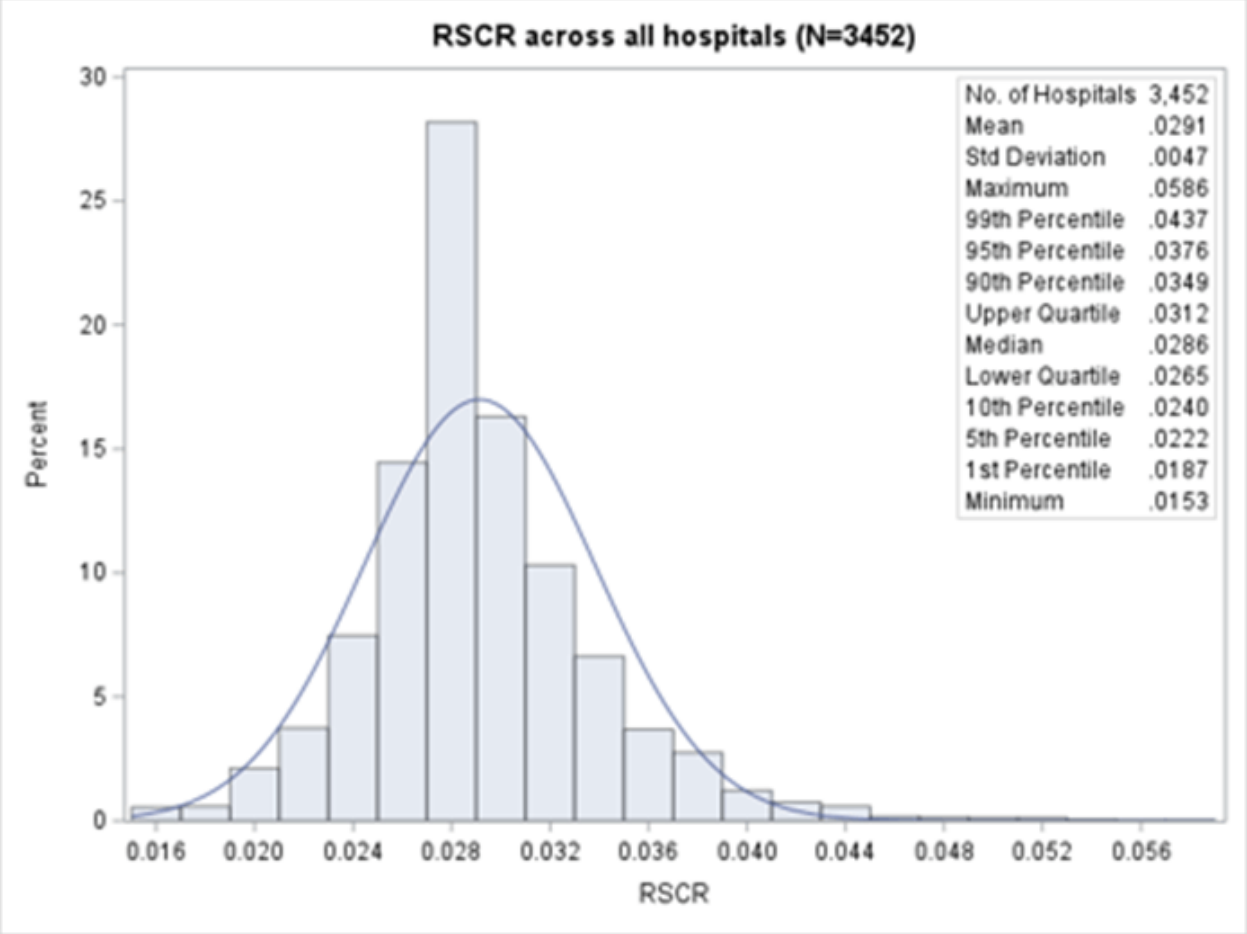

Figure 5. Distribution of RSCRs across all hospitals

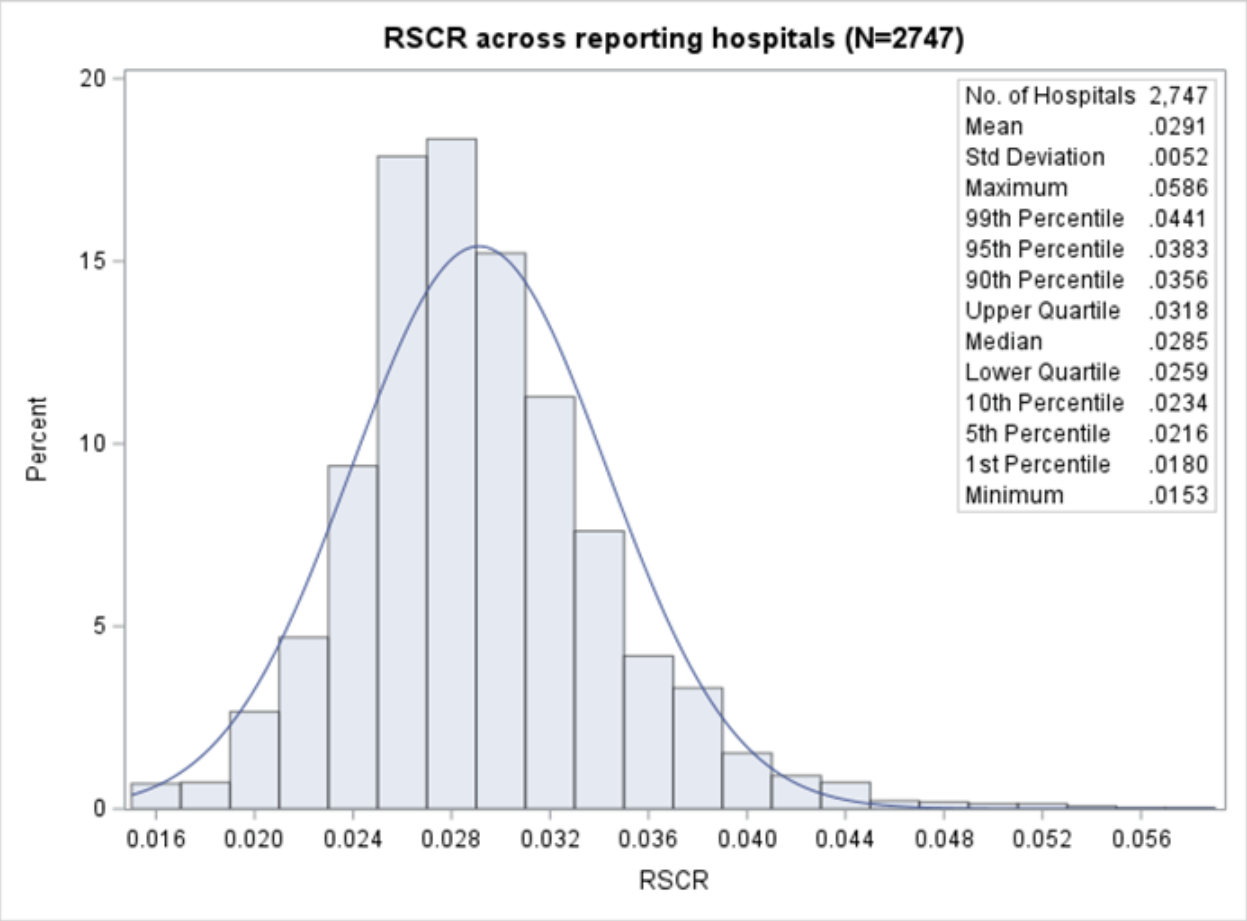

Figure 6. Distribution of RSCRs across all hospitals with at least 25 cases

[Response Ends]

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

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

[Response Begins]

Although the median RSCR is 2.86%, the IQR represents 16% of this value and the delta between the 10th and 90th percentiles is 38% of the median RSCR.

The MOR suggests a meaningful increase in the risk of complications when a patient has a THA/TKA procedure at a higher risk hospital compared to a lower risk hospital. A MOR value of 1.33 indicates that a patient has a 33% increase in the odds of a complication at a higher risk performance hospital compared to a lower risk hospital, indicating the impact of quality on the outcome rate.

The variation in rates suggest there are meaningful differences in the quality of care received across hospitals for THA/TKA procedures. This evidence supports measurement to reduce the variation.

[Response Ends]

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

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

[Response Begins]

The THA/TKA complications measure used claims-based data for respecification and testing. There was no missing data in the development and testing data.

[Response Ends]

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

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

[Response Begins]

Unlike data sources such as surveys or clinical electronic health records, claims data is not prone to missingness at the data element level. Therefore, frequencies of the data elements are not provided.

[Response Ends]

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

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

[Response Begins]

As noted above, unlike data sources such as surveys or clinical electronic health records, claims data is not prone to missingness at the data element level. Therefore, frequencies of the data elements are not provided.

[Response Ends]

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

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

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

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

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

[Response Begins]

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (Testing Dataset, Table 4 below). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in Section sp.17 (Denominator Exclusions).

[Response Ends]

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

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

[Response Begins]

The Testing Dataset (Table 4) below, shows the distribution of exclusions among hospitals with 1 or more admissions, representing <0.75% of the overall cohort:

Exclusion	N	%	Distribution (%) across hospitals (n=3.452) (Min, 25th, 50th, 75th percentile, Max)
1. Discharged against medical advice (AMA)	192	0.02	(0, 0, 0, 0, 33.33)
2. Without enrollment in Medicare FFS for at least 90 days following the index encounter	7,532	0.72	(0, 0, 0.43, 1.11, 100)
3. Admissions for patients with more than two THA/TKA procedure codes during the index admission	0	0.00	(0, 0, 0, 0, 0)

Table 4. Frequency and Distribution of Exclusions Across Hospitals

[Response Ends]

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

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

[Response Begins]

Exclusion 1 (patients who are discharged AMA) accounts for 0.02% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 2 (patients without at least 90 days of post-discharge enrollment in FFS Medicare for index admissions) accounts for 0.72% of all index admissions excluded from the initial cohort. This exclusion is needed because the 90-day complication outcome cannot be assessed in this group since claims data are used to determine whether a patient has experienced complications. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 3 (patients with more than two THA/TKA procedure codes during the index hospitalization) accounts for 0 index procedures excluded from the initial index cohort. Although clinically possible, it is highly unlikely that patients would receive more than two elective THA/TKA procedures in one hospitalization, which may reflect a coding error. Also, because a very small percent of patients are excluded (0% in this analysis), this exclusion is unlikely to affect the measure score.

[Response Ends]

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

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

34

[Response Ends]

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

[Response Begins]

See risk model specification in section 2b.24 and the attached code set file.

[Response Ends]

Attachment: 3732_3732_3732_IP.OP THA_TKA Complication_CodeSetFileforNQF.6.27.22_(1)-508_(1).xlsx

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Published literature

Internal data analysis

Other (specify)

[Other (specify) Please Explain]

Discussions with patients, clinicians and equity subject matter experts.

[Response Ends]

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

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

[Response Begins]

Selecting Risk Variables

Our goal in selecting risk factors for adjustment was to develop a parsimonious model that included clinically relevant variables strongly associated with the risk of complication in the 90 days following an index procedure. We generally use a two-stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of additional risk variables, such as social risk factors.

The goal of risk adjustment is to account for patient characteristics, such as age and comorbid conditions, at the time of admission that are clinically relevant, have strong relationships with the outcome, and are outside of the control of the reporting entity without obscuring important quality differences. Our conceptual model assumes that hospitals directly control or influence many factors that impact the occurrence of complications. These include: immediate perioperative medical and surgical care; staff experience, expertise and training (including recognizing and reducing implicit bias); procedural workflows and volumes; care coordination and discharge protocols; and institutional safety culture. Factors that hospitals may or may not be able to influence through their protocols and care include: surgical setting (inpatient vs. outpatient); patients' clinical comorbidities; patients' demographic factors such as gender, age, frailty, health behaviors, and functional status; and patients' social risk factors such as exposure to racism, housing, transportation, access to care, insurance, income, education, and literacy. Many of these factors are difficult to measure uniformly on all Medicare beneficiaries and therefore proxy measures can be useful. We sought to account for risk factors that are measurable and less under the influence of the hospital, clinical comorbidities and frailty primary among these. We strove to not account for factors that are under the hospital's control. For factors in between these categories, for which the hospital may have some influence, we examined reasonably proxy metrics and discussed these with our TEP. CMS made the final decision for such factors.

Approach

Since the IP/OP 90-Day THA/TKA Complication measure is a re-specification of the existing Hospital-level THA/TKA 90-day Complication measure (to expand the cohort to include patients undergoing eligible elective primary THA/TKA procedures in the HOPD setting), we conducted exploratory analyses to examine the similarities and differences between the inpatient and outpatient cohorts with respect to risk factors used in the existing Hospital-level THA/TKA Complication measure, as well as basic patient demographic characteristics (e.g., age, gender, race). The existing measure risk model includes several factors that are strong predictors of patient frailty and functional status, including malnutrition, osteoporosis and vertebral fractures, dementia, paralysis, and decubitus ulcers.

To assess the appropriateness of using the existing risk model for the inpatient only measure, we examined risk-adjustment variables for patients' comorbid conditions identified in both inpatient and outpatient claims for the 12 months prior to the index encounter, as well as those risk factors coded as Present On Admission (POA) for patients in the inpatient cohort (Krumholz et al, 2019). Since POA indicators are not available in outpatient claims, conditions that may represent adverse outcomes due to care received during the index encounter are not adjusted for in the model for eligible procedures performed in the HOPD setting.

We then assessed risk factor frequencies across the inpatient and HOPD settings; we also examined the association with the complication outcome in univariate analyses by setting.

We then performed calibration plots, examined predictive ability, and c-statistics to assess model calibration, prediction, and discrimination (see Sections 2b.26, 2b.27, 2b.28, 2b.29, and 2b.31) for additional details on model testing).

Next, we examined additional risk variables, including clinical setting (inpatient vs. HOPD), non-White race as a proxy for exposure to racism, and social risk factors representing patient- and community-level social determinants of health.

This resulted in a final risk-adjustment model that included 34 variables.

Clinical Setting

This measure includes patients undergoing procedures in both the inpatient and HOPD settings. Initial empiric analyses indicated that complication rates were lower after HOPD procedures compared to inpatient procedures, despite only modest differences in clinical risk factor frequency across setting. Through discussions with internal clinical and health policy experts, the Clinical Working Group and the TEP, we concluded that there are multiple factors influencing the decision to perform elective THA/TKA procedures in the inpatient vs. HOPD setting. These factors include clinical risk assessment by the surgical team (including frailty), hospital policies and resources, patient preference, and social

determinants of health such as transportation, access to care, housing situation, home support, health literacy and income. We assessed the validity of the clinical setting by examining the correlation between setting and length of stay.

We then assessed model performance using three approaches to risk adjustment: 1) using a single risk model with regression coefficients estimated from the combined cohort of inpatient and outpatient procedures (Single Combined IP/OP Model); 2) using two distinct models, one for each setting with identical risk factors but allowing regression coefficients to vary by setting; and 3) a single combined model (as in #1) but with an indicator variable added reflecting inpatient vs. HOPD setting. As noted in Sections 2b.26, 2b.27, 2b.28, 2b.29, and 2b.31 below, we assessed model performance through calibration plots and statistics, c-statistics, and predictive ability.

Social Risk Factors

We weigh social risk factor adjustment using a comprehensive approach that evaluates the following:

- A summary of the literature that describes the relationship between patient-level risk variables and outcomes, specifically for THA/TKA procedures;
- A conceptual model for influence of social risk factors on measure outcome;
- Feasibility of testing meaningful social risk factors in available data (section 2a.08); and
- Empiric testing of social risk factors (section 2b.02).

Below, we summarize the findings of the literature review and conceptual pathways by which social risk factors may influence risk of the outcome, as well as the statistical methods for social risk factor empiric testing. Our conceptualization of the pathways by which patients' social risk factors affect the outcome is informed by the literature cited below and IMPACT Act–funded work by the National Academy of Science, Engineering and Medicine (NASEM) and the Department of Health and Human Services Assistant Secretary for Policy and Evaluation (ASPE).

Literature summary for social risk factor association with THA/TKA outcomes

Much of the literature around social risk and THA/TKA focuses on access to the procedure, which may be limited by eligibility criteria such as body mass index (BMI) and HbA1c. For example, women, patients with lower socioeconomic status, and non-Hispanic Black patients have lower odds of being eligible for lower-extermity arthroplasty compared with their counterparts (men, patients with higher socioeconomic status, and non-Hispanic Whites) (Wang et al., 2018). However, even after accounting for clinical comorbidities and socioeconomic status, research has shown that Black patients were still less likely to undergo TKA, suggesting there are additional factors underlying this disparity. (MacFarlane et al., 2018) Below, however, we focus on the relationships between social risk factors and outcomes following a procedure, rather than access to the procedure. Outcomes commonly assessed in the literature include specific complications (including pulmonary embolism, deep vein thrombosis, excessive bleeding, and prosthetic joint and other infections), mortality, readmission, and length of stay; below we focus on all of the outcomes except for length of stay which is not part of the outcome definition for the IP/OP 90-Day THA/TKA Complication Measure.

Race

Many studies have evaluated the relationship between race and outcomes and the evidence for an association is mixed. A recent (2022) systematic review found that most studies show higher rates of complications for Black and Hispanic patients compared with White patients. (Alvarez et al., 2022) For example, several recent studies (Cusano et al., 2021; Johnson et al., 2020) have found that after controlling for comorbidities, Black and Hispanic patients were more likely to have higher rates of complications following THA/TKA procedures. A 2020 study in the Kaiser Permanente System found similar results, showing worse outcomes for Black patients compared with White patients. (Hinman et al., 2020) Some of these disparities could be due to worse pre-operative function as well as higher odds of receiving care at low-quality, low-volume hospitals. (Alvarez, et al., 2022; Cai, et al., 2012; Lavernia, et al., 2004; Lavernia, et al., 2015; Slover, et al., 2010; SooHoo, et al., 2011; SooHoo, et al., 2008)

On the other hand, a 2018 study that used a large, international clinical registry (ACS-NSQIP) found that outcomes (a composite of complications) between White and Black men and women were similar; women had a higher rate of complications due to a higher rate of transfusions (Wang et al, 2018). In addition, a 2019 study in a US total joint replacement registry that examined insured patients, found that while 90-day ED visit rates were higher for Black, Hispanic, and Asian patients, rates of infection, venous thromboembolism, readmission, and mortality were similar or

lower compared with White patients (Okike et al, 2019). In our own empiric results for this measure, we find that unadjusted complication rates are similar for Black and Non-White patients, but somewhat higher for other (non-Black, non-White) patients (see section 2b.25 below).

However, more recently, researchers found that when controlling for socioeconomic status using the Area Deprivation Index (ADI), outcomes for Black and White patients were similar (Abstract presented at American Academy of Orthopaedic Surgeons 2022 Annual Meeting), suggesting that socioeconomic risk variables may be the more dominant driver of outcomes.

Socioeconomic risk factors

A recent study from the Cleveland Clinic used the ADI to examine disparities in THA/TKA outcomes. Study authors found that patients with higher ADI scores (greater levels of social risk) had higher readmission rates compared with patients with lower ADI scores, but that 90-day emergency department visits and reoperations did not differ significantly between the two groups (Khlopas et al, 2022). As mentioned above, when White and Black patients are similar in terms of their ADI scores, outcomes are also similar, suggesting that socioeconomic status, rather than race, may be driving differences in outcomes.

In a study examining the association between dual eligibility and THA/TKA outcomes, after controlling for comborbidites, patients with dual eligibility were more likely to visit the ED within 90 days compared with non-duals (Koresselet al, 2022).

Other factors

In addition to race and socioeconomic risk factors, other factors such as frailty, age, gender, and provider density have been examined for their relationship to THA/TKA outcomes. As described above, variables differ in their association with the specific complications. For example, age is associated with a greater risk of cardiac complications (Elsiwy et al, 2019). Women are more likely to experience venous thromboembolism (Zhang et al, 2015), whereas men are at higher risk of death, AMI, pneumonia, and surgical site infections (Basques et al, 2019). Frailty has been shown to be associated with a higher risk of reoperation and readmission (Runner et al, 2017). Finally, no association has been seen between THA/TKA outcomes and primary care provider density (Mehta et al, 2021).

Causal Pathways for Social Risk Variable Selection

Figure 7 combines information from the social risk factor literature summary, prior published literature, clinical input, and empiric data on clinical risk factors, and feedback from experts in disparities and quality measurement, as well as feedback from our TEP. There are two types of risk factors that are associated with the complications outcome: 1) hospital-level factors that directly impact patient outcomes, and 2) patient-level factors that may be associated with the outcomes. Some of the patient-level factors (box on lower left) have already been accounted for in the measure's risk model whereas other factors, namely social risk factors, are not currently accounted for in the measure's risk model. However, we emphasize that other CMS payment programs (such as CMS' Hospital Readmission Reduction Program and CMMI's Comprehensive Care for Joint Replacement or CJR model) have recently started to account for dual eligibility within the payment program rather than the quality measure(s). Implementation planning has not been finalized; CMMI may opt to do the same for the implementation of this measure.

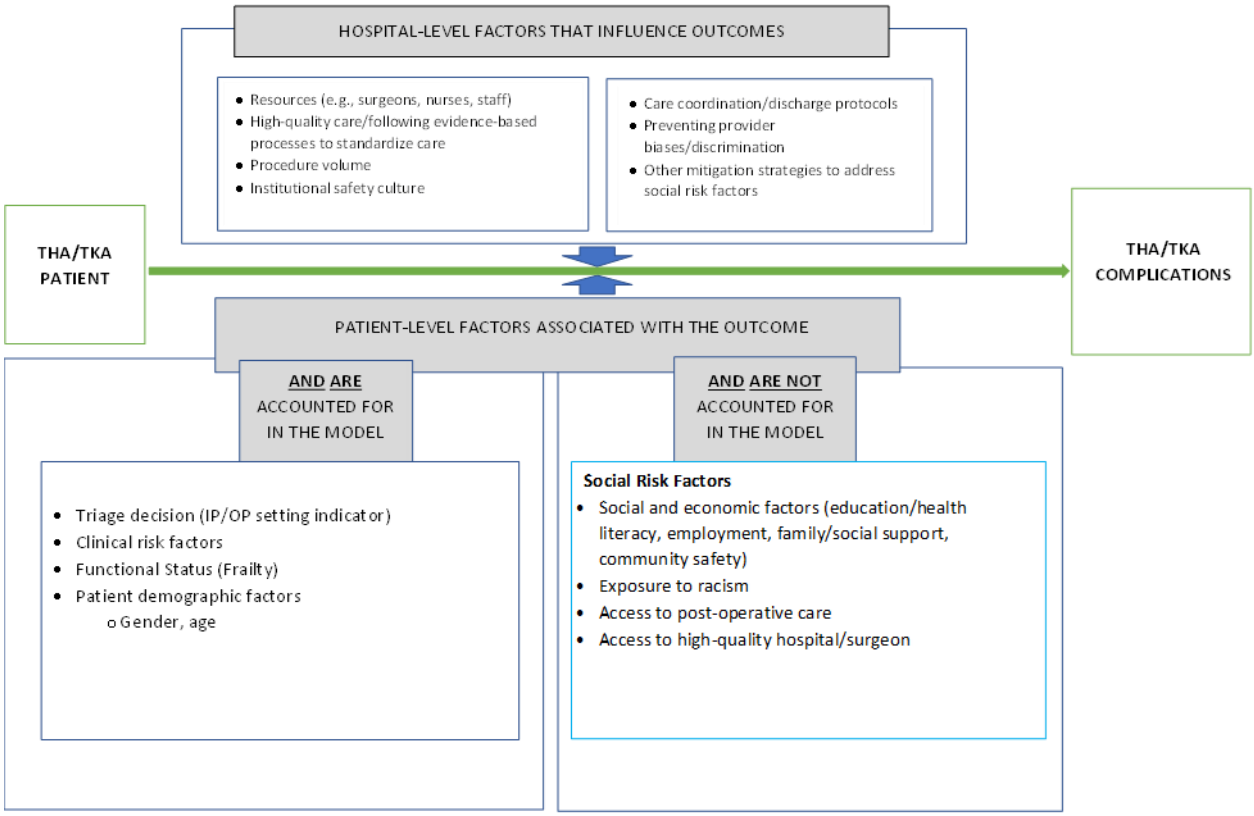

Figure 7: Conceptual model for risk factors and THA/TKA complications

The potential causal pathways by which social risk factors identified above in Figure 7 (variables in the lower right hand box) influence the risk of complication following major surgery, like the factors themselves, are varied, complex, and sometimes overlapping. There are at least four potential pathways that are important to consider:

1. **Patients with social risk factors may have worse health at the time of hospital admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These social risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment.
2. **Patients with social risk factors often receive care at lower quality hospitals**. Patients of lower income, lower education, or unstable housing have inequitable access to high quality facilities, in part, because such facilities are less likely to be found in geographic areas with large populations of poor patients. Thus, patients with low income are more likely to be seen in lower quality hospitals, which can explain increased risk of complications following hospitalization.
3. **Patients with social risk factors may receive differential care within a hospital**. The third major pathway by which social risk factors may contribute to complications risk is that patients may not receive equivalent care within a facility. First, they may receive biased care due to their race or ethnicity, or they may fail to receive the needed differentiated care (such as the provision of lower literacy information for patients with lower education).
4. **Patients with social risk factors may experience worse health outcomes beyond the control of the health care system.** Some social risk factors, such as income or wealth, may affect the likelihood of complications without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing financial priorities which don't allow for adequate recuperation or access to needed treatments, or a lack of access to care outside of the hospital.

The conceptual model (Figure 7) identifies that hospitals can implement mitigating strategies to counteract the impact of patient-level social risk factors, which are outlined in Table 5. We note that some of the evidence for mitigation is directly related to evidence from quality improvement efforts related to THA/TKA procedures. For example, there is some evidence from the implementation of the Comprehensive Care for Joint Replacement (CJR) model, that, while disparities continue to exist for some outcomes, they have narrowed for 90-day readmission (Okewunmi et al, 2022). This may be due to the types of processes that facilities implemented in response to this payment model, including redesigning and standardizing care pathways that include pre-procedure education, discharge planning, preemptive multimodal pain control, and early rehabilitation, which have been shown to improve outcomes, including in patients with social risk factors (Riepen et al, 2021). Other facility-specific care redesign features associated with improved outcomes overall include use of risk prediction tools, care managers to engage with patients before admission, nurse navigators, and 7-day access to an after-hours clinic (Gray et al, 2018). Some of the evidence to support mitigating solutions is generalized from broader research, much of which focuses on preventing readmission (CMS Guide to Reducing Disparities in Readmissions, 2018), which is partially captured by this measure (for readmissions that qualify as complications).

Because several of these social risk factors have similar underlying drivers of poor outcomes (such as education and income, and race and income), we address the mitigating strategies below as a series of topics and the recommended strategies. We also acknowledge that mitigating exposure to racism, in particular, is a complex issue and that hospitals may initially struggle to develop and implement effective approaches. In addition, additional research is needed to identify the most effective approaches (Ricks et al, 2021).

Strategy	Description of interventions	Related social risk factors
Improving post-procedural follow up care	• Advance care transition planning and follow up for patients at high risk • Communicate with patients about importance of follow up care; assist with scheduling appointments • Offer telehealth options • Expand clinic hours to avoid ED use • Engage family/caregivers • Use of nurse navigators	Income Education Exposure to racism Access to post-operative care
Improve access to a usual source of care	• Ensure patient is connected with a usual source of care	Income Education Exposure to racism Access to post-operative care
Reduce language/literacy barriers	• Identify patients at risk (language and literacy barriers) • Ensure access to translation services • Communicate at home or follow up care instructions in patient's native language and in a culturally competent manner • Simplify instructions • Communicate instructions at the appropriate literacy level • Engage family/caregivers	Low health literacy Limited English proficiency

Strategy	Description of interventions	Related social risk factors
Reduce socioeconomic barriers	• Connect patients with community-based resources that address needs (e.g., housing and food insecurity, transportation, employment) • Connect underinsured patients with supplemental insurance • Connect with social support services	Income Education Exposure to racism
Reduce biased care	• Track metrics stratified by race and ethnicity • Quality improvement • Staff training • Diversity of staff, trainees, and Board of Directors	Exposure to racism Income
Improve access to high-quality care	• Recruit, train, and retain high-quality staff • Follow standards of care and use a learning healthcare system • Address workforce shortages and burnout	Exposure to racism Income Education

Table 5. Strategies and interventions to reduce the impact of social risk factors

Based on the literature showing a relationship between social risk factors and complications and assessment of available risk variables that can be linked to claims data (outlined in section 1.8), we tested the impact of adjusting for the following social risk variables on the IP THA/TKA measure:

- Race
- Dual eligible status
- AHRQ SES index

Please refer to section 2a.08 for a detailed description of each social risk factor.

Statistical Methods

We assessed the relationship between the social risk factor variables with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the impact of each variable on model performance and examined model risk prediction for each sub-group of patients. Finally, we examined the impact of adjustment on measure scores, including the relationship between measure scores and the facility proportion of patients with social risk factors.

References

1. Alvarez PM, McKeon JF, Spitzer AI, et al. Race, Utilization, and Outcomes in Total Hip and Knee Arthroplasty: A Systematic Review on Health-Care Disparities. *JBJS Rev*. 2022;10(3):10.2106/JBJS.RVW.21.00161. Published 2022 Mar 1. doi:10.2106/JBJS.RVW.21.00161
2. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. 2017;389(10077):1453-1463. doi:10.1016/S0140-6736(17)30569-X
3. Basques BA, Bell JA, Fillingham YA, Khan JM, Della Valle CJ. Gender Differences for Hip and Knee Arthroplasty: Complications and Healthcare Utilization. *J Arthroplasty*. 2019;34(8):1593-1597.e1. doi:10.1016/j.arth.2019.03.064
4. Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation Cardiovascular quality and outcomes 2014; 7:391-7.
5. Borza T, Oerline MK, Skolarus TA, et al. Association Between Hospital Participation in Medicare Shared Savings Program Accountable Care Organizations and Readmission Following Major Surgery. Ann Surg. 2019;269(5):873-878. doi:10.1097/SLA.0000000000002737.

6. Buntin MB, Ayanian JZ. Social Risk Factors and Equity in Medicare Payment. *New England Journal of Medicine.* 2017;376(6):507-510.
7. Cai X, Cram P, Vaughan-Sarrazin M. Are African American patients more likely to receive a total knee arthroplasty in a low-quality hospital?. *Clin Orthop Relat Res*. 2012;470(4):1185-1193. doi:10.1007/s11999-011-2032-6
8. Cusano A, Venugopal V, Gronbeck C, Harrington MA, Halawi MJ. Where Do We Stand Today on Racial and Ethnic Health Inequities? Analysis of Primary Total Knee Arthroplasty from a 2011-2017 National Database. *J Racial Ethn Health Disparities*. 2021;8(5):1178-1184. doi:10.1007/s40615-020-00875-8
9. Elsiwy Y, Jovanovic I, Doma K, Hazratwala K, Letson H. Risk factors associated with cardiac complication after total joint arthroplasty of the hip and knee: a systematic review. *J Orthop Surg Res*. 2019;14(1):15. Published 2019 Jan 11. doi:10.1186/s13018-018-1058-9
10. Gray CF, Prieto HA, Duncan AT, Parvataneni HK. Arthroplasty care redesign related to the Comprehensive Care for Joint Replacement model: results at a tertiary academic medical center. *Arthroplast Today*. 2018;4(2):221-226. Published 2018 Mar 21. doi:10.1016/j.artd.2018.02.002
11. Hinman AD, Chan PH, Prentice HA, Paxton EW, Okike KM, Navarro RA. The Association of Race/Ethnicity and Total Knee Arthroplasty Outcomes in a Universally Insured Population. *J Arthroplasty*. 2020;35(6):1474-1479. doi:10.1016/j.arth.2020.02.002
12. Hostetter M, Klein S. Confronting racism in health care. Commonwealth Fund. https://www.commonwealthfund.org/publications/2021/oct/confronting-racism-health-care. Published October 18, 2021. Accessed July 19, 2022.
13. Irgit, K., & Nelson, C. L. (2011). Defining Racial and Ethnic Disparities in THA and TKA. Clinical Orthopaedics and Related Research®, 469(7), 1817–1823.
14. Johnson MA, Sloan M, Lopez VS, Andah G, Sheth N, Nelson C. Racial disparities in peri-operative complications following primary total hip arthroplasty. J Orthop. 2020;21:155-160. Published 2020 Mar 26. doi:10.1016/j.jor.2020.03.037
15. Kehle SM, Greer N, Rutks I, Wilt T. Interventions to improve veterans' access to care: a systematic review of the literature. *J Gen Intern Med*. 2011;26 Suppl 2(Suppl 2):689-696. doi:10.1007/s11606-011-1849-8
16. Kerman, H. M., Smith, S. R., Smith, K. C., Collins, J. E., Suter, L. G., Katz, J. N., & Losina, E. (2018). Disparities in Total Knee Replacement: Population Losses in Quality-Adjusted Life-Years Due to Differential Offer, Acceptance, and Complication Rates for African Americans. Arthritis Care & Research, 70(9), 1326–1334.
17. Koressel JE, Perez BA, Kerbel YE, DeAngelis RD, Israelite CL, Nelson CL. Does dual-eligible Medicare/Medicaid insurance status as a surrogate for socioeconomic status compromise total knee arthroplasty outcomes? *The Journal of Arthroplasty*. 2022;37(6).
18. Khlopas A, Grits D, Sax OC, et al. Neighborhood Socioeconomic Disadvantages Associated With Prolonged Lengths of Stay, Nonhome Discharges, and 90-Day Readmissions After Total Knee Arthroplasty. *J Arthroplasty*. 2022;37(6S):S37-S43.e1. doi:10.1016/j.arth.2022.01.032
19. Krumholz, H.M., et al., *Development and Testing of Improved Models to Predict Payment Using Centers for Medicare & Medicaid Services Claims Data*. JAMA Network Open, 2019. **2**(8): p. e198406-e198406.
20. Lavernia CJ, Lee D, Sierra RJ, Gómez-Marín O. Race, ethnicity, insurance coverage, and preoperative status of hip and knee surgical patients. *J Arthroplasty*. 2004;19(8):978-985. doi:10.1016/j.arth.2004.03.028
21. Lavernia CJ, Villa JM. Does Race Affect Outcomes in Total Joint Arthroplasty?. *Clin Orthop Relat Res*. 2015;473(11):3535-3541. doi:10.1007/s11999-015-4481-9
22. MacFarlane LA, Kim E, Cook NR, et al. Racial Variation in Total Knee Replacement in a Diverse Nationwide Clinical Trial. *J Clin Rheumatol*. 2018;24(1):1-5. doi:10.1097/RHU.0000000000000613

23. Mehta B, Brantner C, Williams N, Szymonifka J, Navarro-Millan I, Mandl LA, Bass AR, Russell LA, Parks ML, Figgie MP, Nguyen JT, Ibrahim S, Goodman SM. Primary care provider density and elective total joint replacement outcomes. *Arthroplasty Today*. 2021;10:73–78.
24. Okewunmi, Jeffrey BS1; Mihalopoulos, Meredith BA1; Huang, Hsin-Hui MD, PhD1,2; Mazumdar, Madhu PhD1; Galatz, Leesa M. MD1; Poeran, Jashvant MD, PhD1,2,3,a; Moucha, Calin S. MD1 Racial Differences in Care and Outcomes After Total Hip and Knee Arthroplasties, The Journal of Bone and Joint Surgery: June 1, 2022 - Volume 104 - Issue 11 - p 949-958
25. doi: 10.2106/JBJS.21.00465
26. Okike K, Chan PH, Prentice HA, Navarro RA, Hinman AD, Paxton EW. Association of Race and Ethnicity with Total Hip Arthroplasty Outcomes in a Universally Insured Population. *J Bone Joint Surg Am*. 2019;101(13):1160-1167. doi:10.2106/JBJS.18.01316
27. Re-Engineered Discharge (RED) Toolkit. AHRQ. https://www.ahrq.gov/patient-safety/settings/hospital/red/toolkit/index.html. Published February 2020. Accessed July 19, 2022.
28. Ricks TN, Abbyad C, Polinard E. Undoing Racism and Mitigating Bias Among Healthcare Professionals: Lessons Learned During a Systematic Review [published online ahead of print, 2021 Sep 3]. *J Racial Ethn Health Disparities*. 2021;1-11. doi:10.1007/s40615-021-01137-x
29. Riepen DW, Gelvez D, Collett GA, Nakonezny P, Estrera KA, Huo MH. Standardized total knee arthroplasty pathway improves outcomes in minority patients. *Am J Manag Care*. 2021;27(5):e152-e156. Published 2021 May 1. doi:10.37765/ajmc.2021.88637
30. Runner RP, Bellamy JL, Vu CCL, Erens GA, Schenker ML, Guild GN 3rd. Modified Frailty Index Is an Effective Risk Assessment Tool in Primary Total Knee Arthroplasty. *J Arthroplasty*. 2017;32(9S):S177-S182. doi:10.1016/j.arth.2017.03.046
31. Sheridan SL, Halpern DJ, Viera AJ, Berkman ND, Donahue KE, Crotty K. Interventions for individuals with low health literacy: a systematic review. *J Health Commun*. 2011;16 Suppl 3:30-54. doi:10.1080/10810730.2011.604391
32. Slover JD, Walsh MG, Zuckerman JD. Sex and race characteristics in patients undergoing hip and knee arthroplasty in an urban setting. *J Arthroplasty*. 2010;25(4):576-580. doi:10.1016/j.arth.2009.03.002
33. SooHoo NF, Farng E, Zingmond DS. Disparities in the utilization of high-volume hospitals for total hip replacement. *J Natl Med Assoc*. 2011;103(1):31-35. doi:10.1016/s0027-9684(15)30240-6
34. SooHoo NF, Zingmond DS, Ko CY. Disparities in the utilization of high-volume hospitals for total knee replacement. *J Natl Med Assoc*. 2008;100(5):559-564. doi:10.1016/s0027-9684(15)31303-1
35. Wang AY, Wong MS, Humbyrd CJ. Eligibility Criteria for Lower Extremity Joint Replacement May Worsen Racial and Socioeconomic Disparities. *Clin Orthop Relat Res*. 2018;476(12):2301-2308. doi:10.1097/CORR.0000000000000511
36. Zhang J, Chen Z, Zheng J, Breusch SJ, Tian J. Risk factors for venous thromboembolism after total hip and total knee arthroplasty: a meta-analysis. *Arch Orthop Trauma Surg*. 2015;135(6):759-772. doi:10.1007/s00402-015-2208-8
37. 2018. *Guide to Reducing Disparities in Readmissions*. [ebook] Centers for Medicare & Medicaid Services. Available at: <https://www.cms.gov/about-cms/agency-information/omh/downloads/omh_readmissions_guide.pdf> [Accessed 19 July 2022].

[Response Ends]

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

[Response Begins]

Final risk model variables, including their frequencies and odds ratios, are shown in Table 6.

Variable	Combined IP/OP THA/TKA N=989,587	*
*	n (%)	Adjusted OR (95% CI)
Intercept	*	0.01 (0.01, 0.01)
Claim setting (IP)	658,403 (66.53%)	1.46 (1.41, 1.5)
Years over 65 (continuous)	8.72 (5.72)	1.03 (1.02, 1.03)
Male	378,804 (38.3)	1.02 (0.99, 1.05)
Index admissions with an elective THA procedure	347,216 (35.1)	1.75 (1.71, 1.79)
Number of procedures (two vs. one)	12,624 (1.3)	1.5 (1.35, 1.65)
Metastatic cancer and acute leukemia (CC 8)	6,890 (0.7)	1.16 (1.03, 1.32)
Other major cancers (CC 9-12)	125,787 (12.7)	0.93 (0.89, 0.96)
Respiratory/heart/digestive/urinary/other neoplasms (CC 13-15)	184,836 (18.7)	0.93 (0.9, 0.96)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	261,104 (26.4)	1.09 (1.07, 1.12)
Protein-calorie malnutrition (CC 21)	6,415 (0.6)	1.28 (1.16, 1.42)
Bone/joint/muscle infections/necrosis (CC 39)	28,829 (2.9)	1.24 (1.17, 1.31)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 40)	103,386 (10.4)	1.2 (1.16, 1.24)
Osteoarthritis of hip or knee (CC 42)	963,020 (97.3)	0.94 (0.87, 1.01)
Osteoporosis and other bone/cartilage disorders (CC 43)	242,374 (24.5)	1.07 (1.04, 1.1)
Dementia or other specified brain disorders (CC 51-53)	38,668 (3.9)	1.2 (1.14, 1.26)
Major psychiatric disorders (CC 57-59)	59,741 (6)	1.34 (1.29, 1.4)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	15,580 (1.6)	1.24 (1.15, 1.34)
Cardio-respiratory failure and shock (CC 84), plus ICD-10-CM codes R09.01 and R09.02	27,563 (2.8)	1.34 (1.26, 1.41)
Coronary atherosclerosis or angina (CC 88-89)	231,970 (23.4)	1.16 (1.13, 1.2)
Stroke (CC 99-100)	18,294 (1.8)	1.09 (1.01, 1.17)
Vascular or circulatory disease (CC 106-109)	220,476 (22.3)	1.17 (1.14, 1.21)
Chronic obstructive pulmonary disease (COPD) (CC 111)	107,013 (10.8)	1.37 (1.33, 1.42)
Pneumonia (CC 114-116)	35,438 (3.6)	1.15 (1.09, 1.22)
Pleural effusion/pneumothorax (CC 117)	14,000 (1.4)	1 (0.92, 1.08)

Variable	Combined IP/OP THA/TKA N=989,587	*
Dialysis status (CC 134)	1,962 (0.2)	1.4 (1.19, 1.65)
Renal failure (CC 135-140)	150,937 (15.3)	1.21 (1.17, 1.25)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	19,670 (2)	1.36 (1.28, 1.45)
Trauma (CC 166-168, 170-173)	45,092 (4.6)	1.21 (1.15, 1.27)
Vertebral fractures without spinal cord injury (CC 169)	10,119 (1)	1.2 (1.09, 1.31)
Other injuries, modified (CC 174Y)	250,156 (25.3)	1.11 (1.08, 1.14)
Major complications of medical care and trauma (CC 176-177)	45,321 (4.6)	1.29 (1.23, 1.35)
Morbid obesity (CC 22)	94,620 (9.6)	1.52 (1.47, 1.58)
Other congenital deformity of hip (joint)	87,699 (8.9)	1.1 (1.06, 1.15)
Post-traumatic osteoarthritis	13,644 (1.4)	1.11 (1, 1.22)

Table 6. Final Risk Model Variables and Adjusted Odds Ratios (Logistic Regression Model)

*cell intentionally left empty.

[Response Ends]

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

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

[Response Begins]

Although we analytically aim to separate these pathways to the extent possible, we acknowledge that risk factors often act on multiple pathways, and, as such, individual pathways can be complex to distinguish analytically. Further, some social risk factors, despite having a strong conceptual relationship with worse outcomes, may not have statistically meaningful effects on the risk model. They also have different implications on the decision to risk adjust or not.

Below we characterized each of the three social risk factors (Black race, ASI, Dual-eligibilty) as follows:

- Prevalence of social risk factors among patients in the cohort
- Distribution of patients with social risk factors across hospitals
- Observed outcomes
- Comparison of distribution of measure scores for facilities with low- and high proportion of patients with social risk factors
- Odds ratios for the outcome, both bivariate and multivariate
- Relationship between measure scores and facility-proportion of social risk factors
- Correlations between and differences in measure scores calculated with and without social risk factors
- Social risk factor impact on model C-statistic
- Risk prediction (risk decile plots) for each social risk factor

First, we examined the prevalence and observed outcome rates for each social risk factor (Tables 7, 8, and 9).

Table 7, below, presents results of the analyses representing the impact of exposure to racism using data from the combined IP and OP cohort of the IP/OP 90-Day THA/TKA Complication measure, broken down by race. The results for the combined IP/OP sample show that the unadjusted complication rate for Black beneficiaries (2.92%) was similar to that of White beneficiaries (2.91%), representing a difference (0.01%). The difference in complication rate between White and Other Non-White beneficiaries was larger (-0.28%), with Other Non-White beneficiaries having a lower complication rate compared to their White counterparts.

Exposure to racism	*	*	*	*
*	*	**IP (N=644,737)**	**OP (N=322,338)**	**Both (N=967,075)**
No. of procedures (%)	White	596,281 (92.5)	301,525 (93.5)	897,806 (92.9)
*	Black	30,452 (4.7)	12,158 (3.8)	42,610 (4.4)
*	Other non-White	18,004 (2.8)	8,655 (2.7)	26,659 (2.7)
Unadjusted complication rate	White	3.42%	1.89%	2.91%
*	Black	3.39%	1.76%	2.92%
*	Other non-White	3.04%	1.76%	2.63%
Difference from White	White	REF	REF	REF
*	Black	-0.03%	-0.13%	0.01%
*	Other non-White	-0.38%	-0.13%	-0.28%

Table 7. Prevalence and Observed Complication Rate by Racial Group

*cells intentionally left empty

Table 7 also presents results from analyses for THA/TKA procedures performed in the IP-only and OP-only settings, respectively, broken down by race. Complication rates were highest among the IP-only group and lowest among the OP-only group, regardless of race. Complication rates for Black and Other Non-White race patients were lower (better) than for White patients across all settings except for Black patients in the combined IP/OP setting (where Blacks had an observed complication rate similar to White patients, as noted above).

Similarly, the results of the OP-only analysis show that the unadjusted complication rate for Black and Other Non-White beneficiaries was also lower than that of White beneficiaries (1.76% vs. 1.89%, respectively), representing a larger difference (-0.13%), than what was found in the combined IP/OP analyses. It is worth noting that the findings of lower complication rates for Black and Other Non-White beneficiaries may reflect a number of factors, included restricted access to THA/TKA surgery among persons of color and their overall underrepresentation in patients undergoing THA/TKA procedures.

Table 8 presents the results of the social risk factor analyses for dual eligibility using data from the same IP/OP THA/TKA cohort. The results show that Medicare beneficiaries in the combined IP/OP cohort with dual eligibility status have higher complication rates (4.33%) compared to their counterparts without dual eligibility status (2.83%). Similarly, the results of the analyses on procedures performed in the IP-only and OP-only settings broken down by dual eligibility status, showed that dual eligibility was associated with higher observed THA/TKA complication rates in both the IP and OP settings, compared to their non-dual eligible counterparts in each respective setting. However, the difference was largest in the IP setting (4.88% vs. 3.33%).

Dual-eligibility	*	*	*	*
*	*	IP (N=658,346)	OP (N=331,157)	Both (N=989,503)
No. of procedures (%)	Full dual	25,150 (3.8)	7,837 (2.4)	32,987 (3.3)
*	Non/partial dual	633,196 (96.2)	323,320 (97.6)	956,516 (96.7)
Unadjusted complication rate	Full dual	4.88%	2.54%	4.33%
*	Non/partial dual	3.33%	1.85%	2.83%
Difference from non/partial dual	Full dual	1.55%	0.69%	1.50%
*	Non/partial dual	REF	REF	REF

Table 8. Prevalence and Observed Complication Rate by Dual-Eligibility

*cells intentionally left empty

Table 9 presents results of the social risk factor analyses comparing the unadjusted complication rates for Medicare beneficiaries who have a low AHRQ SES Index (ASI) score (first quartile of ascending-ordered ASI) versus Medicare beneficiaries with a Higher ASI (second-fourth quartile of ascending-ordered ASI). The results of the combined IP/OP model show that beneficiaries with a lower ASI score have a higher complication rate (3.20%) compared to beneficiaries with a higher ASI score (2.84%). Similar findings were also seen in the IP-only and OP-only setting analyses.

Low AHRQ SES Index (ASI) (Q1 vs Q2-Q4)	*	*	*	*
Low AHRQ SES Index (ASI) (Q1 vs Q2-Q4)	*	*	*	*
*	*	IP (N=655,786)	OP (N=330,136)	Both (N=985,922)
No. of procedures (%)	Low ASI	73,088 (11.1)	33,972 (10.3)	107,060 (10.9)
*	Higher ASI	582,698 (88.9)	296,164 (89.7)	878,862 (89.1)
Unadjusted complication rate	Low ASI	3.80%	1.91%	3.20%
*	Higher ASI	3.34%	1.86%	2.84%
Difference from non/partial dual	Low ASI	0.46%	0.05%	0.36%
*	Higher ASI	REF	REF	REF

Table 9. Prevalence and Observed Complication Rate by AHRQ Socioeconomic Status Index (ASI)

*cells intentionally left empty

Next, we compared the distribution of measure scores across quartiles of facility-proportion of patients with social risk factors.

Table 10, Table 11, Figure 8, and Figure 9 show a comparison of the distribution of hospital-level risk standardized complication rate (RSCR) between hospitals with a higher proportion of Black beneficiaries (first quartile of ascending-ordered proportions) and hospitals with a lower proportion of Black beneficiaries (fourth quartile of ascending-ordered

proportions), as well as a comparison of the first and fourth quartiles for the facility proportion of other non-White beneficiaries. Among all the 3,452 hospitals, the mean RSCR of hospitals with fewer Black beneficiaries is 2.96%, slightly lower than the mean RSCR of hospitals with more Black beneficiaries (3.01%). Similarly, the mean RSCR of hospitals with fewer other Non-White beneficiaries is 2.96%, also slightly lower than the mean RSCR of hospitals with more other Non-White beneficiaries (2.98%). Among the 2,733 hospitals with at least 25 index procedures, the mean of RSCR of hospitals with fewer Non-White beneficiaries is also slightly lower than the mean of hospitals with more Non-White beneficiaries (2.99% vs 3.02% for White vs Black, 2.95% vs 2.99% for White vs Other Non-White). Even with these differences in mean values by SRF, the distributions overlap nearly completely (Figures 8 and 9).

Quantile	%Black	%other non-White	RSCR by quartile of proportion of Black	*	RSCR by quartile of proportion of other non-White combined	*
*	*	*	Q1 range of %Black (0-0):	Q4 range of %Black (6.073-100):	Q1 range of %other (0-0):	Q4 range of %other (3.54-100):
No. of hospitals	3,452	3,452	1,205	863	1,023	863
Mean	5.939	4.491	2.96	3.01	2.96	2.98
SD	12.577	11.027	0.352	0.467	0.323	0.472
Max	100	100	5.35	5.07	5.09	6.00
99%	68.86	60	4.20	4.44	4.08	4.44
95%	26.515	18.919	3.60	3.85	3.56	3.86
90%	15.926	10.227	3.41	3.60	3.37	3.54
75%	6.071	3.535	3.11	3.21	3.08	3.17
Median	1.442	1.316	2.89	2.92	2.89	2.88
25%	0	0	2.79	2.75	2.81	2.73
10%	0	0	2.61	2.52	2.64	2.52
5%	0	0	2.51	2.38	2.55	2.29
1%	0	0	2.26	1.94	2.34	2.00
Min	0	0	2.02	1.64	2.06	1.58

Table 10. Risk Standardized Complication Rate (RSCR) by Hospital's Proportion of Non-White THA/TKA Patients, including all hospitals

*cells intentionally left empty

Quantile	%Black	%other non-White	RSCR by quartile of proportion of Black	*	RSCR by quartile of proportion of other non-White combined	*
*	*	*	Q1 range of % Black (0-0):	Q4 range of % Black (5.92-90.32):	Q1 range of % other (0-0.54):	Q4 range of % other (3.36-100):

Quantile	%Black	%other non-White	RSCR by quartile of proportion of Black	*	RSCR by quartile of proportion of other non-White combined	*
No. of hospitals	2,733	2,733	704	683	683	684
Mean	4.941	3.558	2.99	3.02	2.95	2.99
SD	8.692	7.515	0.436	0.517	0.449	0.534
Max	90.323	100	5.35	5.07	5.35	6.00
99%	42.105	41.667	4.37	4.64	4.33	4.64
95%	20.287	12.5	3.76	3.95	3.67	3.95
90%	12.727	8	3.54	3.68	3.52	3.69
75%	5.917	3.361	3.20	3.29	3.19	3.25
Median	1.903	1.515	2.92	2.96	2.90	2.92
25%	0	0.541	2.69	2.67	2.65	2.66
10%	0	0	2.52	2.48	2.50	2.44
5%	0	0	2.41	2.27	2.33	2.23
1%	0	0	2.19	1.87	2.02	1.98
Min	0	0	2.02	1.64	1.64	1.58

Table 11. Risk Standardized Complication Rate (RSCR) by Hospital's Proportion of Non-White THA/TKA Patients, including only hospitals with at least 25 cases

*cells intentionally left empty

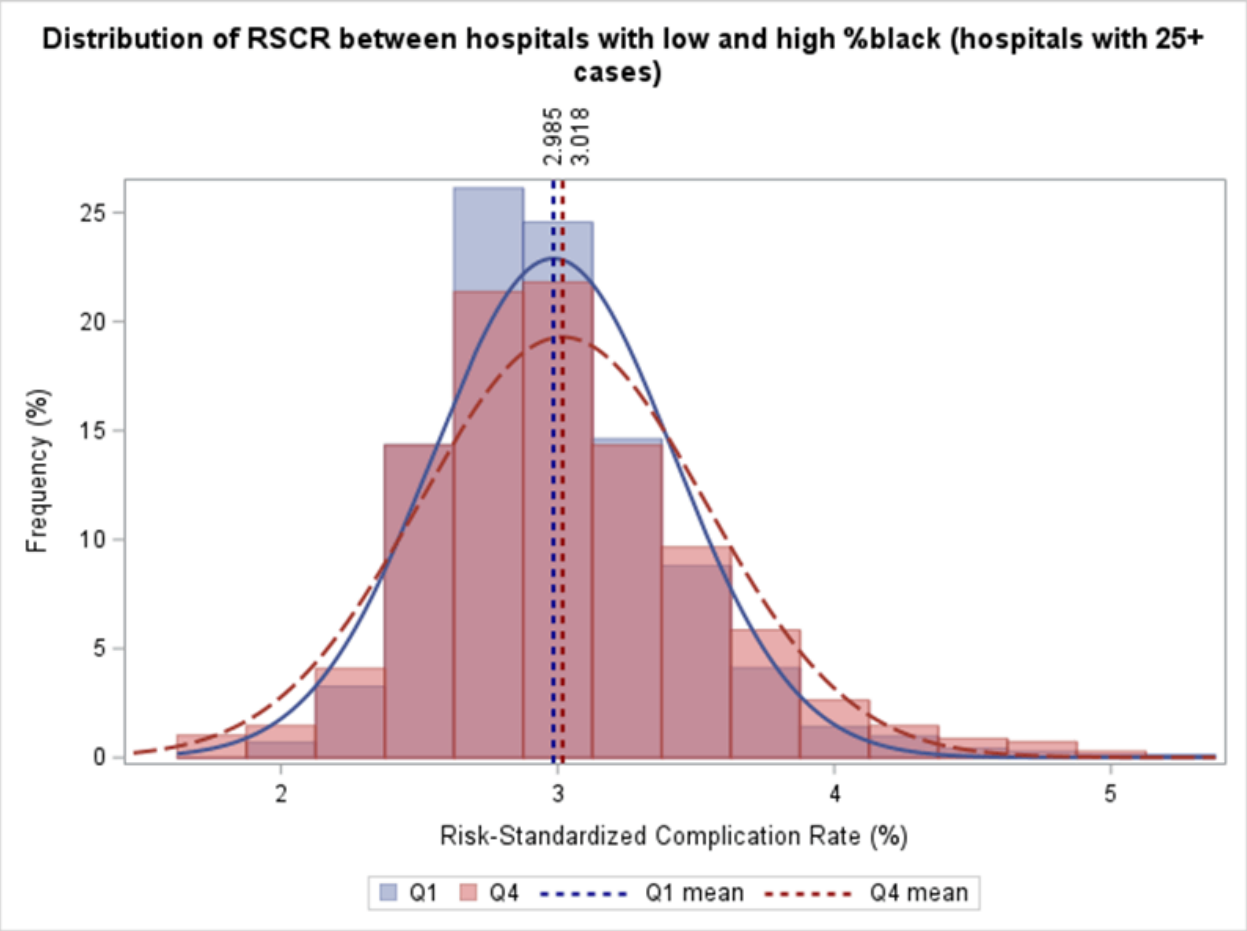

Figure 8. Distribution of RSCR between hospitals with low and high % Black (hospitals with 25+ cases)

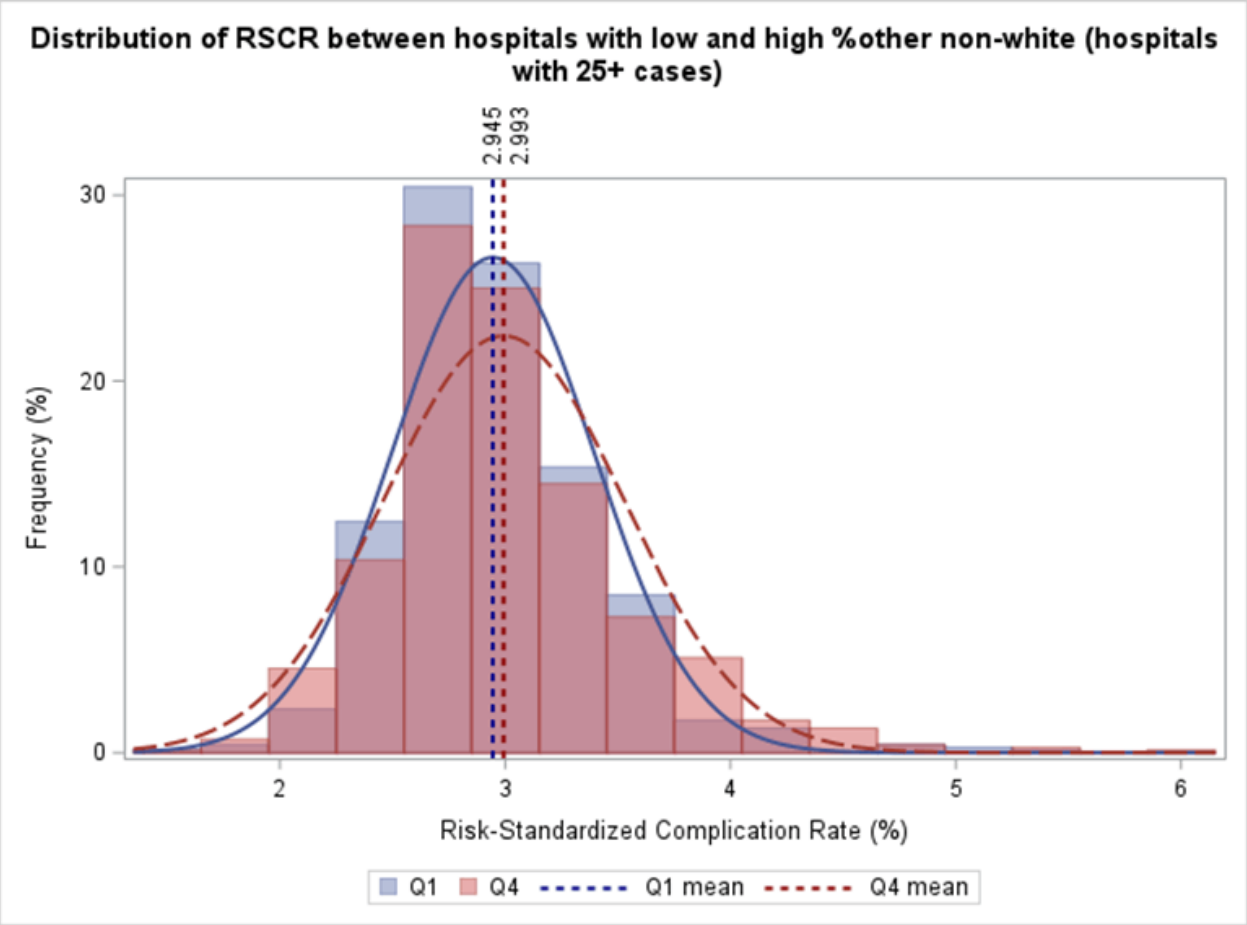

Figure 9. Distribution of RSCR between hospitals with low and high % Black (hospitals with 25+ cases)

Table 12, Table 13, and Figure 10 show the difference of hospital-level RSCR distribution between hospitals with a higher proportion of dual eligible beneficiaries (first quartile of ascending-ordered proportions) and hospitals with a lower proportion of dual eligible beneficiaries (fourth quartile of ascending-ordered proportions). Among all the 3,452 hospitals, the mean RSCR of hospitals with fewer dual eligible beneficiaries is 2.88%, lower than the mean RSCR of hospitals with more dual eligible beneficiaries (3.01%). Similarly, in the 2,747 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer dual eligible beneficiaries is still lower (2.86% vs 3.02%).

Quantile	% dual	RSCR by quartile of proportion of dual	*
*	*	Q1 range of %dual (0-0.97):	Q4 range of %dual (6.71-100):
No. of hospitals	3,452	863	863
Mean	7.013	2.88	3.01
SD	13.738	0.385	0.459
Max	100	5.29	5.86
99%	83.333	4.09	4.77
95%	28.571	3.52	3.87
90%	15.152	3.30	3.53
75%	6.667	3.04	3.15

Quantile	% dual	RSCR by quartile of proportion of dual	*
Median	2.785	2.86	2.87
25%	0.975	2.74	2.76
10%	0	2.47	2.57
5%	0	2.21	2.48
1%	0	1.86	2.12
Min	0	1.53	1.93

Table 12. Complication rate for all hospitals by dual-eligibility

*cells intentionally left blank

Quantile	% dual	RSCR by quartile of proportion of dual	*
*	*	Q1 range of %dual (0-1.33):	Q4 range of %dual (5.79-94):
No. of hospitals	2,747	684	687
Mean	5.216	2.86	3.02
SD	8.619	0.502	0.528
Max	94.000	5.29	5.86
99%	47.826	4.23	4.81
95%	17.188	3.72	4.04
90%	11.111	3.46	3.70
75%	5.792	3.14	3.24
Median	2.823	2.83	2.92
25%	1.333	2.54	2.67
10%	0.439	2.28	2.50
5%	0	2.06	2.34
1%	0	1.75	2.07
Min	0	1.53	1.67

Table 13. Complication rate by dual-eligibility only including hospitals with at least 25 cases

*cells intentionally left empty

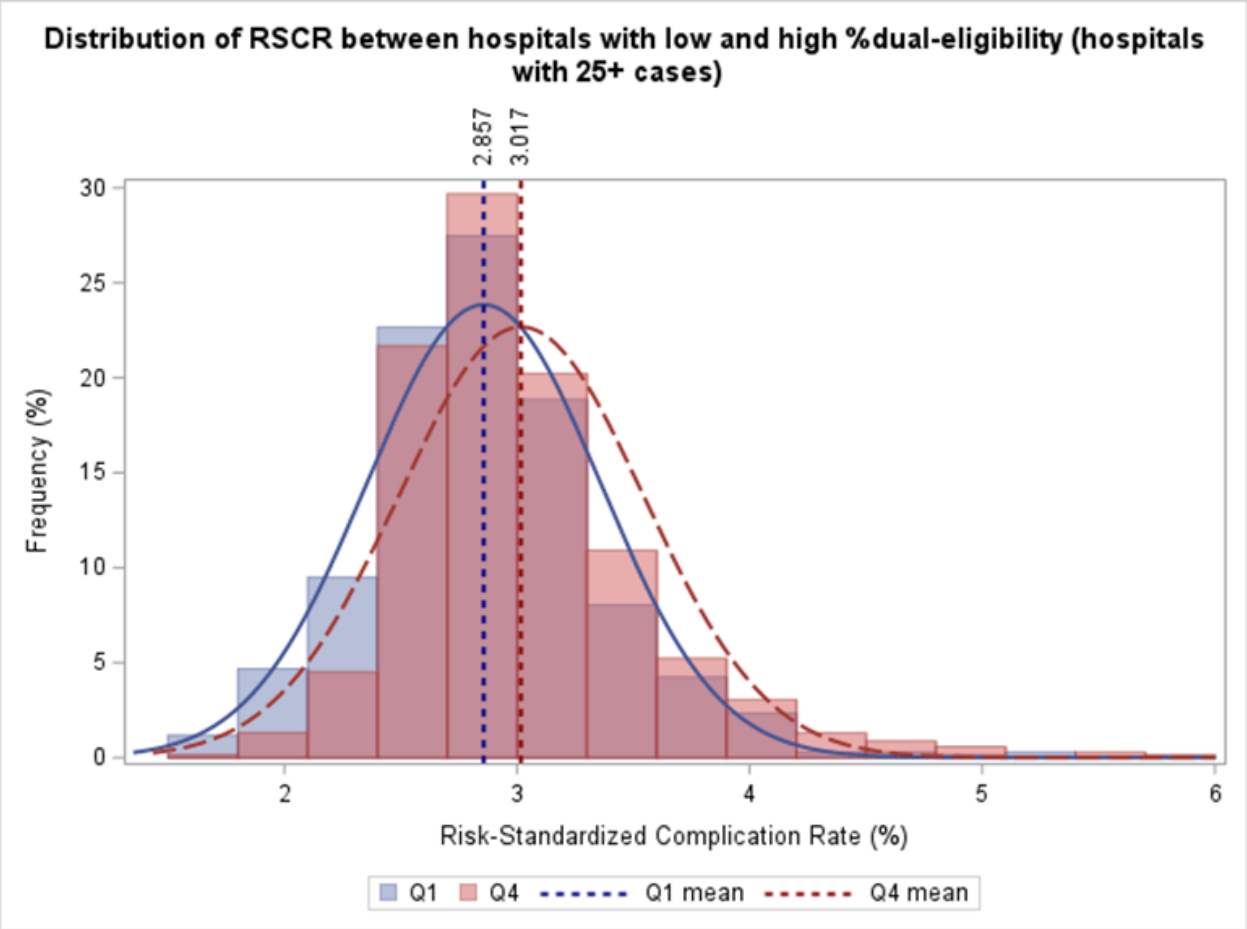

Figure 10. Distribution of RSCR between hospitals with low and high % dual eligibility (hospitals with 25+ cases)

Table 14, Table 15, and Figure 11 show the difference of RSCR distribution between hospitals with a higher proportion of beneficiaries with lower AHRQ SES Index (ASI; first quartile of ascending-ordered proportions) and hospitals with a lower proportion of beneficiaries with lower ASI (fourth quartile of ascending-ordered proportions). Among all the 3,452 hospitals, the mean RSCR of hospitals with fewer beneficiaries who have lower ASI is 2.87%, lower than the mean RSCR of hospitals with more beneficiaries who have lower ASI (2.98%). Similarly, in the 2,740 hospitals with at least 25 index procedures, the mean RSCR of hospitals with fewer beneficiaries who have lower ASI is still lower (2.85% vs 3.00%).

Quantile	% low ASI	RSCR by quartile of proportion of low ASI	*
*	*	Q1 range of %low ASI (0-4.51):	Q4 range of %low ASI (22.99-100):
No. of hospitals	3,424	856	856
Mean	16.668	2.87	2.98
SD	17.932	0.46	0.388
Max	100	5.70	5.40
99%	100	4.28	4.23
95%	52.381	3.71	3.71
90%	40	3.44	3.48
75%	22.97	3.05	3.14

Quantile	% low ASI	RSCR by quartile of proportion of low ASI	*
Median	10.737	2.85	2.89
25%	4.512	2.63	2.78
10%	1.02	2.37	2.58
5%	0	2.21	2.48
1%	0	1.83	2.16
Min	0	1.54	1.86

Table 14. Complication rate for all hospitals by AHRQ SES Index

*cells intentionally left empty

Quantile	% dual	RSCR by quartile of proportion of dual	*
*	*	Q1 range of %dual (0-4.80):	Q4 range of %dual (20-81.08):
No. of hospitals	2,740	684	686
Mean	14.477	2.85	3.00
SD	13.528	0.518	0.478
Max	81.081	5.70	5.40
99%	61.702	4.41	4.40
95%	42.796	3.81	3.84
90%	33.942	3.50	3.61
75%	20	3.10	3.25
Median	10.212	2.79	2.94
25%	4.808	2.53	2.69
10%	2.174	2.28	2.49
5%	1.11	2.15	2.31
1%	0	1.70	1.99

Table 15. Complication rate by AHRQ SES Index only including hospitals with at least 25 cases

*cells intentionally left empty

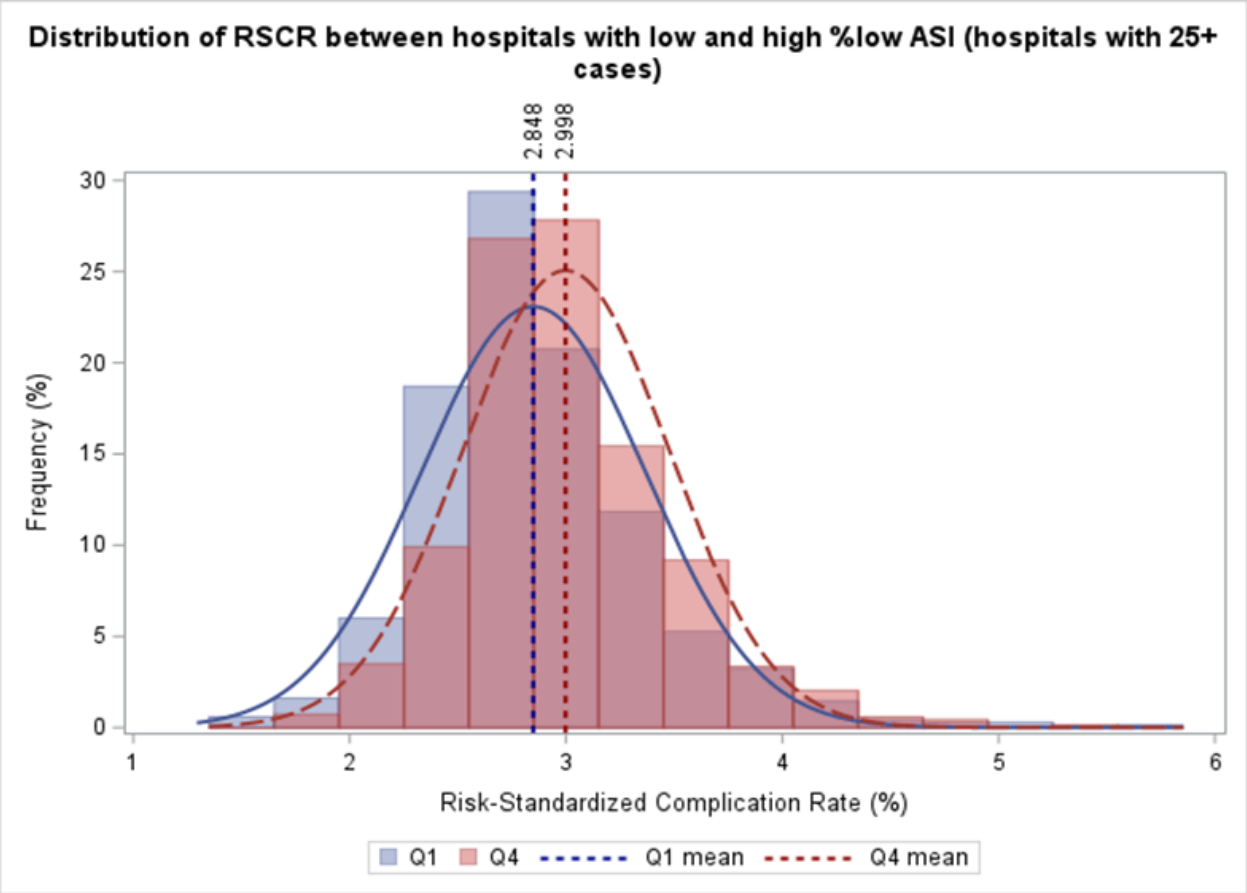

Figure 11. Distribution of RSCR between hospitals with low and high % ASI (hospitals with 25+ cases.)

We then examined odds ratios for the outcome for each social risk factor alone, and in the presence of all of the other variables in the risk model.

Table 16 presents the odds ratios and corresponding 95% confidence intervals for the SRFs described in Tables 7-9 which were estimated using a univariate patient-level model and a multivariate hierarchical model. The results of both the univariate and hierarchical models show that the Black-White Exposure to Racism contrast was not significantly related to complications; the significantly greater odds of a complication found for the Other-Non-White vs. White comparison in the univariate model was attenuated and no longer significant in the hierarchical multivariate model. However, the findings for the dual-eligibility and low-ASI-score contrasts, respectively, showed that both groups had significantly greater (although attenuated) odds of experiencing a complication compared to their counterparts after adjusting for age, sex, and clinical risk factors.

*	From univariate patient-level model	From multivariate hierarchical model (adjusted for age, sex, clinical risk factors and accounting for clusters within hospitals)
Social Risk Factor	OR (95% CI)	OR (95% CI)
Exposure to racism (Black vs White)	1.01 (0.95-1.07)	0.96 (0.90-1.02)
Exposure to racism (other non-White vs White)	0.90 (0.84-0.97)	0.98 (0.91-1.06)
Dual-eligibility (dual vs non-dual)	1.55 (1.47-1.64)	1.19 (1.12-1.26)
Low AHRQ SES Index (ASI) (Q1 vs Q2-Q4)	1.13 (1.09-1.17)	1.06 (1.02-1.10)

Table 16. Association between SRF & Complications

*cell intentionally left blank

We then examined the relationship between social risk factors and facility-proportion of patients with social risk factors

Figures 12 and 13 show the association between the facility-proportion of patients with social risk factors and measure scores, for dual-eligibility (Figure 12) and low ASI (Figure 13) variables. We limited the analyses to these two variables because they were the only variables that were significantly associated with the outcome in a multivariable analysis (Table 16). The results show that for both social risk factors, there no significant relationship between measure scores and the facility-proportion of patients with social risk factors, for facilities with the highest proportion (highest quartile) of patients with social risk factors.

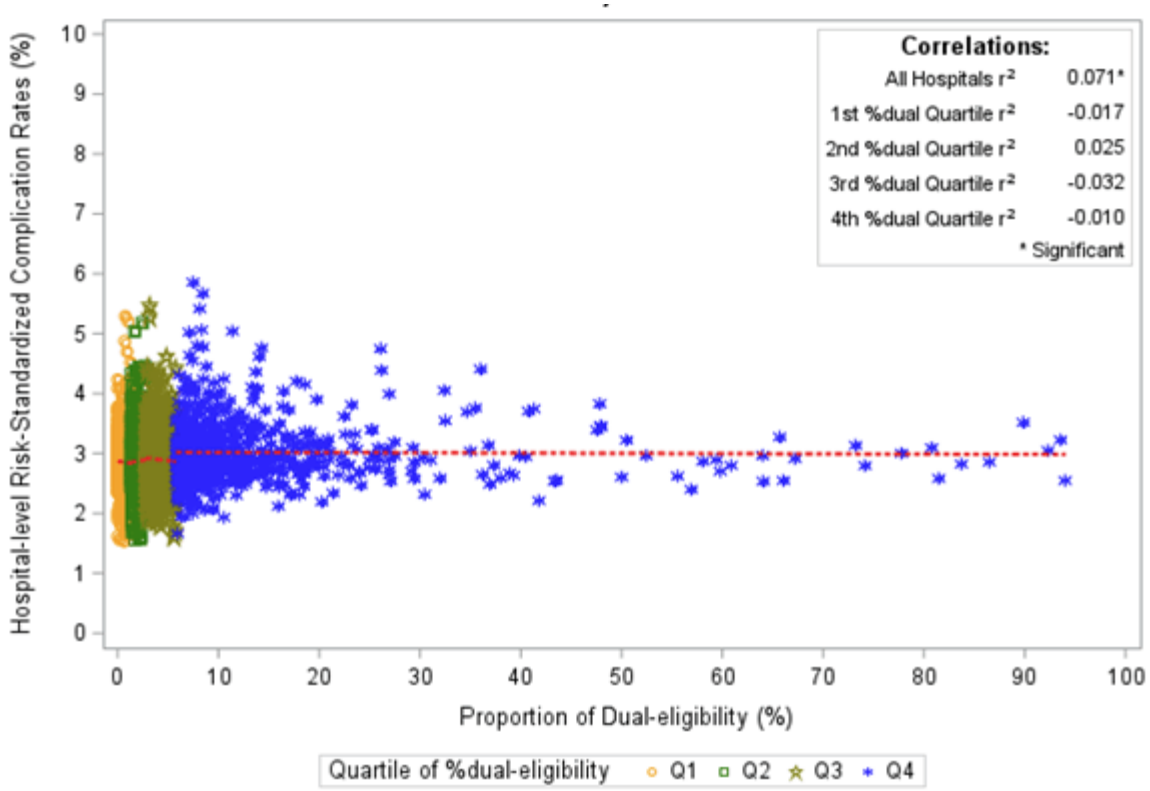

Figure 12. Association between RSCR and Dual Eligibility

*Significant, p<.05

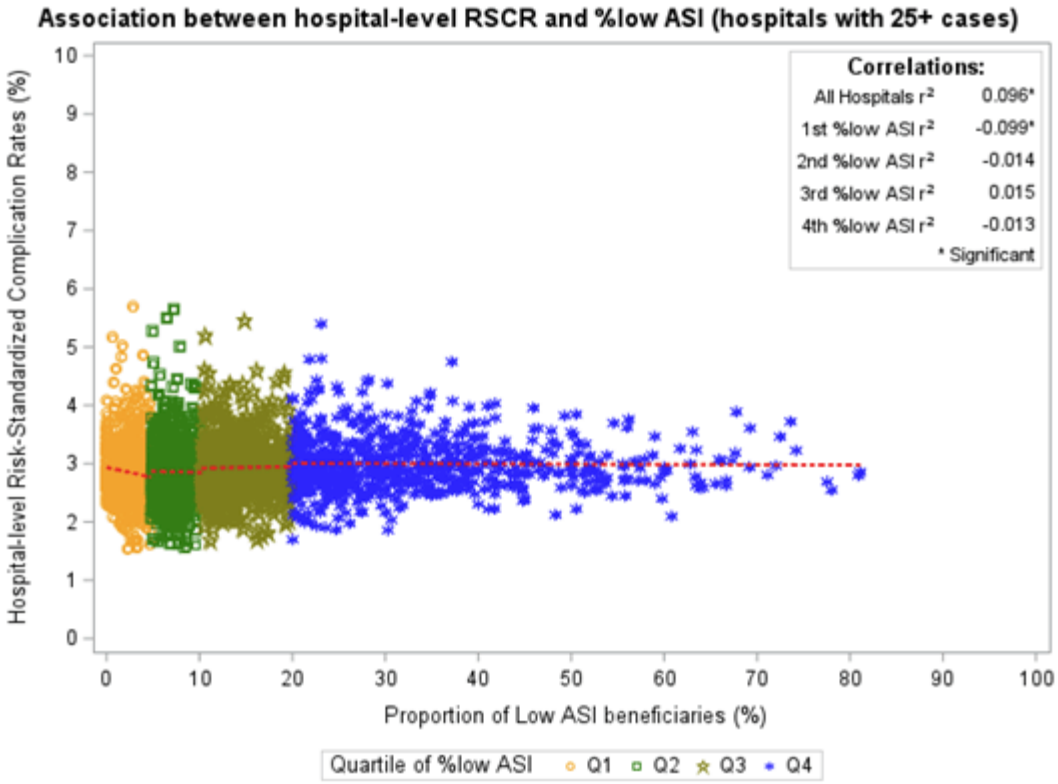

Figure 13. Association between RSCR and ASI

*Significant, p<.05

We then examined the impact of adjusting for each social risk factor on measure scores.

Figures 14-19 show the potential change in RSCRs across hospitals with at least 25 index procedures if social risk factors were included in the model separately. In all the histograms, the potential change in RSCRs is centered around 0, which indicates that the inclusion of social risk factor variables in the model does not drastically affect the RSCRs across hospitals with at least 25 index procedures. The scatter plots show that the Spearman rho correlation coefficients between RSCRs without social risk factor variables and RSCRs with social risk factor variables are all close to 1, with p-value < 0.001, which demonstrates that the inclusion of social risk factor variables in the model would not have a large impact on hospital-level measure results.

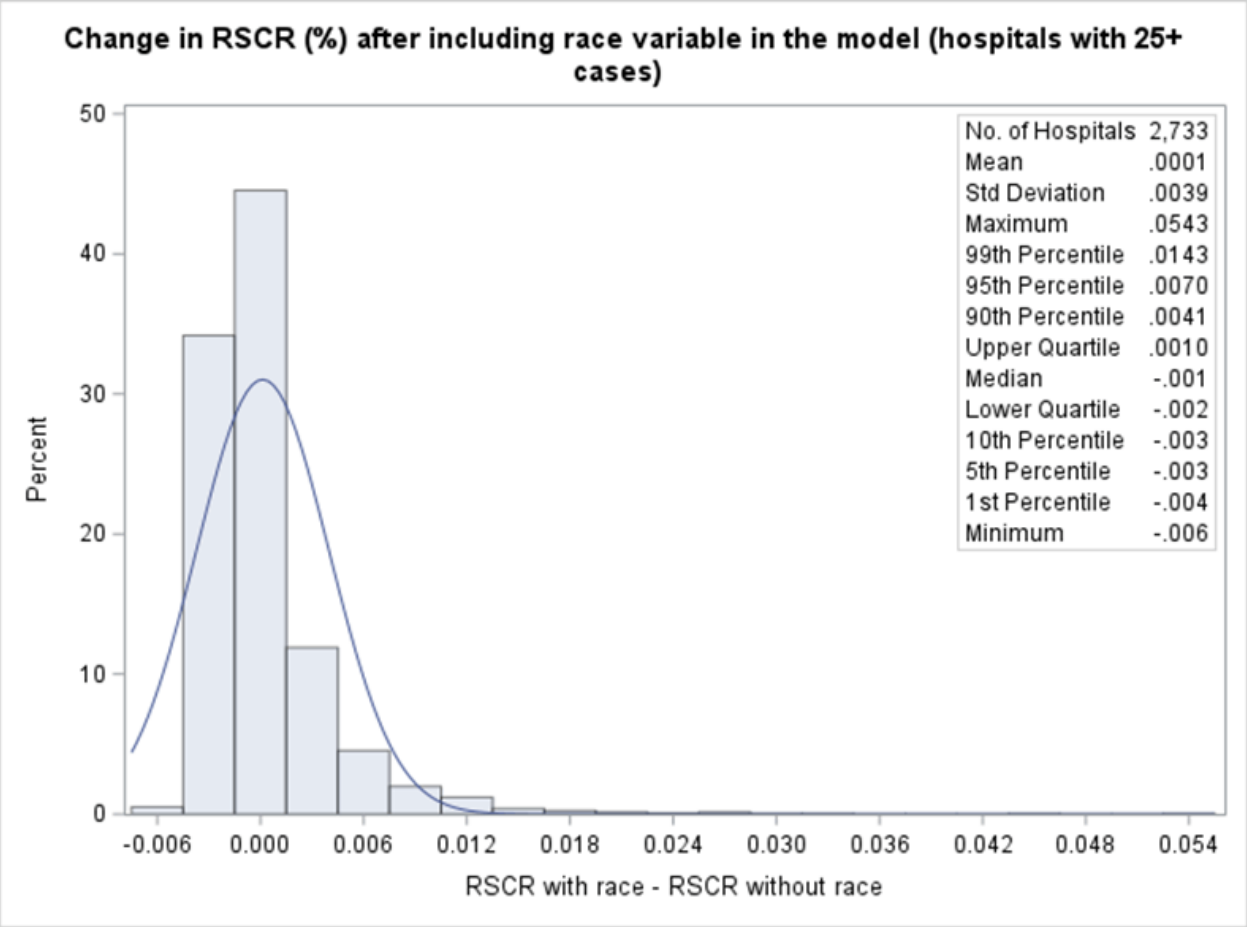

Figure 14. Change in RSCRs calculated with and without race (Black) in the model (for hospitals with at least 25 cases)

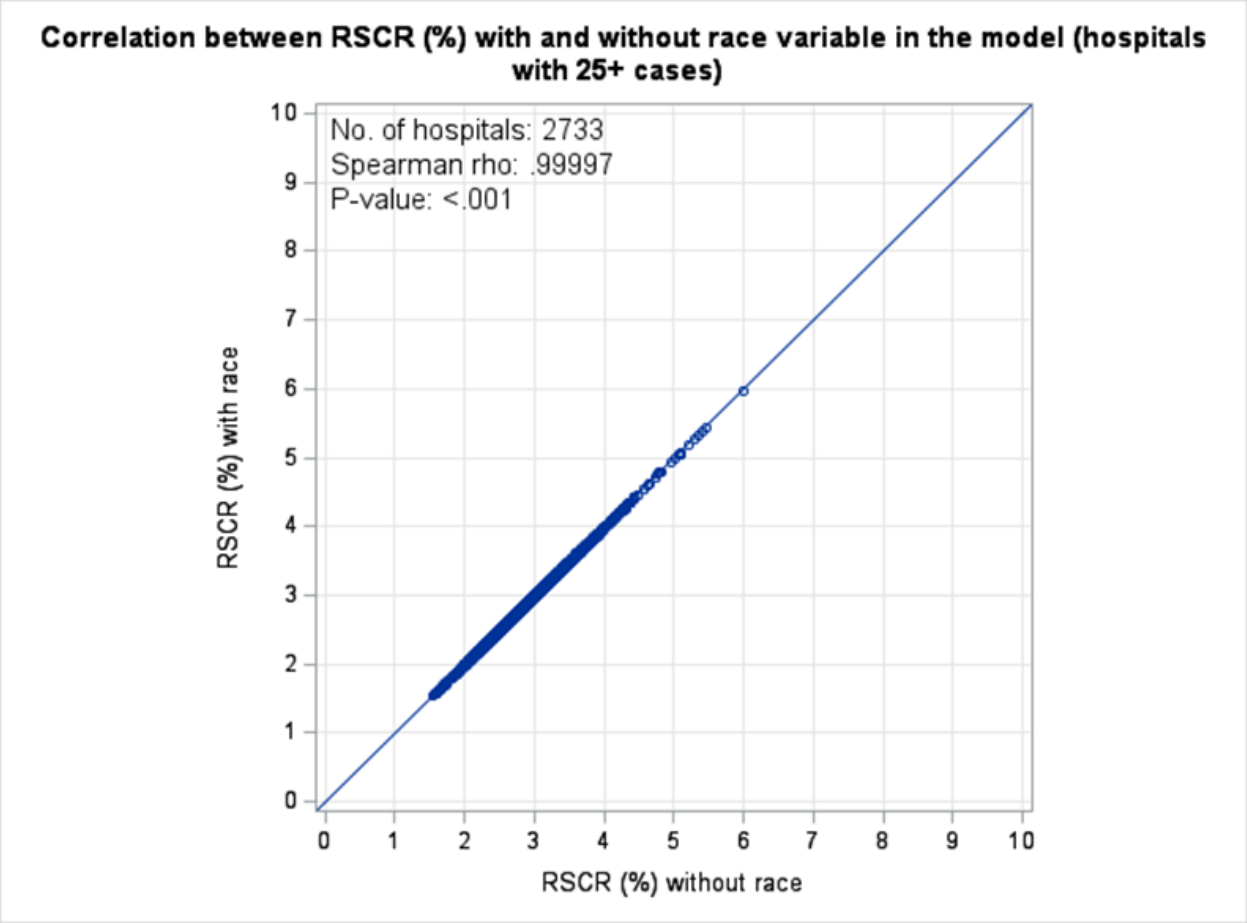

Figure 15. Correlation between RSCRs calculated with and without the race (Black) variable in the model (for hospitals with at least 25 cases)

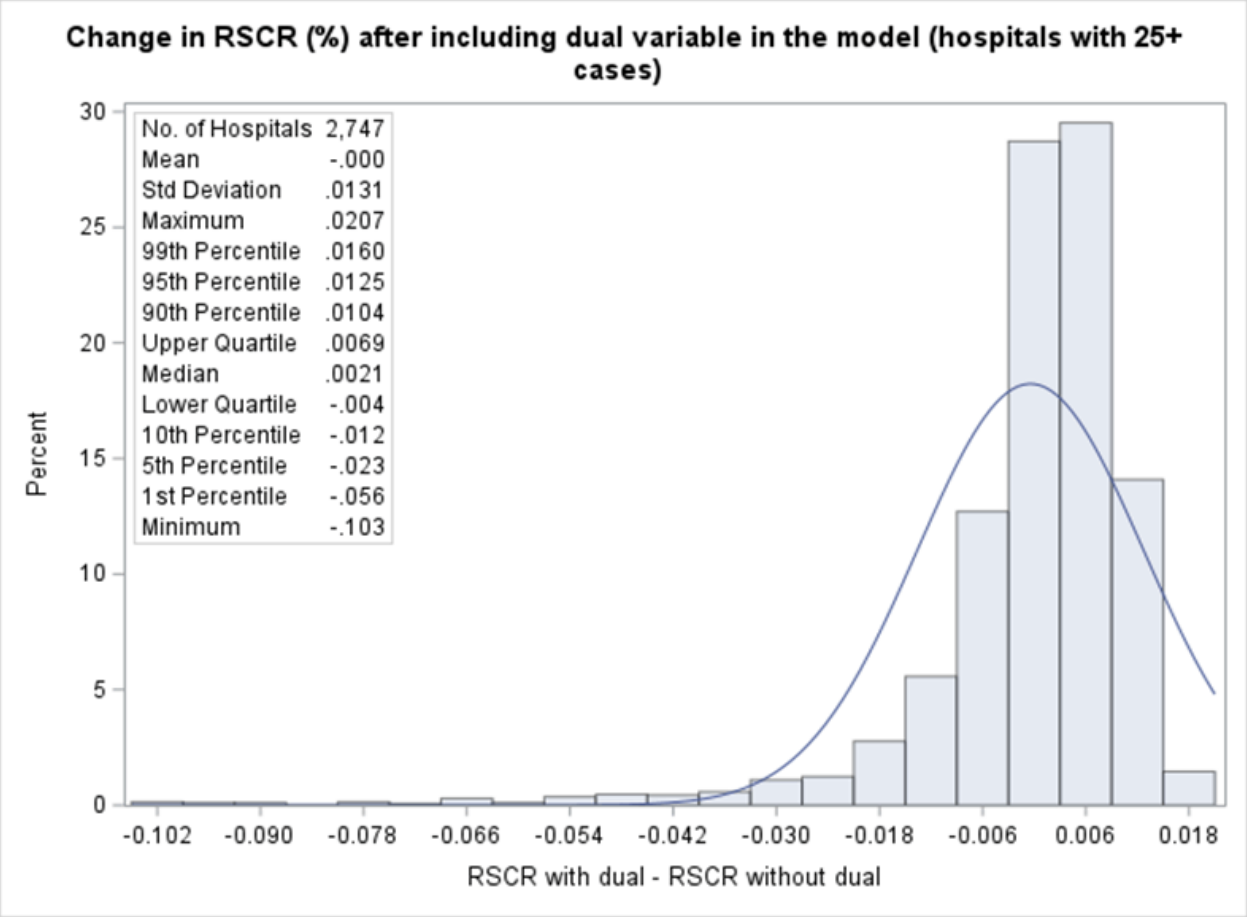

Figure 16. Change in RSCRs calculated with and without dual eligibility in the model (for hospitals with at least 25 cases)

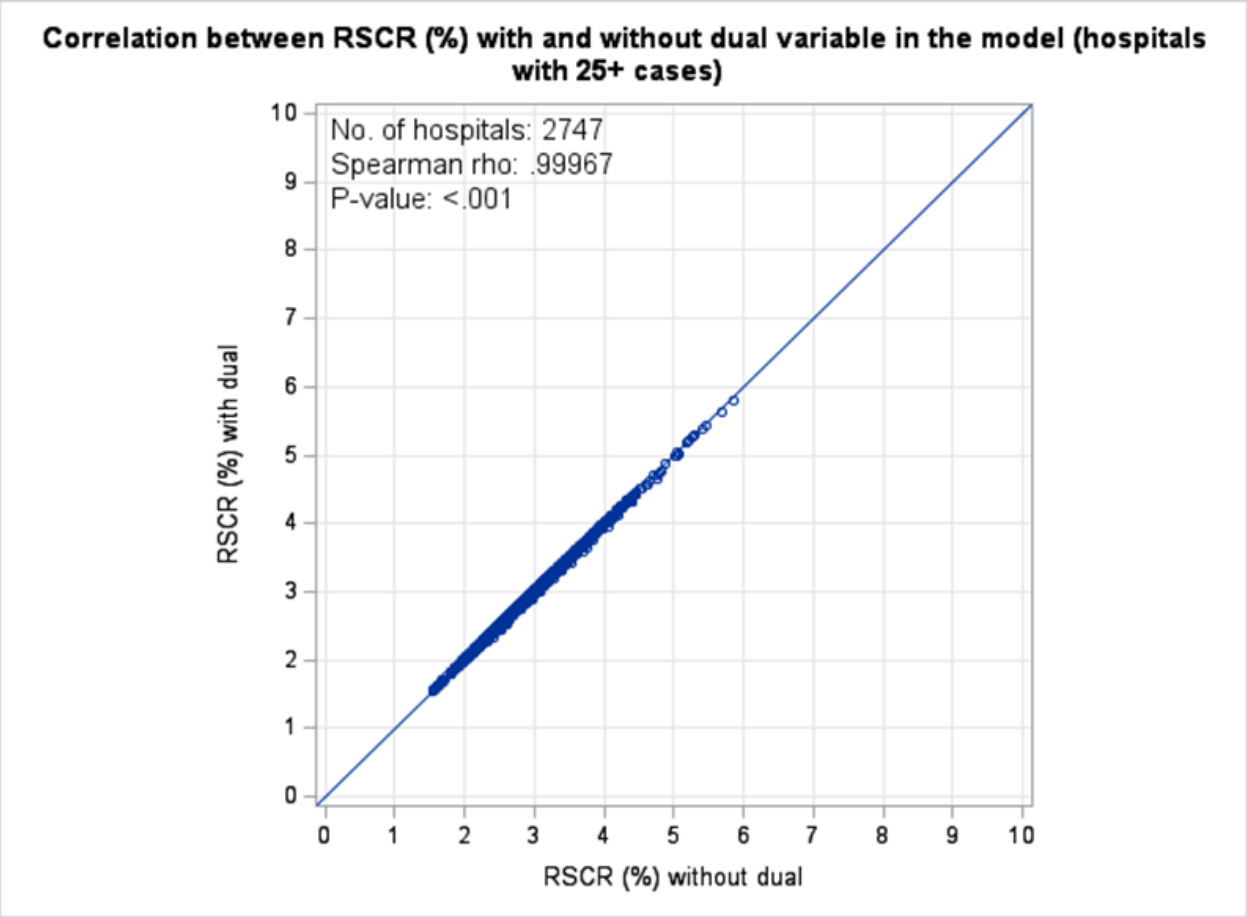

Figure 17. Correlation between RSCRs calculated with and without the dual eligibility variable in the model (for hospitals with at least 25 cases)

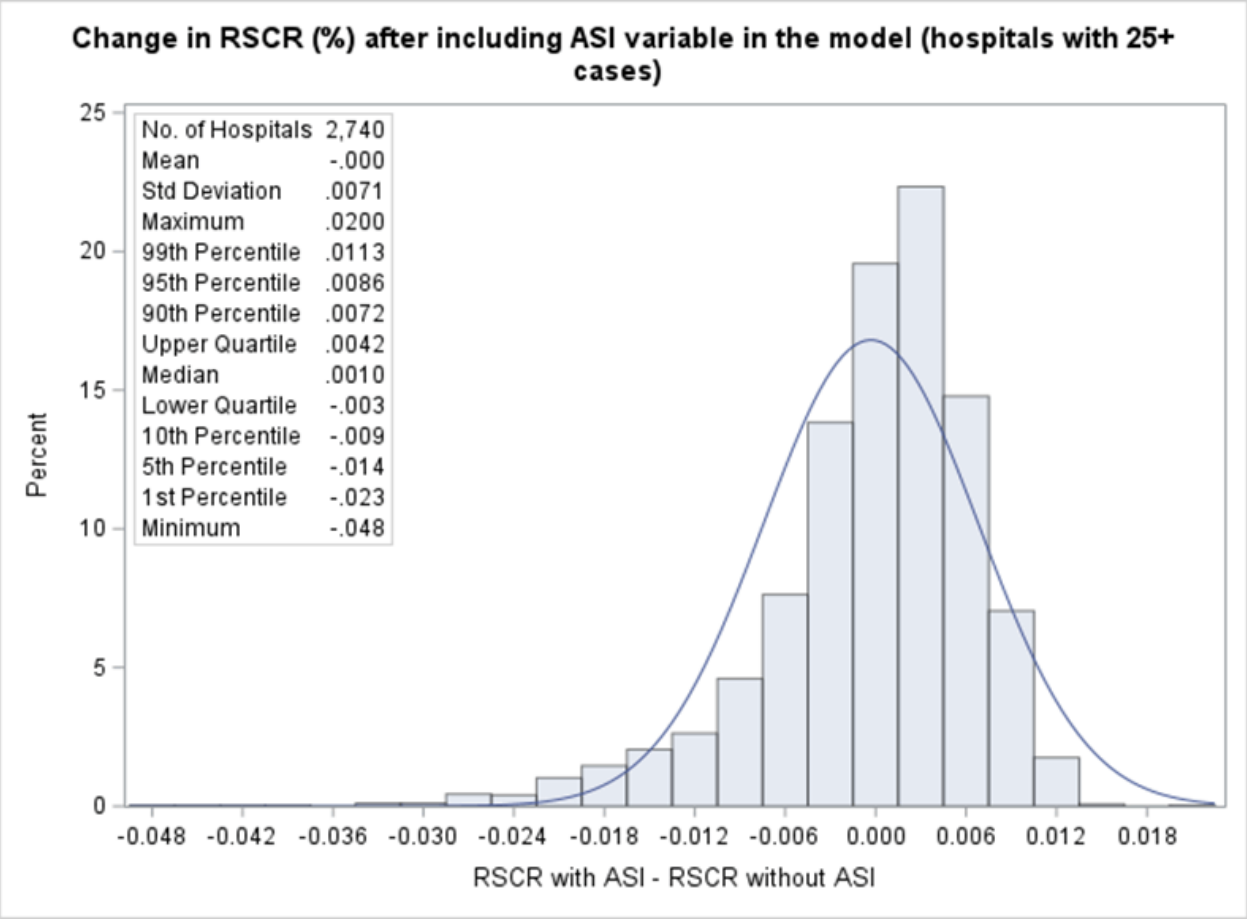

Figure 18. Change in RSCRs calculated with and without the ASI variable in the model (for hospitals with at least 25 cases)

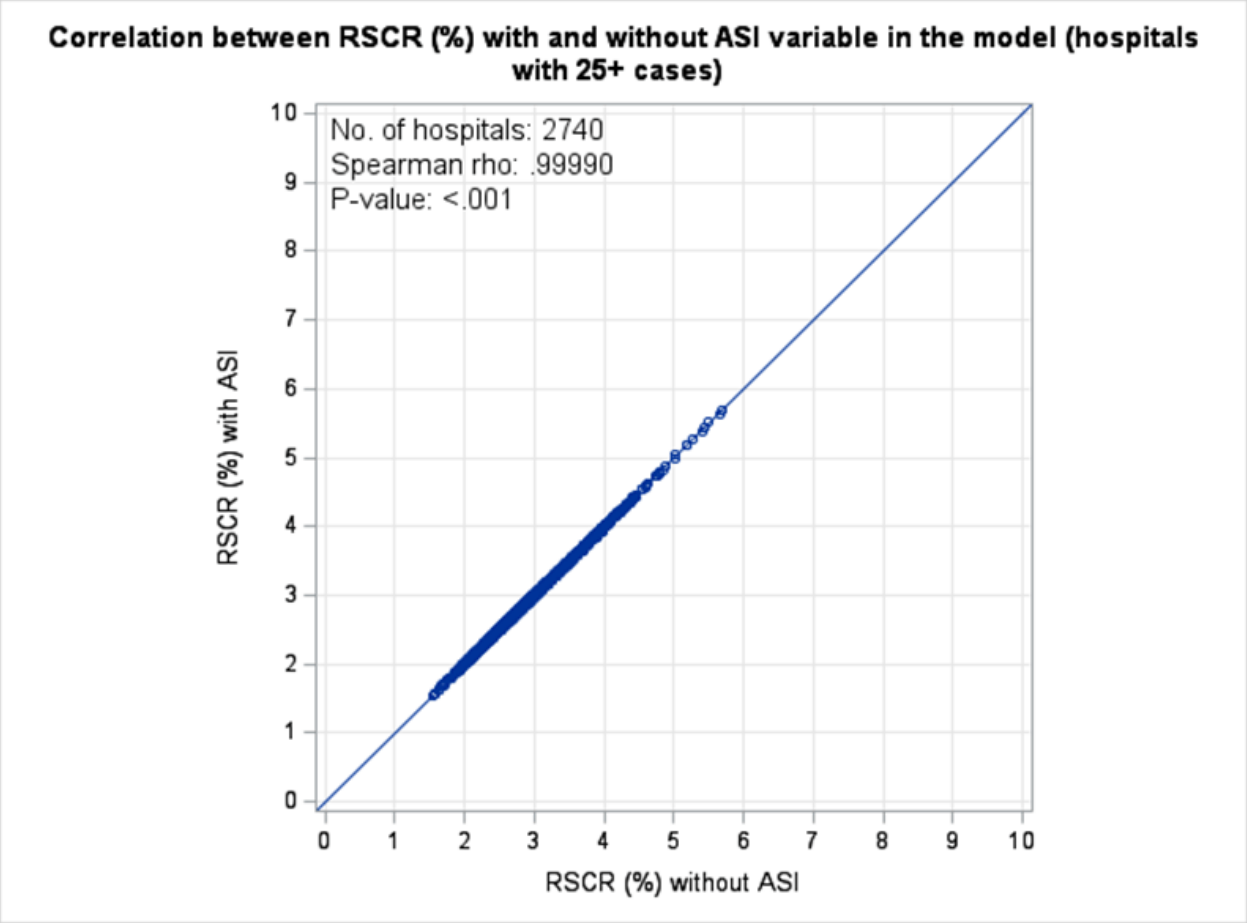

Figure 19. Correlation between RSCRs calculated with and without the ASI variable in the model (for hospitals with at least 25 cases)

Social risk factor adjustment summary

In summary our analyses show that there is a relatively low prevalence of patients with social risk factors in the measure cohort and that facilities with the highest proportions of patients with social risk factors are not unfairly characterized by the measure. Specifically:

- The prevalence of patients with social risk factor variables in the cohort is lower for race (Black) and dual eligibility (4.4% and 3.5%, respectively), compared with the ASI variable (10.9%).
- Unadjusted (observed) complication rates were:
    - Lower or similar for patients of Black or other non-White race, compared with White race
    - Higher for patients with dual eligibility or low ASI
- Odds ratios for the outcome in the presence of all of the risk-model variables were:
    - Significant, and less than 1 for the Black variable
    - Significant, and more than 1 for the DE and low ASI variables
- The risk model shows good risk prediction for subsets of patients with social risk factors (see section 2b.27 below, Figures 23-25) with the exception of the dual-eligibility risk variable, where the model underpredicts risk; however measure scores for facilities with the highest proportion of patients with dual eligibility are not systematically higher (see below).
- For all social risk factor variables, measure scores calculated with and without the social risk factors were highly correlated, with small differences.
- The distribution of measure scores between the lowest (first quartile) vs. the fourth quartile for the facility proportion of patients with social risk factors overlapped for all social risk factors.

- There was no correlation between measure scores and the facility-proportion of patients for either the dual eligibility or ASI variables for facilities with the highest proportion of patients with each social risk factor. (There was also no correlation for the race [Black] variable, data not shown.)

The decision to adjust for social risk is a complex decision and influenced by the intended measure application. In addition, CMMI has indicated a desire to maintain alignment with the original inpatient-only measure (NQF #1550). After examining the empiric analytic results and receiving feedback for the TEP, CMS/CMMI has decided not to risk adjust the measure for SRFs but will consider accounting for social risk through stratification by race and/or social risk at the time of implementation. This could include accounting for social risk in any payment calculations through measure score adjustment, measure score stratification, payment adjustment and/or stratification by dual eligibility, as CMS has done in other payment programs (such as the Hospital Readmission Reduction Program or HRRP) and CMMI has done in payment models (such as the Comprehensive Care for Joint Replacement, or CJR model). CMMI recently began accounting for dual eligibility in its CJR payment model, which uses the existing inpatient only THA/TKA complication measures upon which this specification is based. We therefore anticipate any future decision by CMS/CMMI to implement this current measure would similarly account for dual eligibility and/or other drivers of social health within the payment model to avoid unintended consequences of measurement while simultaneously preserving its ability to track disparities in outcomes with a measure not adjusted for social drivers of health."

We believe the absence of significant racial disparities in these data reflect the complex realities of how patients of non-White race or with social risk factors access elective procedures such as THA/TKA. CMS/CMMI are committed to implementing the measure to prevent unintended consequences of worsening access and/or outcome disparities among patients undergoing THA/TKA procedures.

[Response Ends]

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

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

[Response Begins]

Approach to Assessing Model Performance

We computed two summary statistics for assessing model performance (Harrell and Shih, 2001) for the expanded cohort:

Discrimination Statistics

(1) Area under the receiver operating characteristic (ROC) curve (the C-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile)

References

1. Harrell FE and Shih YC, Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* 17 (2001), pp. 17–26.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

Model performance statistics for the risk model are provided in Table 17.

- C-statistic for the risk model is 0.66.
- Predicative ability from the lowest to highest decile is 1.18-7.10%

*	Risk factors included: same risk factors as in Hospital HK-C measure	*	*	Risk factors included: same risk factors as in Hospital HK-C measure **plus** setting of procedure (inpatient vs outpatient)
*	Single model combining inpatient and outpatient	Separate model for inpatient	Separate model for outpatient	Single model combining inpatient and outpatient
C-statistic	0.659	0.651	0.638	0.664
(95% CI)	(0.656, 0.662)	(0.648, 0.655)	(0.631, 0.646)	(0.661, 0.667)
Predictive Ability[1]	1.180% - 7.10%	1.438% - 8.01%	0.880% - 4.28%	1.034% - 7.191%

Table 17. Model Performance

[1]Observed complication rate: First decile (%) – Last decile (%)

*cells intentionally left empty

[Response Ends]

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

[Response Begins]

Please see section 2b.29 below for detailed information on model calibration.

[Response Ends]

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

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

[Response Begins]

Figures 20-22 show risk decile plots for inpatient and outpatient combined, inpatient only, and outpatient only.

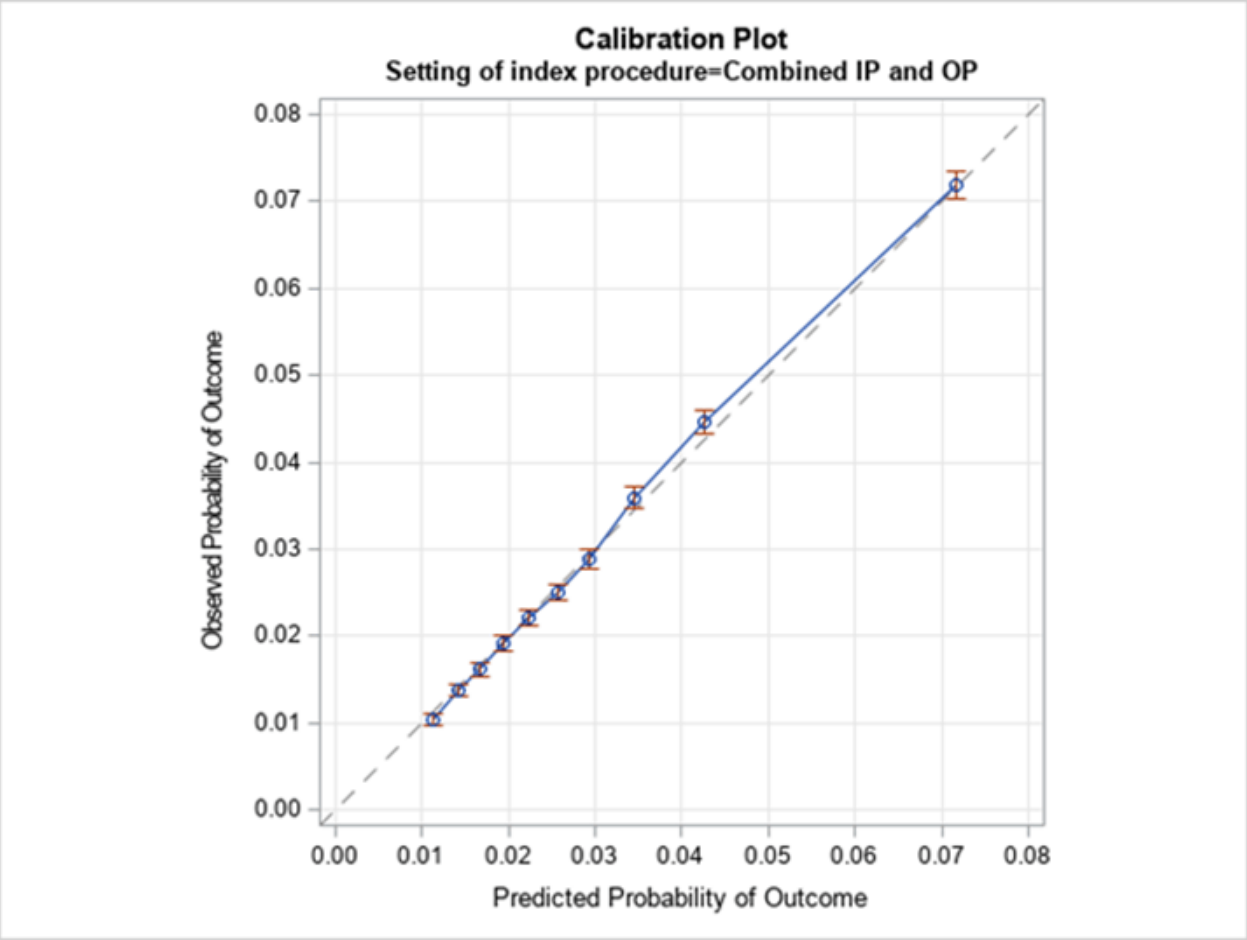

Figure 20. Calibration plot for all procedures (hospital inpatient and outpatient)

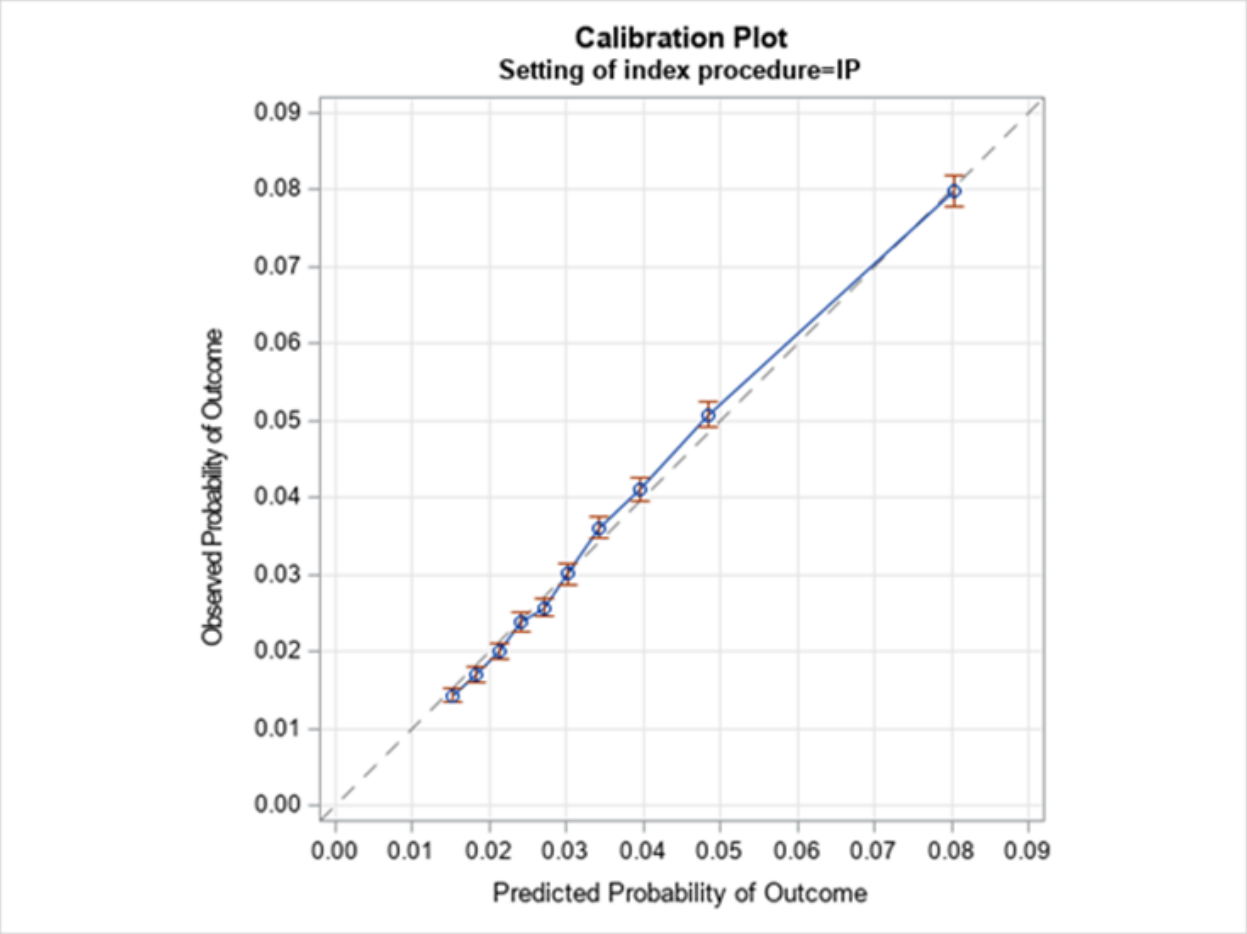

Figure 21. Calibration plot for inpatient procedures only

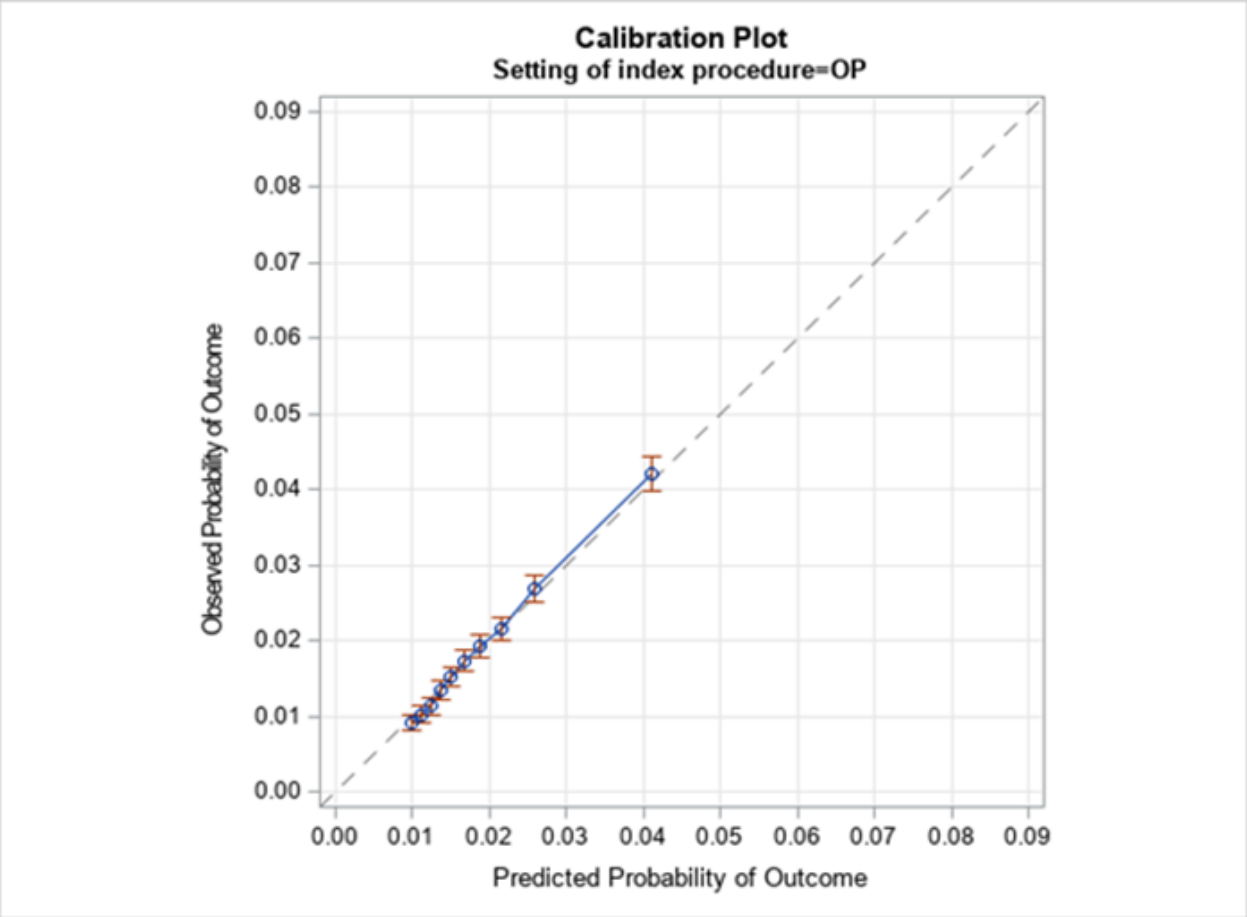

Figure 22. Calibration plot for outpatient procedures only

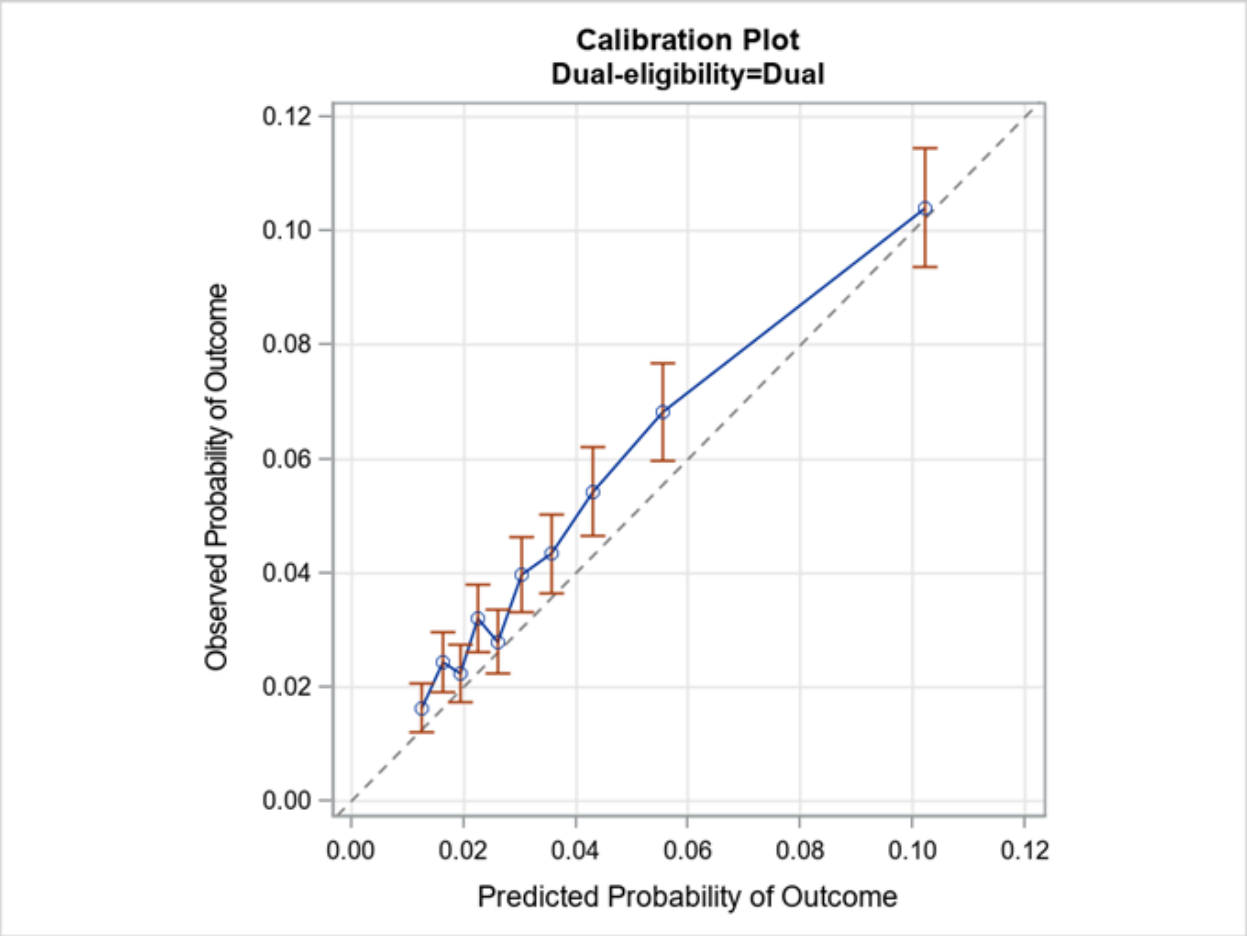

Figure 23. Calibration plot for patients with dual eligibility

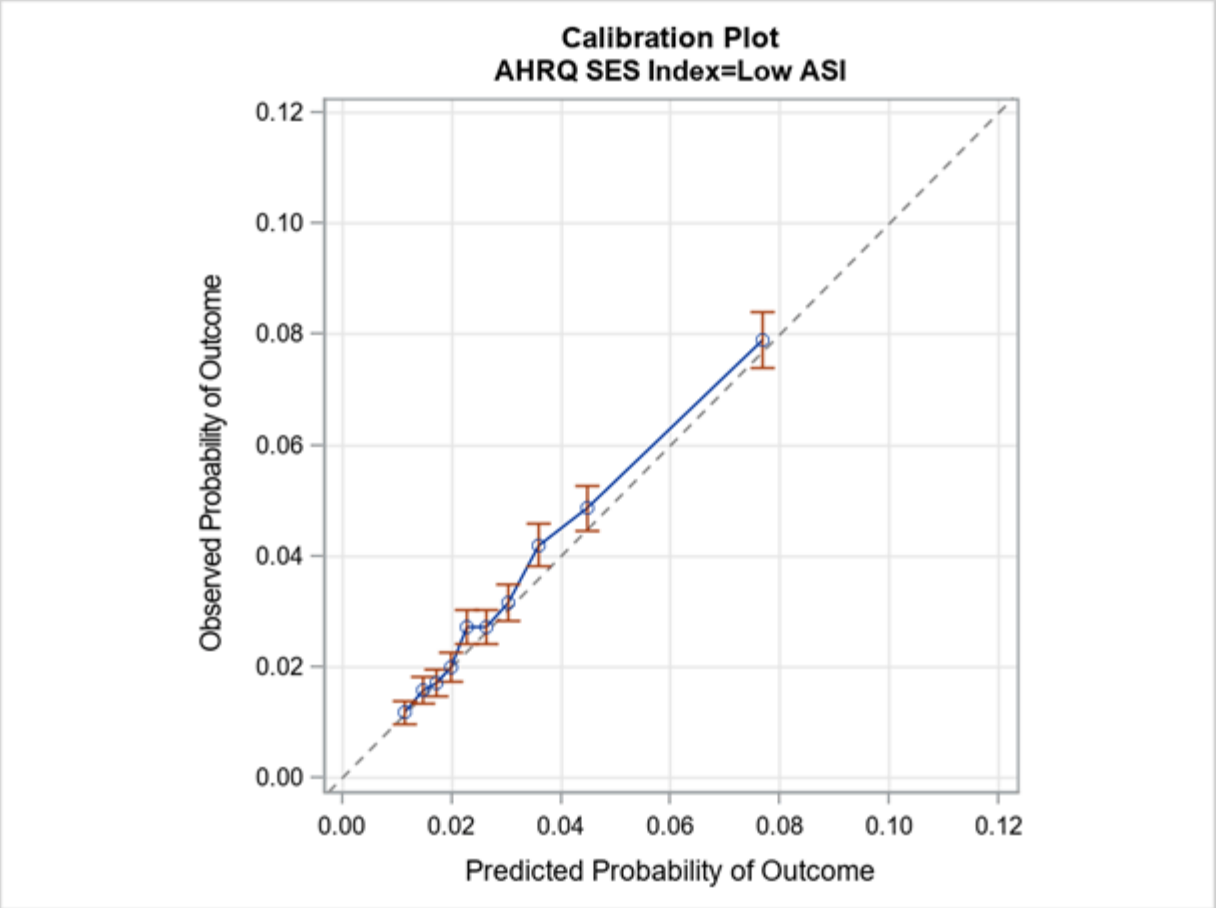

Figure 24. Calibration plot for patients with low ASI

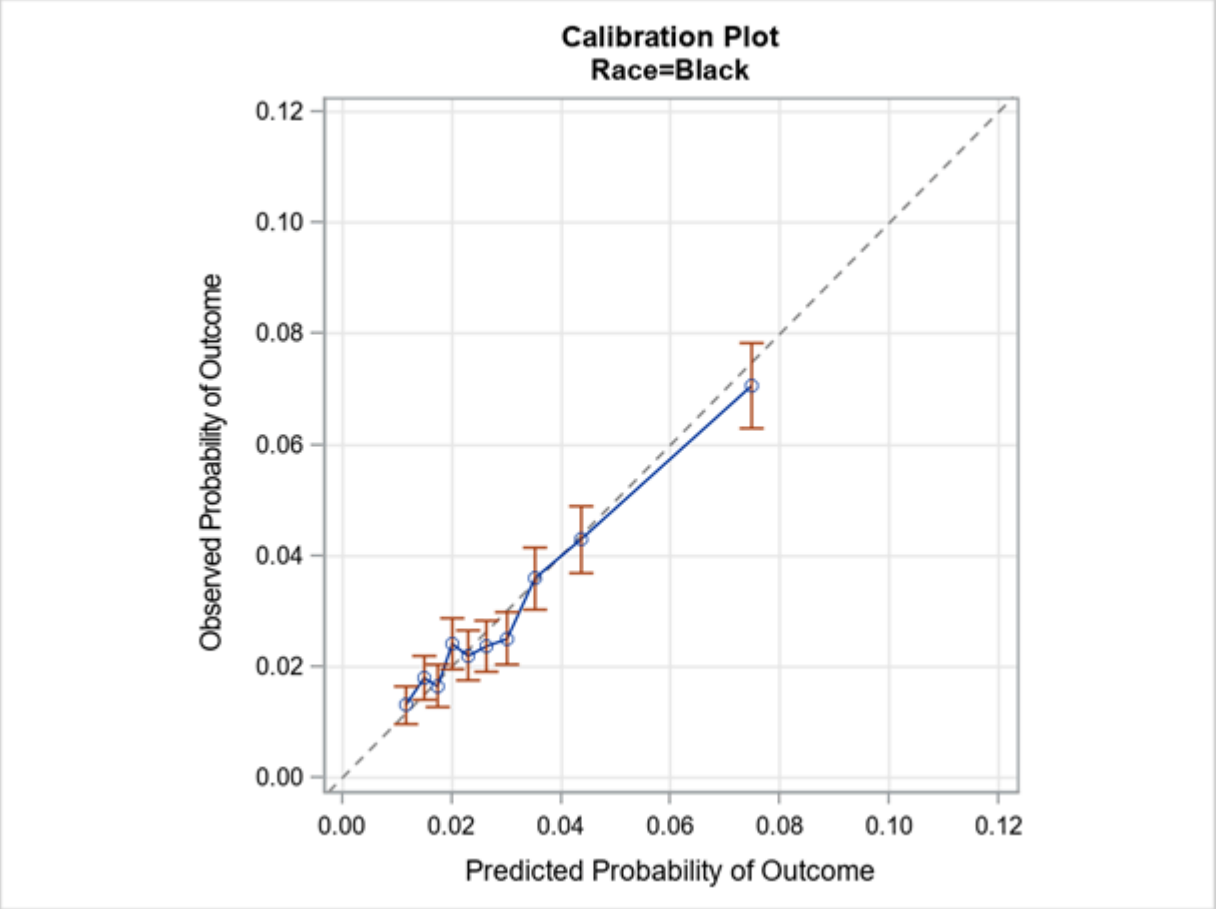

Figure 25. Calibration plot for patients with Exposure to racism

[Response Ends]

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

[Response Begins]

This measure is currently not risk stratified. Implementation planning, including potential stratification, has not been finalized.

[Response Ends]

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

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

[Response Begins]

We evaluated model performance by comparing the C-statistic, predictive ability, and internal calibration plots of each respective model. We considered model performance to be similar between the inpatient- and outpatient-cohort models using the following criteria: 1) the 95% confidence interval of the C-statistic for the two cohorts overlapped, and 2) the calibration plots were similar between the cohorts. Model discrimination (measured by C-statistic), and predictive ability (assessed as the comparison of highest to lowest deciles to evaluate the ability of model to distinguish high-risk subjects from low-risk subjects) were better when the model was applied in the combined and inpatient-only settings compared to the outpatient-only setting. The C-statistic was lower for the inpatient-only (0.651) and outpatient-only (0.638) subgroup models, respectively, compared to that of the single combined model with setting indicator (0.664) (Table 17). However, the predictive ability of the single combined model improved when the setting-specific variable was included in the model [Predictive ability (lowest decile %, highest decile %): 1.268%, 7.096% vs. 1.034%, 7.191%].

After extensive discussions with our Clinical Working Group and TEP and based on the empiric results, the measure will use a single, combined risk model that includes a setting indicator.

With regard to model performance in subgroups of patients with social risk factors, the risk-decile plot for patients with the low ASI and exposure to race variables were similar to that for all patients, suggesting the base model is well calibrated for theses subgroups. However, risk-decile plots show that the base model underpredicts risk for patients with dual eligibility.

[Response Ends]

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

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

[Response Begins]

Not applicable.

[Response Ends]

Criterion 3. Feasibility

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

[Response Begins]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

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

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

[Response Begins]

ALL data elements are in defined fields in a combination of electronic sources

[Response Ends]

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

[Response Begins]

Currently, administrative claims data offer greater coding accuracy (due to payment incentives and federal auditing programs) and lower provider burden than the same data obtained from the electronic health record data.

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A. This measure is not an eCQM.

[Response Ends]

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

[Response Begins]

N/A. This new measure is based on claims data.

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]

N/A. There are no fees, licensing, or other requirements associated with using any aspect of this new measure.

[Response Ends]

Criterion 4: Use and Usability

4a. Use

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

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

Not in use

[Not in use Please Explain]

This outcome measure is being submitted for initial endorsement and thus is not currently being used in any accountability program.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Payment Program

[Response Ends]

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

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

[Response Begins]

N/A; this outcome measure is being submitted for initial endorsement and thus is not currently being used in any accountability program. This measure is intended for a To-Be-Determined cross-setting payment model, such as CMMI's existing Comprehensive Care for Joint Replacement (CJR) model. It serves to ensure that measurement is aligned in both inpatient and outpatient settings, eliminating potential unintended consequences in measurement, as well as to reduce costs and create efficiencies. A measure that combines measurement across settings avoids potentially rewarding hospitals for shifting clinical settings without considering, or conflicting with, patients' individual needs.

[Response Ends]

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

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

[Response Begins]

As noted in 4a.03, this is a new measure and thus is not currently being publicly reported or used in an accountability application. However, the legacy measure upon which this measure is based—Hospital-level Risk-standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (NQF #1550)—was successfully implemented in the CMMI Comprehensive Joint Replacement (CJR) program. Therefore, we cannot identify any barriers to the implementation of the current measure, which aligns more closely with that program.

[Response Ends]

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

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

[Response Begins]

This is a new measure that has not been implemented yet, thus information on performance results or improvement have not been shared with hospitals, the measured entities. However, information about measure results was provided to, and feedback was obtained from a TEP (20 total members, four of which were patients) and a Clinical Working Group (four clinical expert members representing each of the four national TKA and/or THA professional societies). TEP members were selected through a publicly posted call for TEP on the CMS website and patients were recruited through partnerships with Rainmakers. Clinical Working Group members were nominated by the American Academy of Orthopaedic Surgeons, the American Association of Hip and Knee Surgeons, the Hip Society and the Knee Society. Feedback was obtained via teleconference calls. Patients engaging in this work were provided with preparation calls that reviewed the meeting materials ahead of the meeting date and debrief calls that allowed them to share any thoughts after the scheduled meeting. All meeting materials were sent in advance to allow individuals time to review the performance results and data. A summary of the feedback is provided in Section 1a.02 (Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful) of this form.

Public comments were also solicited by email notification to CMS listserv groups, email to relevant stakeholders and stakeholder organizations, and posting on the CMS Public Comment website. The measure methodology report and TEP summary report were provided as background information.

[Response Ends]

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

[Response Begins]

This is a new outcome measure being submitted for initial endorsement; it has not yet been implemented and is not currently being used in an accountability program. Thus, information on performance results and processes for providing that information have not yet been developed.

However, throughout measure development, we have obtained feedback and shared measure development and testing results with members of the Clinical Working Group and TEP, as well as through a Public Comment period.

To date, the TEP has provided input on, and supported, the measure concept, which includes extending monitoring of THA/TKA complications across the inpatient and hospital outpatient setting, as well as our approach to risk model development and testing and the results thereof. In addition, the TEP reviewed our approach to, and the results of, the social risk factor analyses, measure scores and reliability and validity testing. Similarly, we received support and input from the Clinical Working Group on the measure concept, risk model development, testing, and the results thereof, as well as the final measure scores and reliability and validity testing. The call for Public Comment requested feedback on recommended approaches for accounting for race, ethnicity, and/or social risk factors in implementation of the measure, key metrics to monitor after implementation to prevent unintended consequences, including worsening disparities; and topics for CMS/CMMI to consider during future measure reevaluation.

Once the measure is implemented, efforts will be made to evaluate measure performance regularly as described in the MMS Blueprint.

[Response Ends]

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

[Response Begins]

This new outcome measure is being submitted for initial endorsement; it has not yet been implemented and is not currently being used in any accountability program.

To date, feedback was obtained via five teleconference meetings with the Clinical Working Group, two teleconference meetings with the TEP, and a 30-day Public Comment period where feedback was received from three medical associations and societies (American Medical Association [AMA], American Association of Hip and Knee Surgeons [AAHKS], American Society of Anesthesiologists [ASA]). The TEP, Clinical Working Group, and public comments indicated strong support for the measure specifications and provided recommendations for ongoing evaluation, such as consideration of shifts in outpatient procedure volume, the impact of social determinants of health, and disparities in access.

In addition, measured entities (acute care hospitals) and other stakeholders or interested parties submit questions or comments about the measure/measure development through the QualityNet Q&A tool available at https://cmsqualitysupport.servicenowservices.com/qnet_qa, following these steps:

1. Access the tool at https://cmsqualitysupport.servicenowservices.com/qnet_qa?id=ask_a_question
2. Select "Inpatient Claims-Based Measures" from the drop-down menu in the Program field
3. Click into the Topic field and select "Understanding measure methodology" under "Complication"
4. Complete all other mandatory fields, the CAPTCHA, and click "Submit Question"

Experts on measure specifications, calculation, or implementation, prepare responses to those inquiries and reply directly to the sender. We consider issues raised through the Q&A process about measure specifications (or measure calculation for measures in reevaluation).

[Response Ends]

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

[Response Begins]

The feedback received from the TEP patient representatives showed their interest in the measure and support for the measure rationale and decision to expand the measure to the hospital outpatient setting, and showed understanding of measure development decisions, such as the use of a settings indicator. Patients also provided feedback about how closely the measure mirrored the actual experience of having a total hip or knee replacement and of experiencing a

complication. Patient representatives also shared their thoughts about the value of the measure for patients, stating that it would give them the ability to be better advocates for their own health, as well as allow them to make better decisions before having a total hip or knee replacement. Finally, patient representatives stressed the importance of making the patient experience a priority throughout measure development and implementation.

[Response Ends]

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

[Response Begins]

The TEP, which includes multiple clinicians and orthopedic stakeholders (in addition to patients), the Clinical Working Group, which is comprised of four clinicians, and our clinical expert indicated strong support for an elective primary THA/TKA measure that considers the outpatient setting. They recommended ongoing evaluation of the risk model and analyses on the social determinants of health. Similarly, feedback received during public comment was also supportive of the respecified measure's expansion to the outpatient setting. Additional comments included recommendations for continued monitoring of the impact of SDOH and potential unintended consequences, as well as the impact of minimum reliability thresholds and the absence of present on admission (POA) coding on outpatient claims.

[Response Ends]

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

[Response Begins]

This new outcome measure is being submitted for initial endorsement; it has not yet been implemented and is not currently being used in any accountability program. TEP, Clinical Working Group, and Public Comment feedback has been considered throughout measure development and respecification of the Hospital-level THA/TKA Complication measure.

[Response Ends]

4b. Usability

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

[Response Begins]

This new outcome measure is being submitted for initial endorsement; it has not yet been implemented or currently being used in any accountability program, thus data on trends in performance is not available. However, performance data from the existing (inpatient-only) Hospital-level THA/TKA complication measure (NQF #1550) that serves as the basis of the current measure, shows the potential utility of the new measure as a tool for performance improvement (Bozic et al, 2020).

The median RSCR for the existing (inpatient-only) Hospital-level THA/TKA complication measure for the 27-month period between April 1, 2018 – October 2, 2019, and July 1, 2020 – March 31, 2021, reflecting the shortened measurement period for 2022 public reporting, (Bernheim et al, 2022) was 2.4%. The median RSCR appears to have increased by 0.3 absolute percentage points from April 2018-March 2019 (median RSCR: 2.3%) to July 2020-March 2021 (median: RSCR: 2.6%). However, the median RSCR for the combined 27-month period (median RSCR: 2.4%) is consistent with the results for the 3-year period between April 1, 2016 – March 31, 2019, prior to the COVID-19 pandemic.

Yet these results also highlight the variability in the median inpatient-only RCSR as a result of the pandemic and the need for continued monitoring of THA/TKA complications and quality of care. Furthermore, it provides a strong rationale for

the development of the current measure, which extends monitoring to include outcomes associated with THA/TKA procedures performed in the hospital outpatient setting complication. Given the significant shift in the volume of procedures being performed in hospital outpatient department, as outlined in Section 1b.03, the performance results produced by the new IP/OP THA/TKA Complication measure have the potential to shed light on performance gaps across settings and ensure that high quality care is the norm irrespective of setting in which a THA or TKA procedure is performed.

References

1. Bozic K, Yu H, Zywiel MG, et al. Quality Measure Public Reporting Is Associated with Improved Outcomes Following Hip and Knee Replacement. *J Bone Joint Surg Am*. 2020;102(20):1799-1806. doi:10.2106/JBJS.19.00964
2. Bernheim, S., J.N. Grady, and J. Debuhr, *Procedure Specific Complication Measure Updates and Specifications Report: Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) Risk-Standardized Complication Measure (Version 11.0).* 2022.

[Response Ends]

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

[Response Begins]

N/A; this new outcome measure is not yet implemented. No unexpected findings were noted during measure development or testing.

[Response Ends]

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

[Response Begins]

N/A; this new outcome measure is not yet implemented. No unexpected findings were noted during measure development or testing.

[Response Ends]

Criterion 5: Related and Competing Measures

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

(Can search and select measures.)

[Response Begins]

1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)

1551: Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)

3474: Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)

3493: Risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) for Merit-based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups

3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)

3639: Clinician-Level and Clinician Group-Level Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA and TKA) Patient-Reported Outcome-Based Performance Measure (PRO-PM)

[Response Ends]

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

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A.

[Response Ends]

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

[Response Begins]

Yes

[Response Ends]

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

[Response Begins]

To the extent feasible, we have harmonized with existing related measures, specifically with CMS' existing hospital-level THA/TKA complication and readmission measures. Importantly, while this measure represents the same outcome and a similar patient population (patients undergoing elective primary THA/TKA procedures) as Measure NQF#1550: *Hospital-Level Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA),* the goal of this measure is to serve as a more accurate quality assessment tool for payment models that cross care settings. Therefore, this measure will never directly compete with NQF#1550 in a CMS program because it is intended for use only in applications that include inpatient and outpatient hospital settings.

[Response Ends]

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

Provide analyses when possible.

[Response Begins]

This measure has no competing measures.

[Response Ends]