NATIONAL QUALITY FORUM
Driving measurable health improvements together

Memo

November 17, 2020

To: Consensus Standards Approval Committee (CSAC)

From: Patient Experience and Function (PEF) Project Team

Re: PEF Spring 2020 Measures[a]

CSAC Action Required

The CSAC will review recommendations from the PEF project at its November 17-18, 2020 meeting and vote on whether to uphold the recommendations from the Committee.

This memo includes a summary of the project, measure recommendations, themes identified and responses to the public and member comments and the results from the National Quality Forum (NQF) member expression of support. The following documents accompany this memo:

1. **Patient Experience and Function Spring 2020 Draft Report**. The draft report has been updated to reflect the changes made following the Standing Committee's discussion of public and member comments. The complete draft report and supplemental materials are available on the project webpage.
2. **Comment Table**. NQF staff has identified themes within the comments received. This table lists eight comments received during the post-meeting comment period and the NQF/Standing Committee responses.

Background

The healthcare quality measurement ecosystem has increasingly embraced a paradigm that places the patient at the center of quality measurement. The incorporation of patient experience measures and patient function measures are two important components of patient-centered measurement.[1] The Centers for Medicare & Medicaid Services (CMS) Meaningful Measures Initiative includes the identification of measures that capture patients' experiences with clinicians and providers—one of 19 measurement areas for focusing national healthcare quality improvement efforts.[2] This falls under the measurement priority associated with strengthening person and family engagement as partners in their care. Ensuring that each person and family is engaged within a care partnership is critical to achieving better patient outcomes.[3]

Care coordination measures also represent a fundamental component for the success of this integrated approach, providing a multidimensional framework that spans the continuum of care and promotes quality care delivery, better patient experiences, and more meaningful outcomes.[4-6] Well-coordinated care encompasses effective communication among all patient and provider inputs of the care spectrum, and ensures that accountable structures and processes are in place for the integration of comprehensive

[a] This memo is funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001.

plans of care across providers and settings.[7-9]

Patient Experience and Function is a NQF measure topic area encompassing patient functional status, satisfaction, and experience of care, as well as issues related to care coordination. Central to the concepts associated with patient experience with their overall care is the patient's health-related quality of life and many factors that influence it, including communication, care coordination, transitions of care, and use of health information technology.[10-12]

Draft Report

The PEF Spring 2020 draft report presents the results of the evaluation of four measures considered under the Consensus Development Process (CDP). Four are recommended for endorsement.

- **NQF 2614** CoreQ: Short Stay Discharge Measure (American Health Care Association (AHCA)/National Center for Assisted Living (NACL))
- **NQF 2615** CoreQ: Long-Stay Resident Measure (AHCA)
- **NQF 2616** CoreQ: Long-Stay Family Measure (AHCA)/NACL)
- **NQF 3559** Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA) (Centers for Medicare & Medicaid Services (CMS)/Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE))

The measures were evaluated against the 2019 version of the measure evaluation criteria.

	Maintenance	New	Total
Measures under consideration	3	1	4
Measures recommended for endorsement	3	1	4

CSAC Action Required

Pursuant to the CDP, the CSAC is asked to consider endorsement of four candidate consensus measures.

Measures Recommended for Endorsement

- NQF 2614 CoreQ: Short Stay Discharge Measure (AHCA/NACL)

Overall Suitability for Endorsement: Yes-15; No-2

- NQF 2615 CoreQ: Long-Stay Resident Measure (AHCA)

Overall Suitability for Endorsement: Yes-14; No-2

- NQF 2616 CoreQ: Long-Stay Family Measure (AHCA/NACL)

Overall Suitability for Endorsement: Yes-14; No-2

- NQF 3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA) (Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE))

Overall Suitability for Endorsement: Yes-14; No-2

Comments and Their Disposition

NQF received eight comments from three member organizations and individuals pertaining to the draft report and to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Patient Experience and Function project webpage.

Comment Themes and Committee Responses

Comments about specific measure specifications and rationale were forwarded to the developers, who were invited to respond.

The Standing Committee reviewed all of the submitted comments (general and measure specific) and developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

Themed Comments

Theme 1 – Consensus Development Process

The American Medical Association (AMA) and the Federation of American Hospitals (FAH) noted concerns with the Consensus Development Process (CDP). Both stakeholders acknowledged that NQF 2614 – 2616 originate from the same developer and use similar specifications and testing approaches, but they disagreed with the Standing Committee's motion to carry evaluation criteria ratings from NQF 2614 to 2615 and subsequently, NQF 2615 to 2616 with no discussion. They disagreed with the notion that the evidence, reliability, validity, feasibility, use and usability are similar enough across NQF 2614 – 2616 to carry votes forward with no discussion, and they stated that the evidence, sampling strategies, testing results, target populations, different survey tools, feasibility, and usability differ enough to prompt discussion prior to a vote. The AMA and FAH requested clarification on these voting actions that were taken by the Standing Committee for NQF 2615 and 2616.

An additional process concern was raised, specifically by the FAH, about an omission of a pre-evaluation comment in the Spring 2020 Patient Experience and Function Draft Report. Additionally, a request to make the pre-evaluation comment available to the Standing Committee was included.

NQF Response

NQF thanked the commenters for the procedural question related to the carrying of the vote for measures 2614-2616. When considering measures with comparable submissions, NQF Committees may elect to carry the vote on a given criterion provided that there is unanimous support for doing so; a single dissention is sufficient to move to a full vote. This is not a departure from NQF's CDP. In fact, due to the concerns also identified by FAH, it requires a much higher degree of consensus than other CDP procedures. It is also not uncommon. The same measures were passed using the option to carry the vote for certain criteria during their previous evaluation in 2016, as were several other measures since.

Developer Response

No response was required.

Measure-Specific Comments

2614 Core Q: Short Stay Discharge Measure

"The American Geriatrics Society (AGS) wishes to provide comment on Measure 2614. The exclusion of those who are readmitted to acute care, transferred to another skilled nursing facility (SNF) or long-term

acute care (LTAC) facility, or remain in an SNF longer than 100 days removes subsets of patients who are more likely to have lower satisfaction with their short term SNF stay. We believe that these exclusions make the current measure less meaningful. We have additional reservations about the chosen cut-point for the score as 3 or above. A response of 2 is designated as 'average' in the Core Q questionnaire. We believe that this is not necessarily a negative response and could still indicate satisfaction with care."

Committee Response

The Committee thanked the AGS for its thoughtful review of these measures. The Committee considered their comment as well as the developer's response during the course of our meeting. The Committee has discussed the exclusions with the developer and is comfortable with their approach. The Committee also discussed the chosen cut-points for the measure and is satisfied with the scoring methodology used by the developer.

Developer Response

"Thank you for taking the time to review and comment on our measure. AHCA/NCAL greatly values feedback from patients, family members, and the public. We believe it to be paramount to measure the satisfaction of short and long-stay residents as well as family members (or designated parties), which is why we developed and tested measures NQF 2615 (CoreQ: Long-stay resident measure) and NQF 2616 (CoreQ: Long-stay family measure), to assess the satisfaction of long-stay residents and their family members/designated parties. The scoring of the CoreQ was tested using different cut-points. Facilities were very stable in their position relative to others when different cut-points were used. However, with a cut-point of two, many more facilities score at the highest levels. As such, if they subsequently improve their satisfaction, this is not well reflected in the change in CoreQ score. For this purpose, a cut-point of three works best. Also, "average" may not necessarily be a "bad" score - but with input from nursing home representatives during the testing of the CoreQ scoring, it was considered important to move the industry beyond average."

2615 Core Q: Long-Stay Resident Measure and 2616 Core Q: Long-Stay Family Measure

"The American Geriatrics Society (AGS) wishes to provide comment on NQF 2615 and 2616. We have reservations about the chosen cut-point for the score as three or above. A response of two is designated as 'average' in the Core Q questionnaire. We believe that this is not necessarily a negative response and could still indicate satisfaction with care."

Committee Response

The Committee thanked the AGS for its thoughtful review of these measures. The Committee considered their comment as well the developer's response during the course of its meeting and is satisfied with the developer's response and does not wish to take further action in reconsideration of the measure.

Developer Response

"Thank you for taking the time to review and comment on our measure. AHCA/NCAL greatly values feedback from patients, family members, and the public. The scoring of the CoreQ was tested using different cut-points. Facilities were very stable in their position relative to others when different cut-points were used. However, with a cut-point of two, many more facilities score at the highest levels. As such, if they subsequently improve their satisfaction this is not well reflected in the change in CoreQ score. For this purpose, a cut-point of three works best. Also, "average" may not necessarily be a "bad" score - but with input from nursing home representatives during the testing of the CoreQ scoring, it was considered important to move the industry beyond average."

2614 – 2616: Short Stay Discharge, Long-Stay Resident, Long-Stay Family Measures

"I am commenting as a Patient/Public individual interested and engaged in quality measurement and improvement. Thank you for offering the public opportunity to comment on this PEF NQF set of measures. I find the three QC measures appropriate for use, though would like to add that I have heard on a number of occasions that patient experience does not = patient satisfaction. Patient satisfaction is the outcome measured in these three measures. Are these designed to be used in place of CAHPS as they are more specific to stays in LTC/SNFs etc.? Family member identified for measurement query might vary as well. One family member might be pleased another not so much, just food for thought."

Committee Response

The Committee acknowledged the patient/public individual's comment and discussed it during the post-comment web meeting. The Committee agreed with the commenter and the developer that NQF 2614 – 2616 are satisfaction measures rather than patient experience measures, and that these satisfaction measures should not be conflated with patient experience measures.

Developer Response

"Thank you for taking the time to review and comment on our measure. AHCA/NCAL greatly values feedback from patients, family members, and the public. We agree that experience and satisfaction are two different, albeit related, constructs. NQF 2614: CoreQ: Short Stay Discharge Satisfaction measure, was specified and tested to be a patient satisfaction measure, only. NQF 2615: CoreQ: Long-Stay Resident Satisfaction measure, was specified and tested to be a resident satisfaction measure, only. NQF 2616: CoreQ: Long-Stay family measure, was specified and tested to be a family/designated party satisfaction measure, only."

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

"The FAH appreciates the opportunity to comment on measure NQF 3559, Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA), prior to the Standing Committee's evaluation. The FAH supports the development and implementation of patient-reported outcomes performance measures (PRO-PMs), but we also believe that additional questions and work remain before their widespread use. For instance, the degree to which multiple PRO-PMs could lead to survey fatigue for patients, the potential impact additional PRO-PMs may have on the reporting of well-established measures such as Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), and what level of data collection burden for an individual PRO-PM is acceptable for a hospital or other healthcare provider.

Specifically, on review of the measure specifications, the FAH notes that multiple data points beyond the typical clinical variables are required to ensure that the measure results are adequately risk adjusted. The FAH supports the inclusion of these data points, but we are concerned that the developer has not provided sufficient information on how these data are collected and what additional workload and time will be required. For example, several of the data elements needed for risk adjustment are derived from patient-reported surveys, which must be collected within 0-90 days pre-operative. No information was provided on the processes used by the hospitals such as whether it required coordination with orthopedic practices or if the burden of the additional data collection was placed on hospital staff on the day of surgery. To what extent did these requirements impact clinical workflows and were additional staff resources required? What additional costs might an individual hospital encounter as a result of implementation of this PRO-PM? Alternatively, from the patient's perspective, did the additional questions seem relevant and was the point in time during which these additional data were collected appropriate? It would also be useful to understand whether there is a potential for individuals to

prioritize the completion of one survey over another and therefore lead to negative unintended consequences on response rates for other PRO-PMs such as HCAHPS? The FAH believes that these questions should have been addressed during the development of this PRO-PM and this detail should have been provided within the measure submission rather than the generalized statements that we see in the responses under the feasibility criterion.

In addition, while the FAH strongly supports the inclusion of health literacy in the risk adjustment model, we believe that the risk adjustment approach used by many developers considers the identification and testing of social risk factors as supplementary to clinical risk factors. This approach was identified as a concern by the NQF Disparities Standing Committee. Given that this was a new measure, it provided an opportunity for the measure developer to include these factors within the testing of the model rather than the previous approach of "adding on" factors after the model is developed. This type of approach would assist hospitals and others in understanding how their inclusion could impact the model and provide additional information for groups examining this issue such as the NQF and Office of the Assistant Secretary for Planning and Evaluation (ASPE). As a result, the FAH believes that this measure lacks sufficient information on the potential impact these social risk variables have on the risk adjustment model."

Committee Response

The Committee thanked the FAH for its thoughtful review of the measures. The Committee considered their comment as well the developer's response during the course of our meeting. The Committee agreed that the development and implementation of PRO-PMs is vital. However, the Committee did not agree with FAH that this measure would be especially burdensome to implement. The Committee noted that the results of this measure are being collected across the country and is a standard of care for surgeons. The Committee has discussed the risk model with the developer and is comfortable with their approach.

Developer Response

"The Federation of American Hospitals submitted a public comment on June 12, 2020 for NQF 3559, stating support for PRO-PM development and implementation but noting concerns about data collection and hospital burden, survey fatigue for patients, the proposed measure's impact on response to other established PRO-PMs such as HCAHPS, and the inclusion of social risk variables. We would like to thank the Federation of American Hospitals for their comment and have provided responses to their concerns below.

The Centers for Medicare and Medicaid Services (CMS) is in the process of developing a strategic plan for implementing this measure into public reporting, informed by stakeholder input and with careful consideration of hospital burden. While PRO-PMs require providers to integrate data collection into clinical workflows, this integration provides opportunity for PROs to inform clinical decision making and benefit patients by engaging them in discussions about outcomes. The manner in which data collection is integrated into clinical workflows will be at the discretion of the hospital, which allows them freedom and flexibility to choose an approach that meets the needs of their facility.

We do not expect this PRO-PM to contribute to survey fatigue or to negatively impact other PRO-PMs. The Patient-Reported Outcome Measure (PROM) instruments that are used to calculate pre- and postoperative scores for this THA/TKA PRO-PM were carefully selected, with extensive stakeholder input, to be low burden for patients. Each patient undergoing a primary elective THA/TKA will be asked to fill out one joint-specific PROM (six questions on the HOOS, JR for hip replacement patients and seven questions on the KOOS, JR for knee replacement

patients) and one global health PROM (10 to 14 questions). In addition, it is unlikely that patients would be put in a position to prioritize the THA/TKA PRO-PM under consideration over other established PRO-PMs, such as the HCAHPS, as the assessment timelines for these two measures are completely different, with no overlap. The PRO data for the PRO-PM under consideration are collected 90 to 0 days before undergoing an elective, primary THA/TKA and 270 to 365 days after the THA/TKA, whereas the HCAHPS survey is administered to a random sample of adults within two days to six weeks after hospital discharge. Additionally, while sample sizes and patient caseloads vary by hospitals, we anticipate based on HCAHPS minimum caseload (for example, 100 completed surveys in a typical annual Hospital Value Based Purchasing program performance period) that most patients receiving the PROM surveys for this THA/TKA PRO-PM would not have also completed the HCAHPS survey for their hospital stay. Nationally, THA/TKA procedures comprise a small portion of hospital inpatient discharges, so we anticipate that relatively few patients would have completed both survey instruments covering their hospital stays.

The risk model for NQF 3559 was developed with clinical risk factors identified through a systematic literature review and orthopedic input. In consultation with the Technical Expert Panel (TEP), clinical consultants, and through detailed public comments, including from specialty societies, we focused on candidate risk-adjustment variables of interest that were clinically relevant and had an evidence-based relationship with clinical outcomes following elective primary THA or TKA. This included health literacy, identified by clinical experts and stakeholders as an important predictor and associated with a range of social risk factors. While health literacy is associated with other aspects of social risk, our TEP and measure development team felt that health literacy held particular relevance for a measure based upon PRO data. Thus, health literacy was evaluated during the initial risk model development along with clinical candidate risk variables. An additional assessment of the impact of social risk as captured by dual eligibility, the Agency for Health Research and Quality (AHRQ) socioeconomic status (SES) Index (socioeconomic status), and non-white race, was performed after the risk model was finalized. The addition of each of these three social risk variables provided no statistically significant change to our model performance, variable coefficients, or the model outcome. These variables were not included in the present model, but we believe future assessment in reevaluation will be important. These social risk variables were, however, statistically significantly associated with response to PRO surveys—whether patient-reported outcomes were obtained for patients undergoing primary elective THA/TKA—and so were included in the calculation of stabilized inverse probability weights used to account for potential response-bias."

"The American Medical Association (AMA) appreciates the opportunity to comment and vote on NQF 3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA). The AMA supports the assessment of patient-reported outcomes but believes that the burden of data collection both to the hospital and the patient must be adequately addressed and the continued multi-step approach to risk adjustment must be reconsidered. On review of the draft report, we were unable to find sufficient discussion of the validity and usability of this measure and ask that the Committee reconsider the proposed endorsement of this measure in light of our concerns outlined below.

On review of the measure specifications, we note that the information required for the numerator and risk variables includes multiple data elements from additional patient-reported surveys and that these data are expected to be collected between 90 to 0 days prior to surgery. The AMA supports the inclusion of many of these variables within the risk model given their relevance to how patients may or may not be able to achieve improvement but questions whether the developer (CMS) adequately assessed the

feasibility and potential data collection burden both to the hospital and patient. Specifically, the responses to the questions on feasibility do not discuss how the testing sites coordinated data collection across settings or whether the responsibility of the additional items was placed on the hospital. This question is particularly important since the specifications require hospitals to collect data for one measure from 90 days pre-operatively to up to one-year post-operative. Perhaps more importantly, we would have liked to see an assessment from the patient's perspective on whether the timing and number of items solicited throughout this process were appropriate and does not result in survey fatigue. For example, if these data were collected on the morning of the surgery, could stress and anxiety have impacted responses or would the number of surveys throughout the pre-, intra-, and post-operative timeframes lead them to be less likely to complete other surveys such as HCAHPS? We believe that it is critical to understand the potential impact and burden that could be experienced. While it may seem reasonable for one measure, if this measure is an example of how future measures could be specified, what is the potential long-term impact on patients and hospitals as more and more patient-reported outcome performance measures are implemented?

The AMA strongly supports the inclusion of health literacy in the risk model but remains concerned that CMS continues to test social risk factors after the assessment of clinical and demographic risk factors and it is unclear why this multi-step approach is preferable. On review of the Evaluation of the NQF Trial period for Risk Adjustment for Social Risk Factors report,[1] it is clear that the approaches to testing these data should be revised to strategies such as multi-level models or testing of social factors prior to clinical factors and that as access to new data becomes available, it may elucidate more differences that are unrelated to factors within a hospital's control. Additional testing that evaluates clinical and social risk factors at the same time or social prior to clinical variables rather than the current approach with clinical factors prioritized should be completed.

The AMA believes that additional information on these concerns is needed prior to endorsement of this measure. We respectfully ask the Standing Committee to consider these comments and not recommend the measure for endorsement until they are adequately addressed."

[1] National Quality Forum. *Evaluation of the NQF Trial period for Risk Adjustment for Social Risk Factors*. Final report. July 18, 2017. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=85635. Last accessed December 18, 2018.

Committee Response

The Committee thanked the AMA for this comment. These concerns were addressed in the Committee response to the FAH comment above.

Developer Response

"The American Medical Association submitted a public comment on August 25, 2020 for NQF 3559, stating support for PRO-PM development but noting concerns about data collection and hospital burden, the inclusion of social risk variables, intended use of the measure, and sufficient NQF discussion regarding the validity of the measure. We would like to thank the American Medical Association for their comment and have provided responses to their concerns below.

The issue of burden has been carefully considered during measure development and will be carefully considered for measure implementation. The Centers for Medicare and Medicaid Services (CMS) is in the process of developing a strategic plan for implementing this measure into public reporting, informed by stakeholder input and with careful consideration of hospital burden. While patient-reported outcome performance measures (PRO-PMs) require providers to integrate data collection into clinical workflows, this integration provides opportunity for PROs to inform clinical decision making and benefit patients by engaging them in discussions

about outcomes. The manner in which data collection is integrated into clinical workflows will be at the discretion of the hospital, which allows them freedom and flexibility to choose an approach that meets the needs of their facility. CMS's Comprehensive Care for Joint Replacement (CJR) program, the source of the data for measure development and testing, served as proof of concept for patient-reported outcome data collection and submission.

We do not expect this PRO-PM to contribute to survey fatigue or to negatively impact the collection of data for other PRO-PMs. The Patient-Reported Outcome Measure (PROM) instruments that are used to calculate pre- and postoperative scores for this THA/TKA PRO-PM were carefully selected, with extensive stakeholder input, to be low burden for patients. Each patient undergoing a primary elective THA/TKA will be asked to fill out one joint-specific PROM (6 questions on the HOOS, JR for hip replacement patients and seven questions on the KOOS, JR for knee replacement patients) and one global health PROM (10 to 14 questions). In addition, it is unlikely that patients would be less likely to complete other PROM surveys such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), as the assessment timelines for these two measures are completely different, with no overlap. As noted, the PRO data for the PRO-PM under consideration are collected 90 to zero days before undergoing an elective, primary THA/TKA and 270 to 365 days after the THA/TKA, whereas the HCAHPS survey is administered to a random sample of adults within two days to six weeks after hospital discharge. Additionally, while sample sizes and patient caseloads vary by hospitals, we anticipate based on HCAHPS minimum caseload (for example, 100 completed surveys in a typical annual Hospital Value Based Purchasing program performance period) that most patients receiving the PROM surveys for this THA/TKA PRO-PM would not have also completed the HCAHPS survey for their hospital stay. Nationally, THA/TKA procedures comprise a small portion of hospital inpatient discharges, so we anticipate that relatively few patients would have completed both survey instruments covering their hospital stays.

And while the addition of PRO-PMs over time will require continued consideration of patient and hospital burden, the value of patient-reported outcomes to patient-centered care, improved patient care and improved hospital quality is widely supported. This measure was developed with extensive input from patients, who indicated strong support for a patient-reported outcomes-based performance measure following elective primary THA and TKA. They stated that patients expected a significant amount of improvement in pain levels and functional status, which helped define improvement thresholds for the measure outcome. They reviewed improvement threshold considerations, the preoperative and postoperative data collection periods, risk adjustments models, and testing results, and found the measure to be valuable. They provided feedback on the presentation of the measure outcome for patient use.

The risk model for NQF 3559 was developed with clinical risk factors identified through a systematic literature review and orthopedic input. In consultation with the Technical Expert Panel (TEP), clinical consultants, and through detailed public comments, including from specialty societies, we focused on candidate risk-adjustment variables of interest that were clinically relevant and had an evidence-based relationship with clinical outcomes following elective primary THA or TKA. This included health literacy, identified by clinical experts and stakeholders as an important predictor and associated with a range of social risk factors. While health literacy is associated with other aspects of social risk, our TEP and measure development team felt that health literacy held particular relevance for a measure based upon PRO data. Thus, health literacy was evaluated during the initial risk model development along with clinical candidate risk variables. An additional assessment of the impact of social risk as captured by dual eligibility, the Agency for Health Research and Quality (AHRQ) socioeconomic status (SES) Index

(socioeconomic status), and non-white race, was performed after the risk model was finalized. The addition of each of these three social risk variables provided no statistically significant change to our model performance, variable coefficients, or the model outcome. These variables were not included in the present model, but we believe future assessment in reevaluation will be important. These social risk variables were, however, statistically significantly associated with response to PRO surveys—whether patient-reported outcomes were obtained for patients undergoing primary elective THA/TKA—and so were included in the calculation of stabilized inverse probability weights used to account for potential response-bias. We will continue to assess the impact of social risk factors on measure over time.

Finally, extensive discussion regarding the validity of measure was conducted by the Scientific Methods Panel, which subsequently voted in support of the measure on both reliability and validity."

"The American Geriatrics Society (AGS) wishes to provide comment on NQF 3559. While this seems like a potentially useful PROM, it is unclear how it would be implemented. There should be some consideration given to the amount of time hospital staff would devote to collecting this measure as well as the costs involved, both of which will be borne directly by the hospitals. We suggest that this measure could be replaced by data which hospitals may already be collecting (such as ADLs, IADLs, or pain scores) and therefore do not pose as much of a burden on reporting hospitals."

Committee Response

The Committee thanked the AGS for their comment. These concerns were addressed in the Committee response to the FAH comment above.

Developer Response

"The American Geriatrics Society submitted a public comment on September 3, 2020 for Measure 3559, stating support for the potential usefulness of the measure but expressing concern about implementation and hospital burden. We would like to thank the American Geriatrics Society for their comment and have provided responses to their concerns below.

The Centers for Medicare and Medicaid Services (CMS) is in the process of developing a strategic plan for implementing this measure into public reporting, informed by stakeholder input and with careful consideration of hospital burden. While patient-reported outcomes performance measures (PRO-PMs) require providers to integrate data collection into clinical workflows, this integration provides opportunity for patient reported outcomes (PROs) to inform clinical decision making and benefit patients by engaging them in discussions about potential outcomes. The manner in which data collection is integrated into clinical workflows will be at the discretion of the hospital, which allows them freedom and flexibility to choose an approach that meets the needs of their facility.

This performance measure, built on established and validated PROM survey instruments, will standardize across hospitals the data collected, the assessment periods for data collection, and the risk adjustment used for evaluation of hospital quality following elective primary THA and TKA procedures. While most hospitals do collect functional and pain information from orthopedic patients (such as ADLs, IADLS, and pain scores), assessment of hospital-level performance for public reporting requires a standard approach across all hospitals. The patient reported outcome measure (PROM) instruments that are used to calculate pre- and postoperative scores for this THA/TKA PRO-PM were carefully selected, with extensive stakeholder input from clinicians, to be low burden and joint specific. The clinicians felt, and data demonstrated, that joint-specific functional status tools such as the HOOS, JR and KOOS, JR

are more relevant for clinical decision making and are more responsive than PROMs that are not as specific. In addition, this measure was developed with extensive input from patients, who indicated strong support for a PRO-PM following elective primary THA and TKA. They stated that patients expected a significant amount of improvement in pain levels and functional status, which helped define improvement thresholds for the measure outcome. They reviewed improvement threshold considerations, the preoperative and postoperative data collection periods, risk adjustments models, and testing results. Patients provided feedback on the presentation of the measure outcome for patient use and found the measure to be valuable."

"As a patient/public partner engaged in quality improvement, my comments are as follows. I find it interesting that NQF 3559 measure did not have an intended use identified pre NQF endorsement phase. It seems to me that having the use case open allows for ease for endorsement though also some might feel that the intended use of measures should be identified in development therefore decreasing opportunities for questions later. I did appreciate the longevity of time for follow up as in these types of total hip/knee arthroplasty surgeries, the 30/60/90-day outcomes might not reflect actual patient experience or function post-surgery that impacts their quality of life. Thank you for this."

Committee Response

The Committee acknowledged the patient/public individual's comment and discussed it during the post-comment meeting. The Committee agreed with the commenter and the developer that NQF 2614 – 2616 are satisfaction measures. The Committee also agreed that these satisfaction measures are not patient experience measures, and they should not be conflated with patient experience measures.

Developer Response

"An independent member of the public submitted a public comment for NQF 3559. They expressed support for the measure and the timeframe for post-operative data collection, noting that the longer follow up is beneficial for measuring patient experience and quality of life. They also expressed a concern about the intended use of the measure, which was not identified during the development of the measure. We would like to thank this individual for their comment and have provided a response to their concern below.

During NQF evaluation of new measures, fully identified "Usability and Use" is not required for NQF endorsement. The Centers for Medicare and Medicaid Services (CMS) is in the process of developing a strategic plan for implementing this measure into public reporting, informed by stakeholder input and with careful consideration of hospital burden. CMS's Comprehensive Care for Joint Replacement (CJR) program, the source of the data for measure development and testing, served as proof of concept for patient-reported outcome data collection and submission.

This measure was developed with extensive stakeholder input and support from patients, orthopedic groups, and clinical experts. Patients indicated strong support for a patient-reported outcomes-based performance measure following elective primary THA and TKA. They stated that patients expected a significant amount of improvement in pain levels and functional status, which helped define improvement thresholds for the measure outcome. They reviewed improvement threshold considerations, risk adjustments models, and testing results, and found the measure to be valuable. They provided feedback on the presentation of the measure outcome for patient use."

Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ('support' or 'do not support') for each measure submitted for endorsement consideration to inform the Committee's recommendations. One NQF member provided their expression of non-support. Appendix C details the expression of support.

Removal of NQF Endorsement

Two measures previously endorsed by NQF have not been re-submitted, and endorsement has been removed.

Measure	Measure Description	Reason for Removal of Endorsement
0726 Patient Experience of Psychiatric Care as Measured by the Inpatient Consumer Survey (ICS)	The Patient Experience of Psychiatric Care as Measure by the Inpatient Consumer Survey (ICS) was developed to gather patient´s evaluation of their inpatient psychiatric care. The survey is composed of the following six individual measures or domains: Measure #1: Outcome of care Measure #2: Dignity Measure #3: Rights Measure #4: Participation in treatment Measure #5: Hospital environment Measure #6: Empowerment Each measure is scored as the percentage of patients (adolescents aged 13-17 and adults aged 18 and older) at time of discharge or at annual review who respond positively to the domain on the survey for a given month. Survey questions are based on a standard 5-point Likert scale, evaluated on a scale from strongly disagree to strongly agree. As a note, the words domain and measure are used interchangeably during the application.	A shorter version of the measure is being developed.
3452 Access to Independence Promoting Services for Dual Eligible Beneficiaries	The composite measure is generated from responses to optional supplemental survey items used in the Medicare Advantage and Prescription Drug Plans (MA & PDP) Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. Currently, these supplemental items are required only when the survey is fielded among dual eligible beneficiaries enrolled in Medicare-Medicaid Plans (MMPs). The composite measure is the equally weighted average of the following three indicators: • Indicator 1: Access to Medical Equipment This indicator generates a score based on a target population of individuals who reported a	The measure is no longer supported by the steward.

Measure	Measure Description	Reason for Removal of Endorsement
	need for medical equipment and indicated how easy it was to get or replace the medical equipment through their health plan during the last six months. • Indicator 2: Access to Personal Aide Assistance This indicator generates a score based on a target population of individuals who reported a need for home health care or assistance and indicated how easy it was to get personal care or aide assistance through their health plan during the last six months. • Indicator 3: Access to Counseling or Treatment This indicator generates a score based on a target population of individuals who reported a need for treatment or counseling for personal or family problems and indicated how easy it was to get treatment or counseling through their health plan during the last six months.	

References

[1] Agency for Healthcare Research and Quality (AHRQ). Priorities of the national quality strategy website. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr15/priorities.html. Last accessed February 2020.

[2] Centers for Medicare & Medicaid Services. Meaningful Measures Hub website. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/MMF/General-info-Sub-Page#Measurement_Areas. Last accessed February 2020.

[3] Frandsen BR, Joynt KE, Rebitzer JB, et al. Care fragmentation, quality, and costs among chronically ill patients. *AJMC.* 2015; 21(5):355-362.

[4] Tricco AC, Antony J, Ivers NM, et al. Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services: a systematic review and meta-analysis*. CMAJ*.

2014;186(15): E568-578.

[5] Turchi RM, Antonelli RC, Norwood KW Jr, et al. Patient-and family-centered care coordination: a framework for integrating care for children and youth across multiple systems. *Pediatrics*.

2014;133(5): e1451-e1460.

[6] Pronovost P, Weast B, Schwarz M, et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. *J Crit Care.* 2003;18(4):201-205.

[7] Gnanasakthy A, Mordin M, Evans E, et al. A review of patient-reported outcome labeling in the United States (2011-2015). *Value Health.* 2017;20(3):420-429.

[8] Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. *Med Decis Making*. 2015;35(1):114-131.

[9] Berkowitz SA, Parashuram S, Rowan K, et al. Johns Hopkins Community Health Partnership (J-CHiP) Team. Association of a care coordination model with health care costs and utilization: The Johns Hopkins Community Health Partnership (J-CHiP). *JAMA Netw Open*. 2018;1(7): e184273.

[10] Schultz EM, Pineda N, Lonhart J, et al. A systematic review of the care coordination measurement landscape. *BMC Health Serv Res*. 2013; 13:119.

[11] Rosenthal MB, Alidina S, Friedberg MW, et al. A difference-in-difference analysis of changes in quality, utilization and cost following the Colorado multi-payer patient-centered medical home pilot. *J of Gen Intern Med*. 2016;31(3):289-296.

[12] Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *New Engl J Med*. 2009;360(14):1418-1428.

Appendix A: CSAC Checklist

The table below lists the key considerations to inform the CSAC's review of the measures submitted for endorsement consideration.

Key Consideration	Yes/No	Notes
Were there any process concerns raised during the CDP project? If so, briefly explain.	Yes	Member commenters, AMA and the FAH noted concerns with the CDP. Both stakeholders acknowledged that NQF 2614 – 2616 originate from the same developer and use similar specifications and testing approaches, but they disagreed with the Standing Committee's motion to carry evaluation criteria ratings from NQF 2614 to 2615 and subsequently, NQF 2615 to 2616 with no discussion. NQF staff acknowledged the process-based concern and provided clarity on the Standing Committee's voting decision to carry the vote for measures 2614-16. Furthermore, NQF staff explained that Standing Committees may elect to carry the vote on a given criteria provided that there is unanimous support for doing so when considering measures with comparable submissions; a single dissention is sufficient to move to a full vote. NQF staff noted that this is not a departure from nor an uncommon action of NQF's CDP. The same measures were passed using the option to carry the vote for certain criteria during their previous evaluation in 2016, as were several other measures since.
Did the Standing Committee receive requests for reconsideration? If so, briefly explain.	No	
Did the Standing Committee overturn any of the Scientific Methods Panel's ratings of Scientific Acceptability? If so, state the measure and why the measure was overturned.	No	
If a recommended measure is a related and/or competing measure, was a rationale provided for the Standing Committee's recommendation? If not, briefly explain.	n/a	
Were any measurement gap areas addressed? If so, identify the areas.	No	

Key Consideration	Yes/No	Notes
Are there additional concerns that require CSAC discussion? If so, briefly explain.	No	

Appendix B: Measures Not Recommended for Endorsement

The Standing Committee recommends all candidate measures in the spring 2020 cycle for endorsement.

Appendix C: NQF Member Expression of Support Results

One NQF members provided their expression of non-support for one measure under consideration. Results for the measure are provided below.

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

Member Council	Support	Do Not Support	Total
Health Professional	0	1	1

Appendix D: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable

Measures Recommended

2614 CoreQ: Short-stay Discharge Measure

Submission

Description: The measure calculates the percentage of individuals discharged in a six-month time period from a SNF, within 100 days of admission, who are satisfied (see: S.5 for details of the time frame). This patient-reported outcomes measure is based on the CoreQ: Short-stay Discharge questionnaire that utilizes four items.

Numerator Statement: The measure assesses the number of patients who are discharged from a SNF, within 100 days of admission, who are satisfied. The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the four questions on the CoreQ: Short-stay Discharge questionnaire.

Denominator Statement: The denominator includes all of the patients that are admitted to the SNF, regardless of payor source, for post-acute care, that are discharged within 100 days; who receive the survey (e.g., people meeting exclusions do not receive a questionnaire); and who respond to the CoreQ: Short-stay Discharge questionnaire within the time window.

Exclusions: Exclusions used are made at the time of sample selection and include:

(1) Patients who died during their SNF stay;

(2) Patients discharged to a hospital, another SNF, psychiatric facility, inpatient rehabilitation facility or long-term care hospital;

(3) Patients with court-appointed legal guardian for all decisions;

(4) Patients discharged to hospice;

(5) Patients who left the nursing facility against medical advice (AMA);

(6) Patients who have dementia impairing their ability to answer the questionnaire, defined as having a BIMS score on the MDS of 7 or lower. [Note: we understand that some SNCCs may not have information on cognitive function available to help with sample selection. In that case, we suggest administering the survey to all residents, and assume that those with cognitive impairment will not complete the survey or have someone else complete it on their behalf, which, in either case, will exclude them from the analysis.]

(7) Patients who responded after the two-month response period; and

(8) Patients whose responses were filled out by someone else.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post-Acute Care

Type of Measure: Outcome: PRO-PM

Data Source: Instrument-Based Data. The collection instrument is the CoreQ: Short-stay Discharge questionnaire and Resident Assessment Minimum Data Set (MDS) version 3.0.

Measure Steward: AHCA/NCAL

STANDING COMMITTEE MEETING 06/23/2020, 06/24/2020

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence; 1b. Performance Gap)

1a. Evidence: **Pass-16; No Pass-0**; 1b. Performance Gap: **H-3**; **M-11**; **L-1**; **I-2**

Rationale:

- The Committee began the discussion of evidence for NQF 2614 by acknowledging that the developer's submission has changed minimally since the 2016 review; the evidence submitted is essentially the same, with the exception of a portion that was added in relation to the meaningfulness of the measure to patients.
- Developer provided updated statistics broken out by quarter with each quarter representing a rolling 12-month period of data.

2614 CoreQ: Short-stay Discharge Measure

- There were no statistically significant differences by race or education, but significant differences by age and gender.
- The Committee noted stable performance rates over time and had a brief discussion related to the differences in performance expectations for satisfaction and patient experience measures versus clinical measures.
- Committee questioned whether the imputation methods for incomplete surveys affected the performance scores. It was noted that regardless of the method used for imputation—either imputing the maximum or minimum score—the resulting performance scores were essentially the same.
- The Committee noted that there were no statistical differences in measure performance associated with race and expressed concern that this is incongruent with known quality problems by race in nursing facilities.
- The Committee noted that research over the last 20 years has consistently found poorer care in facilities with high minority populations and that nursing homes remain segregated, with black patients concentrated in poorer-quality homes (as measured by staffing ratios, performance, and financial vulnerability).

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-10**; **L-0**; **I-1**; 2b. Validity: **H-2**; **M-11**; **L-3**; **I-1**

<u>Rationale</u>:

- Developer used the same testing from the 2016 submission
- Measure developer performed both data element level and score level reliability testing
- Data element reliability testing included test-retest analysis on a convenience sample of 100 patients
    - Developer calculated the distribution of responses by question in the original round of surveys, and then again in the follow-up surveys (they should be distributed similarly);
    - Developer subsequently calculated the correlations between the original and follow-up responses by question (they should be highly correlated).
- The stability of the facility-level score was tested using bootstrapping with 10,000 repetitions of the facility score calculation.
    - Developer presented the percent of facility resamples where the facility score is within 1 percentage point, 3 percentage points, 5 percentage points, and 10 percentage points of the original score.
- Data element reliability testing showed very high levels of agreement and no statistically significant difference in the responses to each question between the original and re-test results.
- Validity testing of the CoreQ: Short Stay Discharge questionnaire included both data element level and score level testing.
- Data element level
    - Exploratory factor analysis (EFA) was used to further refine the pilot instrument. This was an iterative process that included using Eigenvalues from the principal factors (unrotated) and correlation analysis of the individual items.
    - Correlation analysis and a factor analysis conducted
    - Face validity evaluated via literature review and review of 12 commonly used satisfaction surveys; also examined face validity of domains and the response scale, using 40 patients in 5 nursing homes. The Flesch-Kinkaid scale was used to determine if patients understood the questions.
    - Also examined correlation between the four items in the measure and all the items on the pilot instrument.

2614 CoreQ: Short-stay Discharge Measure

- Measure score level
- Convergent validity testing was performed. Developers examined correlations between the CoreQ measure scores and i) measures of regulatory compliance and other quality metrics from the Certification and Survey Provider Enhanced Reporting (CASPER) data, ii) several other quality metrics from Nursing Home Compare, iii) risk adjusted discharge to community measure and iv) risk adjusted PointRight® Pro 30™ Rehospitalizations.
- The Committee noted in the discussion of scientific acceptability that the submission was the same as the previous submission in 2016.
- The Committee questioned the exclusion of surveys that were completed by proxy, but noted that many survey firms exclude satisfaction surveys completed by someone other than the person who receives the service.
- The Committee requested the developer to explore the payment source (Medicare, Medicaid) for the purposes of investigating case mix adjustment.
- In the discussion on validity, the Committee noted in the convergent validity analyses conducted by the developer with external measures of quality that many of the measures did not exhibit high convergence with NQF 2614. It was suggested that this is also true of other measures of satisfaction; it is not unsual for such measures to be independent of other measures of quality.

3. Feasibility: H-2; M-14; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- This is a satisfaction survey conducted via mailed survey.
- No fees required to use the measure; the developer did not indicate if there are fees associated with the use of the survey.
- The Committee discussed the feasibility of the measure and noted that it was similar to other measures of patient satisfaction in terms of its overall burden to providers and patients.
- The Committee suggested that the developer consider an electronic version of the surveys.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; 4c. Benefits outweigh evidence of unintended consequences)

4a. Use: Pass-15; No Pass-1; 4b. Usability: H-2; M-14; L-0; I-1

Rationale:

- The measure is used in several accountability applications
- Professional Certification or Recognition Program
  - AHCA Quality Initiative
  - AHCA Quality Awards
- Quality Improvement (external benchmarking to organizations)
  - AHCA NCAL Long Term Care Trend Tracker
- Developer notes that a number of states are implementing the CoreQ survey inside of state incentive programs, including NJ, MA, TN, GA, and others.
- The Committee did not express concerns related to use or usability.

5. Related and Competing Measures

- No competing measures noted.

6. **Standing Committee Recommendation for Endorsement**: **Yes-15**; **No-2**

7. Public and Member Comment

Comments received related to:

2614 CoreQ: Short-stay Discharge Measure
• Intended use and inquiry concerning the relationship with CAHPS measures • Concerns with exclusions and testing of CoreQ scoring
8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020):
9. Appeals

2615 CoreQ: Long-Stay Resident Measure
Submission
Description The measure calculates the percentage of long-stay residents, those living in the facility for 100 days or more, who are satisfied (see: S.5 for details of the time frame). This patient-reported outcomes measure is based on the CoreQ: Long Stay Resident questionnaire that is a three-item questionnaire. **Numerator Statement**: The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long Stay Resident questionnaire. **Denominator Statement**: The denominator includes all of the residents that have been in the SNF for 100 days or more regardless of payer status; who received the CoreQ: Long Stay Resident questionnaire (e.g., people meeting exclusions do not receive the questionnaire), who responded to the questionnaire within the two-month time window, who did not have the questionnaire completed by somebody other than the resident, and who did not have more than one item missing. **Exclusions**: Exclusions made at the time of sample selection are the following: (1) Residents who have poor cognition defined by the BIMS score; (2) residents receiving hospice; (3) residents with a legal court-appointed guardian; and (4) residents who have lived in the SNF for less than 100 days. Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (two months after the administration date); b) surveys that have more than one questionnaire item missing; c) surveys from residents who indicate that someone else answered the questions for the resident. (Note this does not include cases where the resident solely had help such as reading the questions or writing down their responses.) **Adjustment/Stratification**: No risk adjustment or risk stratification **Level of Analysis:** Facility **Setting of Care:** Post-Acute Care **Type of Measure**: Outcome: PRO-PM **Data Source**: Instrument-Based Data. The collection instrument is the CoreQ: Long Stay Resident questionnaire, and exclusions are from the Resident Assessment Instrument Minimum Data Set (MDS) version 3.0. **Measure Steward**: American Health Care Association
STANDING COMMITTEE MEETING 07/09/2020 **1. Importance to Measure and Report: The measure meets the Importance criteria** (1a. Evidence; 1b. Performance Gap) 1a. Evidence: **Pass-16; No Pass-0**; 1b. Performance Gap: **H-1**; **M-12**; **L-2**; **I-1** Rationale: • The Committee began the discussion of evidence for NQF 2615 by acknowledging that the developer's submission has changed minimally since the 2016 review; the evidence submitted is essentially the same, with the exception of a portion that was added related to the meaningfulness of the measure to patients. • The Committee noted that drivers for high satisfaction rates include competency of staff, care/concern of staff, and responsiveness of management. • The Committee also acknowledged that the evidence for this measure was similar to NQF 2614 and elected to carry the vote from the previous with little discussion.

2615 CoreQ: Long-Stay Resident Measure

- Developer provided updated statistics broken out by quarter with each quarter representing a rolling 12-month period of data.
- There were no statistically significant differences by race or education, but significant differences by age and gender.
- The performance gap discussion focused on disparities, with the Committee noting that only 6% of the data in the analysis was from Black patients.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>
(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)
2a. Reliability: **H-6**; **M-10**; **L-0**; **I-1**; 2b. Validity: **H-1**; **M-11**; **L-3**; **I-0**
<u>Rationale</u>:

- Developer used the same testing from the 2016 submission.
- Measure developer performed both data element level and score level reliability testing.
- Data element reliability testing included test-retest analysis on a convenience sample of 100 patients.
    - Developer calculated the distribution of responses by question in the original round of surveys, and then again in the follow-up surveys (they should be distributed similarly);
    - Developer subsequently calculated the correlations between the original and follow-up responses by question (they should be highly correlated).
- The stability of the facility-level score was tested using bootstrapping with 10,000 repetitions of the facility score calculation.
    - Developer presented the percent of facility resamples where the facility score is within 1 percentage point, 3 percentage points, 5 percentage points, and 10 percentage points of the original score.
- Data element reliability testing showed very high levels of agreement and no statistically significant difference in the responses to each question between the original and re-test results.
- Validity testing of the CoreQ: Short Stay Discharge questionnaire included both data element level and score level testing.
- Data element level
    - Exploratory factor analysis (EFA) was used to further refine the pilot instrument. This was an iterative process that included using Eigenvalues from the principal factors (unrotated) and correlation analysis of the individual items.
    - Correlation analysis and a factor analysis conducted
    - Face validity evaluated via literature review and review of 12 commonly used satisfaction surveys; also examined face validity of domains and the response scale, using 40 patients in 5 nursing homes. The Flesch-Kinkaid scale was used to determine if patients understood the questions.
    - Also examined correlation between the four items in the measure and all the items on the pilot instrument.
- Measure score level
    - Convergent validity testing was performed.
    - Developers examined correlations between the CoreQ measure scores and i) measures of regulatory compliance and other quality metrics from the Certification and Survey Provider Enhanced Reporting (CASPER) data, ii) several other quality metrics from Nursing Home Compare, iii) risk adjusted discharge to community measure and iv) risk adjusted PointRight® Pro 30™ Rehospitalizations.

2615 CoreQ: Long-Stay Resident Measure

- The Committee acknowledged that the reliability testing for this measure was very similar to NQF 2614 and the Committee chose to carry the vote from NQF 2614 with no discussion.
- The Committee acknowledged some differences in the measure exclusions and suggested that risk-factors for poor satisfaction should not be an exclusion for the measure.
    - As such, it was concerning to the Committee that 34% of patients were excluded from the analysis because of poor cognition.
    - The developer noted that the Brief Interview for Mental Status (BIMS) was used to determine cognition levels.
    - It was suggested that BIMS can serve as a proxy for recall of patient satisfaction with care and that patients that are unable to recall their experience may not be able to accurately provide input related to it.
    - The Committee also discussed not having family members serve as a proxy. It was noted that there is significant variance between measures of patient satisfaction and family satisfaction, hence the need for the third measure of family satisfaction (NQF 2616).

3. Feasibility: H-2; M-14; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- This is a satisfaction survey conducted via mailed survey.
- No fees required to use the measure; the developer did not indicate if there are fees associated with the use of the survey.
- The Committee carried the vote from measure NQF 2614 with no discussion.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use: Pass-15; No Pass-1; 4b. Usability: H-2; M-14; L-0; I-1

Rationale:

- The measure is used in several accountability applications.
- Professional Certification or Recognition Program
    - AHCA Quality Initiative
    - AHCA Quality Awards
- Quality Improvement (external benchmarking to organizations)
    - AHCA NCAL Long Term Care Trend Tracker
- Developer notes that a number of states are implementing the CoreQ survey inside of state incentive programs, including NJ, MA, TN, GA, and others.
- The Committee carried the vote from measure NQF 2614 with no discussion.

5. Related and Competing Measures

- No competing measures noted.

6. Standing Committee Recommendation for Endorsement: Yes-14; No-2

7. Public and Member Comment:

Comments received related to:

- Intended use and inquiry concerning the relationship with CAHPS measures
- Concerns with exclusions and testing of CoreQ scoring

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020):

2615 CoreQ: Long-Stay Resident Measure

9. Appeals

2616 CoreQ: Long-Stay Family Measure

Submission

Description: The measure calculates the percentage of family or designated responsible party for long stay residents (i.e., residents living in the facility for 100 days or more), who are satisfied (see: S.5 for details of the time frame). This consumer reported outcome measure is based on the CoreQ: Long-Stay Family questionnaire that has three items.

Numerator Statement: The numerator assesses the number of family or designated responsible party for long stay residents that are satisfied. Specifically, the numerator is the sum of the family or designated responsible party members for long stay residents that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long-Stay Family questionnaire.

Denominator Statement: The target population is family or designated responsible party of a resident residing in a SNF for at least 100 days. The denominator includes all of the individuals in the target population who respond to the CoreQ: Long-Stay Family questionnaire within the two-month time window (see S.5) who do not meet the exclusion criteria (see S.10).

Exclusions: Please note, the resident representative for each current resident is initially eligible regardless of their being a family member or not. Only one primary contact per resident should be selected.

Exclusions made at the time of sample selection include: (1) family or designated responsible party for residents with hospice; (2) family or designated responsible party for residents with a legal court-appointed guardian; (3) representatives of residents who have lived in the SNF for less than 100 days; and (4) representatives who reside in another country.

Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (more than two months after the administration date); and b) surveys that have more than one questionnaire item missing.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post-Acute Care

Type of Measure: Outcome: PRO-PM

Data Source: Instrument-Based Data. The collection instrument is the CoreQ: Long-Stay Family questionnaire, and for exclusions the Resident Assessment Instrument Minimum Data Set (MDS) version 3.0 is used.

Measure Steward: AHCA/NCAL

STANDING COMMITTEE MEETING 07/09/2020

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence; 1b. Performance Gap)

1a. Evidence: **Pass-16; No Pass-0**; 1b. Performance Gap: **H-1**; **M-12**; **L-2**; **I-1**

Rationale:

- The Committee began the discussion of evidence for NQF 2616 by acknowledging that the developer's submission has changed minimally since the 2016 review; the evidence that was submitted is essentially the same, with the exception of a portion that was added related to the meaningfulness of the measure to patients.
- Developer provided updated statistics broken out by quarter with each quarter representing a rolling 12-month period of data.
- There were no statistically significant differences by race or education, but significant differences by age and gender.

2616 CoreQ: Long-Stay Family Measure

- The Committee began the discussion on evidence by asking the developer if the items of the survey are sufficiently granular to detect important problems in care delivery. The developer noted that their analysis has indicated that they are sufficiently granular, with no differences in satisfaction scores when detailed questions are used instead of general satisfaction questions.
- The Committee also discussed the differences between satisfaction and experience surveys.
- It was also noted that the family view of the care provided for patients can be very different than the patient perspective, acknowledging the need for this measure as a complement to measure NQF 2615.
- The Committee elected to carry the vote from NQF 2615 on both evidence and performance gap.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-10**; **L-0**; **I-1**; 2b. Validity: **H-1**; **M-11**; **L-3**; **I-0**

<u>Rationale</u>:

- Developer used the same testing from the 2016 submission.
- Measure developer performed both data element level and score level reliability testing.
- Data element reliability testing included test-retest analysis on a convenience sample of 100 patients:
    - Developer calculated the distribution of responses by question in the original round of surveys, and then again in the follow-up surveys (they should be distributed similarly);
    - Developer subsequently calculated the correlations between the original and follow-up responses by question (they should be highly correlated).
- The stability of the facility-level score was tested using bootstrapping with 10,000 repetitions of the facility score calculation.
    - Developer presented the percent of facility resamples where the facility score is within 1 percentage point, 3 percentage points, 5 percentage points, and 10 percentage points of the original score.
- Data element reliability testing showed very high levels of agreement and no statistically significant difference in the responses to each question between the original and re-test results.
- Validity testing of the CoreQ: Short Stay Discharge questionnaire included both data element level and score level testing.
- Data element level
    - Exploratory factor analysis (EFA) was used to further refine the pilot instrument. This was an iterative process that included using Eigenvalues from the principal factors (unrotated) and correlation analysis of the individual items.
    - Correlation analysis and a factor analysis conducted.
    - Face validity evaluated via literature review and review of 12 commonly used satisfaction surveys; also examined face validity of domains and the response scale, using 40 patients in 5 nursing homes. The Flesch-Kinkaid scale was used to determine if patients understood the questions.
    - Also examined correlation between the four items in the measure and all the items on the pilot instrument.
- Measure score level

2616 CoreQ: Long-Stay Family Measure

- Convergent validity testing was performed. Developers examined correlations between the CoreQ measure scores and I) measures of regulatory compliance and other quality metrics from the Certification and Survey Provider Enhanced Reporting (CASPER) data, ii) several other quality metrics from Nursing Home Compare, iii) risk adjusted discharge to community measure and iv) risk adjusted PointRight® Pro 30™ Rehospitalizations.
- The Committee considered the issues related to scientific accepatability to be similar to NQF 2615 and chose to carry the vote for both reliability and validity with no discussion.

3. Feasibility: H-2; M-14; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- This is a satisfaction survey conducted via mailed survey.
- No fees required to use the measure; the developer did not indicate if there are fees associated with the use of the survey.
- The Committee carried the vote from measure NQF 2615 with no discussion.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; 4c. Benefits outweigh evidence of unintended consequences)

4a. Use: Pass-15; No Pass-1; 4b. Usability: H-2; M-14; L-0; I-1

Rationale:

- The measure is used in several accountability applications.
- Professional Certification or Recognition Program
  - AHCA Quality Initiative
  - AHCA Quality Awards
- Quality Improvement (external benchmarking to organizations)
  - AHCA NCAL Long Term Care Trend Tracker
- Developer notes that a number of states are implementing the CoreQ survey inside of state incentive programs, including NJ, MA, TN, GA, and others.
- The Committee carried the vote from measure NQF 2615 with no discussion.

5. Related and Competing Measures

- No competing measures noted.

6. Standing Committee Recommendation for Endorsement: Yes-14; No-2

7. Public and Member Comment:

- Comments received during public comment were related to intended use and inquiries concerning the relationship with CAHPS measures.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020):

9. Appeals

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

Submission

Description: This patient-reported outcomes-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcomes data collected prior to and following the elective procedure. The preoperative data collection time frame will be 90

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

to 0 days before surgery, and the postoperative data collection time frame will be 270 to 365 days following surgery.

Numerator Statement: The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows:

- For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and

- For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR).

SCB thresholds were defined using published literature (Lyman and Lee, 2018), and vetted by our Patient Working Group, Technical Expert Panel (TEP), and Technical Advisory Group.

References:

Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.

Denominator Statement: The cohort (target population) includes Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.

Exclusions: Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.

Adjustment/Stratification: Statistical risk model. No risk stratification

Level of Analysis: Facility

Setting of Care: Inpatient/Hospital

Type of Measure: Outcome: PRO-PM

Data Source: Claims, Instrument-Based Data. The PROM surveys used to define the measure outcome are: 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients; and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are presently available in English; not yet in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) 10 Global or the Veterans RAND 12 Item Health Survey (VR-12) are used. The PROMIS Global is available in 16 languages; the VR-12 is available in Spanish, Chinese, and German.

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

STANDING COMMITTEE MEETING 06/23/2020, 06/24/2020

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence; 1b. Performance Gap)

1a. Evidence: **Pass-17; No Pass-0**; 1b. Performance Gap: **H-4**; **M-9**; **L-2**; **I-0**

<u>Rationale</u>:

- Developer notes that there are many studies indicating how providers can improve outcomes of the patients by addressing aspects of pre-, peri-, and postoperative care (Brown et al., 2012; Choong et al., 2009; Galea et al., 2008; Kim, 2019; McGregor et al., 2004; Moffet et al., 2004; Monticone et al., 2013; Walters, 2016).
- Optimal clinical outcomes may be influenced by:
    - The surgeon performing the procedure
    - Team's efforts in the care of the patient
    - Care coordination across provider groups and specialties
    - Patients' engagement in their own recovery (Feng et al, 2018; Saufl et al, 2007).

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

- The Committee began the discussion of evidence by noting that the measure is based on two survey instruments, the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR).
    - The Committee noted that for a patient in the denominator to be included in the numerator as well, they need to have a performance score improvement of 20 on the HOOS, JR or 22 on the KOOS, JR.
    - The Committee noted the analysis conducted by Stephen Lyman, et al. for the analytical basis in selecting these thresholds.
- The Committee expressed concern about the ability of patients to interpret the meanings associated with the thresholds, as well as actual implementation by clinicians at the point of care.
- The Committee emphasized the need to have such measures of functional status directly integrated into workflow to allow them to inform decisions related to the direction of care.
- Developer provided disparities data for n=6,734 patients within the Development Dataset, analyzing race, dual eligibility status, and socioeconomic status (SES).
- Chi-square analyses and multivariate analyses did not reveal a statistically significant association between non-White race or SES.
- Dual eligibility was borderline significant (p=0.058) at the bivariate level and statistically significant within the risk model.
- In the discussion on performance gap, the Committee noted the relatively low representation of non-Whites sampled (8.2%) which is not reflective of the general population.
    - The developer noted that elective procedures result in differential access to care and a disproportionately White population that receives full joint replacements electively.
    - They also noted a propensity score weighting used in their analysis as a means to compensate for this phenomenon.
- The Committee also expressed concern for missing differences in care in that patients with English as a second language may be screened out.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: SMP Rating-High; Vote to uphold the SMP's rating: **Yes-14 No-1**; 2b. Validity: SMP Rating-Moderate; Vote to uphold the SMP's rating: **Yes-10 No-4**

<u>Rationale</u>:

- Reliability testing conducted at the data element and score level.
- Data element reliability testing assessed consistency and test-retest reliability of the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) and Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) instruments.
    - HOOS, JR internal consistency using Person Separation Index (PSI) was 0.86 and 0.87 in the two cohorts tested.
    - HOOS, JR test-retest results produced ICCs between 0.75 and 0.97.
    - KOOS, JR internal consistency using PSI was 0.84 and 0.85 in the two cohorts tested.
    - KOOS, JR test-retest results produced ICCs between 0.75 and 0.93.
- Score level reliability testing consisted of a signal-to-noise analysis. Results from a sample of 123 hospitals yielded a mean of 0.95 and a range from 0.90 to 0.99.
- Validity testing was conducted at both the data element and score levels.
- Data element validity testing included responsiveness, external validity, floor and ceiling effects for both HOOS, JR and KOOS, JR. HOOS, JR responsiveness produced standardized

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

response means relative to other PROMs (HOOS domains, The Western Ontario and McMaster University Arthritis Index [WOMAC] domains) measuring post-surgery hip improvement of 2.38 and 2.03 in the two samples.
    - HOOS, JR external validity used Spearman's correlation analysis with the HOOS and WOMAC instruments and produced 0.87 for both samples.
    - HOOS, JR showed floor (0.6%–1.9%) and ceiling (37%–46%) effects and were comparable to or better than HOOS domains and the WOMAC.
    - KOOS, JR responsiveness produced standardized response means relative to other PROMs (KOOS, WOMAC) measuring post-surgery hip improvement of 1.79 and 1.70 in the two samples.
    - KOOS, JR external validity used Spearman's correlation analysis with the KOOS and WOMAC instruments and produced 0.89 and 0.91 for the two samples.
    - KOOS, JR showed floor (0.4%–1.2%) and ceiling (18.8%–21.8%) effects.
- Score level validity testing included empirical comparisons to another quality measure: NQF 1550 Hospital-level risk-standardized complication rate (RSCR) following elective primary THA/TKA. Comparison of THA/TKA PRO-PM RSIRs to RSCR categories indicates an increasing monotonic trend. Those hospitals in the "RSCR Worse than National Average" category have lower median RSIRs (51.87%) than the median RSIR (66.49%) of hospitals in the "RSCR Same as National Average" category, which is lower than that of hospitals in the "RSCR Better than National Average" category (71.13%).
- The Committee reviewed the evaluation of scientific acceptability of the measure by the NQF SMP.
- The Committee expressed no concern on the data element reliability of the measure but cited some concern on the sources of error, noting that the signal to noise analysis conducted using the beta binomial method described by Adams in 2009 only includes one source of provider error but that this measure potentially has several others, such as low response rate.
- It was noted during the validity discussion that risk factors should not be exclusions for the measure and the Committee expressed concern that the exclusions may rule out complications associated with total joint replacement. The developer clarified that this is not the case and that the measure removes second elective procedures.
- The Committee also discussed the 25-patient threshold for public reporting as well as adjusting for social drivers of health.

3. Feasibility: H-0; M-11; L-4; I-1

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- During the feasibility discussion, the Committee noted the burden of paper scoring methodologies for functional status measures and encouraged to developer to explore digital capture.
- The developer noted that the HOOS, JR and KOOS, JR have lower total items than the full scoring tool to reduce burden and that providers do not score the instrument directly.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; 4c. Benefits outweigh evidence of unintended consequences)

4a. Use: Pass-6; No Pass-10 4b. Usability: H-0; M-8; L-3; I-5

Rationale:

3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
• In the discussion on use and usability, the Committee noted that while the measure was commissioned by CMS, they did not provide an explanation related to the intended use of the measure or a plan for its implementation. • This does not meet the NQF standard and the Committee did not pass the measure on use (not a must-pass criterion for new measures).
5. Related and Competing Measures • No competing measures noted.
6. Standing Committee Recommendation for Endorsement: **Yes-14**; **No-2**
7. Public and Member Comment: Comments received during public comment were related to feasibility and implementation concerns regarding cost, administrative burden (providers), survey fatigue (patients), and the process for data collection. Other comments were related to the risk adjustment approach and the multi-step inclusion of social risk factors as well as concerns with the adequacy of the Standing Committee discussion on validity and usability.
8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020):
9. Appeals

NATIONAL QUALITY FORUM
Driving measurable health improvements together

http://www.qualityforum.org

Patient Experience and Function Spring 2020 Review Cycle

CSAC Review and Endorsement

November 17, 2020

Funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001.

NATIONAL QUALITY FORUM

Standing Committee Recommendations

- Four measures reviewed for Spring 2020
  - One measure reviewed by the SMP
- Four measures recommended for endorsement
  - **NQF 2614** CoreQ: Short-stay Discharge Measure (Maintenance Measure)
  - **NQF 2615** CoreQ: Long-Stay Resident Measure (Maintenance Measure)
  - **NQF 2616** CoreQ: Long-Stay Family Measure (Maintenance Measure)
  - **NQF 3559** Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (New Measure)

NATIONAL QUALITY FORUM

Overarching Issues

- Inequities Within Satisfaction Measures
  - The Committee discussed that there are known differences in the quality of care provided to minorities within these facilities.
    - » Poorer care in facilities with high minority populations.
    - » Nursing homes remain segregated, with black patients concentrated in poorer quality homes (as measured by staffing ratios, performance, and financial vulnerability).
  - Concerning given that NQF 2614-2616 did not exhibit statistically significant differences by race.
    - » The Committee noted that patient satisfaction equates to meet expectations, and questioned the implications associated with equal satisfaction in the face of known quality inequities.
    - » The Committee also noted that many satisfaction measures do not always track well with external measures of quality.

Public and Member Comment and Member Expressions of Support

- Eight comments received
  - One NQF member expression of non-support received

NATIONAL QUALITY FORUM

Questions?

- Project team:
  - Samuel Stolpe, PharmD, MPH, Senior Director
  - Oroma Igwe, MPH, Manager
  - Udobi Onyeuku, MSHA, Analyst
  - Yemsrach Kidane, PMP, Project Manager
- Project webpage: http://www.qualityforum.org/Project_Pages/Patient_Experience_and_Function.aspx
- Project email address: patientexperience@qualityforum.org

THANK YOU.

NATIONAL QUALITY FORUM

http://www.qualityforum.org

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Patient Experience and Function Spring 2020 Cycle: CDP Report

DRAFT REPORT FOR CSAC REVIEW
NOVEMBER 17, 2020

This report is funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order 75FCMC19F0007.

http://www.qualityforum.org

Contents

Executive Summary

The healthcare quality measurement ecosystem has increasingly embraced a paradigm that places the patient at the center of quality measurement. The incorporation of patient experience measures and patient function measures are two important components of patient centered measurement.[1] The Centers for Medicare & Medicaid Services (CMS) Meaningful Measures Initiative includes the identification of measures that capture patients' experiences with clinicians and providers—one of 19 measurement areas for focusing national healthcare quality improvement efforts.[2] This falls under the measurement priority associated with strengthening person and family engagement as partners in their care. Ensuring that each person and family is engaged within a care partnership is critical to achieving better patient outcomes.[3]

Care coordination measures also represent a fundamental component for the success of this integrated approach, providing a multidimensional framework that spans the continuum of care and promotes quality care delivery, better patient experiences, and more meaningful outcomes.[4-6] Well-coordinated care encompasses effective communication among all patient and provider inputs of the care spectrum and ensures that accountable structures and processes are in place for the integration of comprehensive plans of care across providers and settings.[7-9]

Patient Experience and Function (PEF) is a National Quality Forum (NQF) measure topic area encompassing patient functional status, satisfaction, and experience of care, as well as issues related to care coordination. Central to the concepts associated with patient experience with their overall care is the patient's health-related quality of life and many factors that influence it, including communication, care coordination, transitions of care, and use of health information technology.[10-12]

The NQF PEF Committee was established to evaluate measures within this topic area for NQF endorsement. NQF has endorsed over 50 measures addressing patient experience of care, patient functional status, mobility and self-care, shared decision making, patient activation, and care coordination. The majority of the measures within this portfolio are patient-reported outcomes performance measures. During this cycle, the Committee's discussion remained primarily focused on the measures under consideration for maintenance review, but also included a broad discussion related to healthcare disparities and the incongruencies between measures of satisfaction and external measures of quality.

For this project, the Standing Committee evaluated one newly submitted measure, and three measures undergoing maintenance review against NQF's standard evaluation criteria. The Standing Committee recommended all four measures for endorsement. The measures recommended for endorsement are:

- **NQF 2614** CoreQ: Short Stay Discharge Measure (American Health Care Association (AHCA)/National Center for Assisted Living (NACL))
- **NQF 2615** CoreQ: Long-Stay Resident Measure (AHCA)
- **NQF 2616** CoreQ: Long-Stay Family Measure (AHCA)/NACL)
- **NQF 3559** Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA) ((CMS)/Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE))

Brief summaries of the measures currently under review are included in the body of the report; detailed summaries of the Committee's discussion and ratings of the criteria for each measure are in Appendix A.

Introduction

The implementation of patient-centered measures is one of the most important approaches to ensure that the healthcare in the United States (U.S.) reflects the goals, preferences, and values of care recipients. Patient- and family-engaged care is planned, delivered, managed, and continually improved in active partnership with patients and their families (or care partners as defined by the patient) to ensure integration of their health and healthcare goals, preferences, and values.[13] As such, effective engaged care must adapt readily to individual and family circumstances, as well as differing cultures, languages, disabilities, health literacy levels, and socioeconomic backgrounds.[1]

The coordination of care is an essential component to the improvement of patient experiences and outcomes. Poorly coordinated and fragmented care not only compromises the quality of care patients receive, but may also lead to negative, unintended consequences, including medication errors and preventable hospital admissions.[10] For patients living with multiple chronic conditions—including more than two-thirds of Medicare beneficiaries—poor care transitions between different providers can contribute to poor outcomes and hospitalizations.[14] Nearly 15 percent of Medicare beneficiaries discharged from the hospital are readmitted within 30 days, with half of the patients not having yet seen an outpatient doctor for follow-up, and most of these readmissions occur through an emergency department (ED). [12,15] The existing evidence suggests that care in the U.S. is largely uncoordinated, even though evidence also suggests that quality improvement strategies within care can improve performance.[5] Care coordination is positively associated with patient- and family-reported receipt of family-centered care, resulting in greater satisfaction with services, lower financial burden, and fewer ED visits.[3,6,9,14,16]

The capture of patient experience and satisfaction is a hallmark of patient-centered measurement. This is especially important for vulnerable populations. Among those are patients who receive care within nursing facilities. According to data from the Centers for Disease Control and Prevention (CDC), there are over 15,000 nursing homes nationwide serving a population of over 1.3 million residents.[17] Data collected from CMS indicates that the US spends over $160 billion per year in health expenditures related to nursing care facilities and continuing care retirement communities.[18] It is critical to understand the perceptions of patients and their families related to care received within these facilities.

Patient-reported outcomes, such as functional status, are increasingly recognized as critical markers of success for surgical interventions. Total joint replacements are among the most common surgeries performed in the United States. In 2010, the prevalence of total hip and total knee replacement in the total US population was 0.8% and 1.5% respectively, representing 2.5 million patients with total hip replacement and 4.7 million patients with total knee replacement.[19] There has been a substantial rise in prevalence over time with patients receiving total joint arthroplasty at increasingly younger ages. Median total joint arthroplasty costs are estimated at approximately $15,000 per year, representing a significant national economic burden.[20]

NQF Portfolio of Performance Measures for Patient Experience and Function Conditions

The Patient Experience and Function Standing Committee (Appendix C) oversees NQF's portfolio of Patient Experience and Function measures (Appendix B) that includes measures of functional status, communication, shared decision making, care coordination, patient experience, and long-term services and supports. This portfolio contains 49 measures: three process measures, one composite measure, and 45 outcome measures, of which 27 are patient-reported outcome performance measures (see Table 1).

Table 1. NQF Patient Experience and Function Portfolio of Measures

	Process	Outcome/Resource Use	Composite
Functional status change and assessment	1	23	0
Shared decision making	0	3	0
Care coordination	2	5	0
Patient experience	0	10	1
Long-term services and supports	0	4	0
Total	3	45	1

There are other measures of patient experience and function that have been assigned to other portfolios. These include healthcare-associated infection measures (Patient Safety), care coordination measures (Geriatrics and Palliative Care), imaging efficiency measures (Cost and Resource Use), and a variety of condition- or procedure-specific outcome measures (Cardiovascular, Cancer, Primary Care and Chronic Illness and Renal).

Patient Experience and Function Measure Evaluation

On June 23, June 24, and July 9, 2020, the Patient Experience and Function Standing Committee evaluated one new measure and three measures undergoing maintenance review against NQF's standard measure evaluation criteria.

Table 2. Patient Experience and Function Measure Evaluation Summary

	Maintenance	New	Total
Measures under consideration	3	1	4
Measures recommended for endorsement	3	1	4

Comments Received Prior to Committee Evaluation

NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). In addition, NQF solicits comments for a continuous 16-week period during each

evaluation cycle via an online tool located on the project webpage. For this evaluation cycle, the commenting period opened on May 1, 2020 and closed on September 3, 2020. Pre-meeting commenting closed on June 12, 2020. As of that date, no comments were submitted prior to the measure evaluation meeting(s) (Appendix F).

Comments Received After Committee Evaluation

The continuous 16-week public commenting period with NQF member support closed on August 5, 2020. Following the Committee's evaluation of the measures under consideration, NQF received eight comments from three member organizations and individuals pertaining to the draft report and to the measures under consideration. All comments for each measure under consideration have been summarized in Appendix A.

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ('support' or 'do not support') for each measure submitted for endorsement consideration to inform the Committee's recommendations. One NQF members provided their expression of non-support.

Overarching Issue

During the Standing Committee's discussion of the measures, an overarching issue emerged that was factored into the Committee's ratings and recommendations for multiple measures and is not repeated in detail with each individual measure.

Known Inequities Within Satisfaction Measures

Three of the measures reviewed during Committee measure evaluations were CoreQ measures based on survey questions administered to residents and their family members within nursing facilities. The Committee discussed that there are known differences in the quality of care provided to minorities within these facilities. In particular, research over the last 20 years has consistently found poorer care in facilities with high minority populations and that nursing homes remain segregated, with black patients concentrated in poorer-quality homes (as measured by staffing ratios, performance, and financial vulnerability). This was especially concerning given that the three measures did not exhibit statistically significant differences by race. The Committee discussed that measures of patient satisfaction are essentially measures of met expectations and questioned the implications associated with equal satisfaction in the face of known quality inequities. The Committee expressed concern that the measures may mask differences in quality by showing that racial minority patients are still satisfied with poorer quality care. The Committee also noted that many satisfaction measures do not always track well with external measures of quality.

Summary of Measure Evaluation

The following brief summaries of the measure evaluation highlight the major issues that the Committee considered. Details of the Committee's discussion and ratings of the criteria for each measure are included in Appendix A.

2614 CoreQ: Short Stay Discharge Measure (AHCA/NACL): Recommended

Description: The measure calculates the percentage of individuals discharged in a six month time period from a SNF, within 100 days of admission, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Short Stay Discharge questionnaire that utilizes four items.; **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Facility; **Setting of Care**: Post-Acute Care; **Data Source**: Instrument-Based Data

The Standing Committee recommended the measure for continued endorsement. This measure is based on the CoreQ: Short Stay Discharge questionnaire that utilizes four items. The Committee began the discussion of this measure by acknowledging that the developers submission has changed very little since the 2016 review; the evidence submitted is essentially the same, with the exception of a portion that was added related to the meaningfulness of the measure to patients. The Committee noted stable performance rates over time and had a brief discussion related to the differences in performance expectations for satisfaction and patient experience measures versus clinical measures. The developer noted that this is a unique survey instrument, though there are other proprietary vendor survey instruments that measure patient satisfaction in the same population. The Committee questioned whether the imputation methods for incomplete surveys affected the performance scores. It was noted that regardless of the method used for imputation—either imputing the maximum or minimum score—the resulting performance scores were essentially the same. The Committee noted that there were no statistical differences in measure performance associated with race and expressed concern that this is incongruent with known quality problems by race in nursing facilities. The Committee noted that research over the last 20 years has consistently found poorer care in facilities with high minority populations and that nursing homes remain segregated, with black patients concentrated in poorer-quality homes (as measured by staffing ratios, performance, and financial vulnerability). The Committee noted in the discussion of scientific acceptability that the submission was the same as the previous submission in 2016. The Committee questioned the exclusion of surveys that were completed by proxy, but noted that many survey firms exclude satisfaction surveys completed by someone other than the person who receives the service. The Committee requested the developer to explore the payment source (Medicare, Medicaid) for the purposes of investigating case mix adjustment. In the discussion on validity, the Committee noted in the convergent validity analyses conducted by the developer with external measures of quality that many of the measures did not exhibit high convergence with NQF 2614. It was suggested that this is also true of other measures of satisfaction; it is not unsual for such measures to be independent of other measures of quality. The Committee discussed the feasibility of the measure and noted that it was similar to other measures of patient satisfaction in terms of its overall burden to providers and patients. The Committee suggested that the developer should consider an electronic version of the surveys. The Committee did not express concerns related to use or usability.

Comments received during public comment were related to intended use and inquiries concerning the relationship with Consumer Assessment of Healthcare Providers and Systems (CAHPS) measures as well as concerns with exclusions and the testing of CoreQ scoring.

2615 CoreQ: Long-Stay Resident Measure (AHCA): Recommended

Description: The measure calculates the percentage of long-stay residents, those living in the facility for 100 days or more, who are satisfied (see: S.5 for details of the time-frame). This patient reported

outcome measure is based on the CoreQ: Long-Stay Resident questionnaire that is a three item questionnaire.; **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Facility; **Setting of Care**: Post-Acute Care; **Data Source**: Instrument-Based Data

The Standing Committee recommended the measure for continued endorsement. This measure is based on the CoreQ: Long-Stay Resident questionnaire that is a three-item questionnaire. The Committee noted that drivers for high satisfaction rates include competency of staff, care/concern of staff, and responsiveness of management. The Committee also acknowledged that the evidence for this measure was similar to NQF 2614 and elected to carry the vote from that measure with little discussion. The performance gap discussion focused on disparities, with the Committee noting that only 6% of the data in the analysis was from Black patients. The Committee acknowledged that the reliability testing for this measure was very similar to NQF 2614 and the Committee chose to carry the vote from NQF 2614 with no discussion. The Committee acknowledged some differences in the measure exclusions and suggested that risk-factors for poor satisfaction should not be an exclusion for the measure. As such, it was concerning to the Committee that 34% of patients were excluded from the analysis because of poor cognition. The developer noted that the Brief Interview for Mental Status (BIMS) was used to determine cognition levels. It was suggested that BIMS can serve as a proxy for recall of patient satisfaction with care and that patients that are unable to recall their experience may not be able to accurately provide input related to it. The Committee also discussed not having family members serve as a proxy. It was noted that there is significant variance between measures of patient satisfaction and family satisfaction, hence the need for the third measure of family satisfaction (NQF 2616). The Committee carried the vote from NQF 2614 on feasibility, usability and use with no discussion.

Comments received during public comment were related to intended use and inquiries concerning the relationship with CAHPS measures as well as concerns with exclusions and the testing of CoreQ scoring.

2616 CoreQ: Long-Stay Family Measure (AHCA/NACL): Recommended

Description: The measure calculates the percentage of family or designated responsible party for long stay residents (i.e., residents living in the facility for 100 days or more), who are satisfied (see: S.5 for details of the timeframe). This consumer reported outcome measure is based on the CoreQ: Long-Stay Family questionnaire that has three items.; **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Facility; **Setting of Care**: Post-Acute Care; **Data Source**: Instrument-Based Data

The Standing Committee recommended the measure for continued endorsement. This consumer reported outcome measure is based on the CoreQ: Long-Stay Family questionnaire that has three items. The Committee began the discussion on evidence by asking the developer if the items of the survey are sufficiently granular to detect important problems in care delivery. The developer noted that their analysis has indicated that they are sufficiently granular, with no differences in satisfaction scores when detailed questions are used instead of general satisfaction questions. The Committee also discussed the differences between satisfaction and experience surveys. It was also noted that the family view of the care provided for patients can be very different than the patient perspective, acknowledging the need for this measure as a complement to measure NQF 2615. The Committee elected to carry the vote from NQF 2615 on both evidence and performance gap. The Committee considered the issues related to scientific accepatability to be similar to NQF 2615 and chose to carry the vote for both reliability and

validity with no discussion. The Committee also noted that the feasibility, usability and use of the measure was essentially the same to NQF 2615 and proceded to carry the vote for those criteria as well with no discussion.

Comments received during public comment were related to intended use and inquiries concerning the relationship with CAHPS measures.

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

Description: This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.; **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Facility; **Setting of Care**: Inpatient/Hospital; **Data Source**: Claims, Instrument-Based Data

The Standing Committee recommended the measure for endorsement. The Committee began the discussion of evidence by noting that the measure is based on two survey instruments, the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR). The Committee noted that for a patient in the denominator to be included in the numerator as well, they need to have a performance score improvement of 20 on the HOOS, JR or 22 on the KOOS, JR. The Committee noted the evidence provided by the developer for the analytical basis in selecting these thresholds. The Committee expressed concern about the ability of patients to interpret the meanings associated with the thresholds, as well as actual implementation by clinicians at the point of care. The Committee emphasized the need to have such measures of functional status directly integrated into workflow to allow them to inform decisions related to the direction of care. In the discussion on performance gap, the Committee noted the relatively low representation of non-Whites sampled (8.2%) which is not reflective of the general population. The developer noted that elective procedures result in differential access to care and a disproportionately White population that receives full joint replacements electively. They also noted a propensity score weighting used in their analysis as a means to compensate for this phenomenon. The Committee also expressed concern for missing differences in care in that patients with English as a second language may be screened out. The Committee reviewed the evaluation of scientific acceptability of the measure by the NQF Scientific Methods Panel (SMP). The Committee expressed no concern on the data element reliability of the measure but cited some concern on the sources of error, noting that the signal-to-noise analysis conducted using the beta binomial method described by Adams in 2009 only includes one source of provider error but that this measure potentially has several others, such as low response rate. It was noted during the validity criterion discussion that risk factors should not be exclusions for the measure, and the Committee expressed concerns that the exclusions may rule out complications associated with total joint replacement. The developer clarified that this is not the case and also that the measure removes second elective procedures. The Committee also discussed the 25-patient threshold for public reporting as well as adjusting for social drivers of health. During the feasibility discussion, the Committee noted the burden of paper scoring

methodologies for functional status measures and encouraged to developer to explore digital capture. The developer noted that the HOOS, JR and KOOS, JR have lower total items than the full scoring tool to reduce burden and that providers do not score the instrument directly. In the discussion on use and usability, the Committee noted that while the measure was commissioned by CMS, they did not provide an explanation related to the intended use of the measure or a plan for its implementation. This does not meet the NQF standard, and the Committee did not pass the measure on use. Since use is not a must-pass criterion for new measures, the measure still advanced and received a recommendation for overall endorsement.

Comments received during public comment were related to feasibility and implementation concerns regarding cost, administrative burden (providers), survey fatigue (patients), and the process for data collection. Other comments were related to the risk adjustment approach and the multi-step inclusion of social risk factors as well as concerns with the adequacy of the Standing Committee's discussion on the validity and usability criteria.

Measures Withdrawn from Consideration

Two measures previously endorsed by NQF have not been re-submitted for maintenance of endorsement or have been withdrawn during the endorsement evaluation process. Endorsement for these measures will be removed.

Table 3. Measures Withdrawn from Consideration

Measure	Reason for withdrawal
0726 Patient Experience of Psychiatric Care as Measured by the Inpatient Consumer Survey (ICS)	A shorter version of the measure is being developed.
3452 Access to Independence Promoting Services for Dual Eligible Beneficiaries	The measure is no longer supported by the steward.

References

[1] Agency for Healthcare Research and Quality (AHRQ). Priorities of the national quality strategy website. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr15/priorities.html. Last accessed February 2020.

[2] Centers for Medicare & Medicaid Services. Meaningful Measures Hub website. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/MMF/General-info-Sub-Page#Measurement_Areas. Last accessed February 2020.

[3] Frandsen BR, Joynt KE, Rebitzer JB, et al. Care fragmentation, quality, and costs among chronically ill patients. *AJMC.* 2015; 21(5):355-362.

[4] Tricco AC, Antony J, Ivers NM, et al. Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services: a systematic review and meta-analysis*. CMAJ*. 2014;186(15): E568-578.

[5] Turchi RM, Antonelli RC, Norwood KW Jr, et al. Patient-and family-centered care coordination: a framework for integrating care for children and youth across multiple systems. *Pediatrics*. 2014;133(5): e1451-e1460.

[6] Pronovost P, Weast B, Schwarz M, et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. *J Crit Care.* 2003;18(4):201-205.

[7] Gnanasakthy A, Mordin M, Evans E, et al. A review of patient-reported outcome labeling in the United States (2011-2015). *Value Health.* 2017;20(3):420-429.

[8] Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. *Med Decis Making*. 2015;35(1):114-131.

[9] Berkowitz SA, Parashuram S, Rowan K, et al. Johns Hopkins Community Health Partnership (J-CHiP) Team. Association of a care coordination model with health care costs and utilization: The Johns Hopkins Community Health Partnership (J-CHiP). *JAMA Netw Open*. 2018;1(7): e184273.

[10] Schultz EM, Pineda N, Lonhart J, et al. A systematic review of the care coordination measurement landscape. *BMC Health Serv Res*. 2013; 13:119.

[11] Rosenthal MB, Alidina S, Friedberg MW, et al. A difference-in-difference analysis of changes in quality, utilization and cost following the Colorado multi-payer patient-centered medical home pilot. *J of Gen Intern Med*. 2016;31(3):289-296.

[12] Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *New Engl J Med*. 2009;360(14):1418-1428.

[13] Frampton SB, Guastello S, Hoy L, et al. Harnessing Evidence and Experience to Change Culture: A Guiding Framework for Patient and Family Engaged Care. Washington, DC: National Academy of Medicine (NAM); 2017.

[14] CMS. *Medicare Beneficiary Characteristics.* https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Medicare_Beneficiary_Characteristics. Last accessed February 2020.

[15] United Health Foundation. America's Health Rankings website. https://www.americashealthrankings.org/explore/senior/measure/hospital_readmissions_sr/state/ALL. Last accessed February 2020.

[16] Institute of Medicine Roundtable on Evidence-Based Medicine; Yong PL, Saunders RS, Olsen LA, eds. *The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary.* Washington, DC: National Academies Press; 2010.

[17] FastStats. https://www.cdc.gov/nchs/fastats/nursing-home-care.htm. Published May 21, 2020. Last accessed July 2020.

[18] Centers for Medicare & Medicaid Services, Office of the Actuary, National Health Statistics Group, National Health Expenditure Accounts, National health expenditures. National health expenditures, average annual percent change, and percent distribution, by type of expenditure: United States, selected years 1960–2017. July 2020. https://www.cdc.gov/nchs/data/hus/2018/043.pdf.

[19] Maradit Kremers H, Larson DR, Crowson CS, et al. Prevalence of Total Hip and Knee Replacement in the United States. *J Bone Joint Surg Am*. 2015;97(17):1386-1397.

[20] Joint replacement surgery costs vary greatly between hospitals. Modern Healthcare. https://www.modernhealthcare.com/article/20180523/NEWS/180529981/joint-replacement-surgery-costs-vary-greatly-between-hospitals. Published May 23, 2018. Last accessed July 2020.

Appendix A: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable

Measures Recommended

2614 CoreQ: Short Stay Discharge Measure

Submission | Specifications

Description: The measure calculates the percentage of individuals discharged in a six month time period from a SNF, within 100 days of admission, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Short Stay Discharge questionnaire that utilizes four items.

Numerator Statement: The measure assesses the number of patients who are discharged from a SNF, within 100 days of admission, who are satisfied. The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the four questions on the CoreQ: Short Stay Discharge questionnaire.

Denominator Statement: The denominator includes all of the patients that are admitted to the SNF, regardless of payor source, for post-acute care, that are discharged within 100 days; who receive the survey (e.g. people meeting exclusions do not receive a questionnaire) and who respond to the CoreQ: Short Stay Discharge questionnaire within the time window.

Exclusions: Exclusions used are made at the time of sample selection and include:

(1) Patients who died during their SNF stay;

(2) Patients discharged to a hospital, another SNF, psychiatric facility, inpatient rehabilitation facility or long term care hospital;

(3) Patients with court appointed legal guardian for all decisions;

(4) Patients discharged on hospice;

(5) Patients who left the nursing facility against medical advice (AMA);

(6) Patients who have dementia impairing their ability to answer the questionnaire defined as having a BIMS score on the MDS as 7 or lower. [Note: we understand that some SNCCs may not have information on cognitive function available to help with sample selection. In that case, we suggest administering the survey to all residents and assume that those with cognitive impairment will not complete the survey or have someone else complete on their behalf which in either case will exclude them from the analysis.]

(7) Patients who responded after the two month response period; and

(8) Patients whose responses were filled out by someone else.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post-Acute Care

Type of Measure: Outcome: PRO-PM

Data Source: Instrument-Based Data The collection instrument is the CoreQ: Short Stay Discharge questionnaire and Resident Assessment Minimum Data Set (MDS) version 3.0.

Measure Steward: AHCA/NCAL

STANDING COMMITTEE MEETING 06/23/2020, 06/24/2020

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Pass-16; No Pass-0**; 1b. Performance Gap: **H-3**; **M-11**; **L-1**; **I-2**

Rationale:

- The Committee began the discussion of evidence for NQF 2614 by acknowledging that the developer's submission has changed minimally since the 2016 review; the evidence submitted is essentially the same, with the exception of a portion that was added in relation to the meaningfulness of the measure to patients.
- Developer provided updated statistics broken out by quarter with each quarter representing a rolling 12-month period of data.

2614 CoreQ: Short Stay Discharge Measure

- There were no statistically significant differences by race or education, but significant differences by age and gender.
- The Committee noted stable performance rates over time and had a brief discussion related to the differences in performance expectations for satisfaction and patient experience measures versus clinical measures.
- Committee questioned whether the imputation methods for incomplete surveys affected the performance scores. It was noted that regardless of the method used for imputation—either imputing the maximum or minimum score—the resulting performance scores were essentially the same.
- The Committee noted that there were no statistical differences in measure performance associated with race and expressed concern that this is incongruent with known quality problems by race in nursing facilities.
- The Committee noted that research over the last 20 years has consistently found poorer care in facilities with high minority populations and that nursing homes remain segregated, with black patients concentrated in poorer-quality homes (as measured by staffing ratios, performance, and financial vulnerability).

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-10**; **L-0**; **I-1**; 2b. Validity: **H-2**; **M-11**; **L-3**; **I-1**

<u>Rationale</u>:

- Developer used the same testing from the 2016 submission
- Measure developer performed both data element level and score level reliability testing
- Data element reliability testing included test-retest analysis on a convenience sample of 100 patients
    - Developer calculated the distribution of responses by question in the original round of surveys, and then again in the follow-up surveys (they should be distributed similarly);
    - Developer subsequently calculated the correlations between the original and follow-up responses by question (they should be highly correlated).
- The stability of the facility-level score was tested using bootstrapping with 10,000 repetitions of the facility score calculation.
    - Developer presented the percent of facility resamples where the facility score is within 1 percentage point, 3 percentage points, 5 percentage points, and 10 percentage points of the original score.
- Data element reliability testing showed very high levels of agreement and no statistically significant difference in the responses to each question between the original and re-test results.
- Validity testing of the CoreQ: Short Stay Discharge questionnaire included both data element level and score level testing.
- Data element level
    - Exploratory factor analysis (EFA) was used to further refine the pilot instrument. This was an iterative process that included using Eigenvalues from the principal factors (unrotated) and correlation analysis of the individual items.
    - Correlation analysis and a factor analysis conducted
    - Face validity evaluated via literature review and review of 12 commonly used satisfaction surveys; also examined face validity of domains and the response scale, using 40 patients in 5 nursing homes. The Flesch-Kinkaid scale was used to determine if patients understood the questions.
    - Also examined correlation between the four items in the measure and all the items on the pilot instrument.
- Measure score level
- Convergent validity testing was performed. Developers examined correlations between the CoreQ measure scores and i) measures of regulatory compliance and other quality metrics from the Certification and Survey Provider Enhanced Reporting (CASPER) data, ii) several other quality metrics from Nursing Home Compare, iii) risk adjusted discharge to community measure and iv) risk adjusted PointRight® Pro 30™ Rehospitalizations.

2614 CoreQ: Short Stay Discharge Measure

- The Committee noted in the discussion of scientific acceptability that the submission was the same as the previous submission in 2016.
- The Committee questioned the exclusion of surveys that were completed by proxy, but noted that many survey firms exclude satisfaction surveys completed by someone other than the person who receives the service.
- The Committee requested the developer to explore the payment source (Medicare, Medicaid) for the purposes of investigating case mix adjustment.
- In the discussion on validity, the Committee noted in the convergent validity analyses conducted by the developer with external measures of quality that many of the measures did not exhibit high convergence with NQF 2614. It was suggested that this is also true of other measures of satisfaction; it is not unsual for such measures to be independent of other measures of quality.

3. Feasibility: H-2; M-14; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- This is a satisfaction survey conducted via mailed survey.
- No fees required to use the measure; the developer did not indicate if there are fees associated with the use of the survey.
- The Committee discussed the feasibility of the measure and noted that it was similar to other measures of patient satisfaction in terms of its overall burden to providers and patients.
- The Committee suggested that the developer consider an electronic version of the surveys.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Pass-15; No Pass-1**; 4b. Usability: **H-2; M-14; L-0; I-1**

Rationale:

- The measure is used in several accountability applications
- Professional Certification or Recognition Program
  - AHCA Quality Initiative
  - AHCA Quality Awards
- Quality Improvement (external benchmarking to organizations)
  - AHCA NCAL Long Term Care Trend Tracker
- Developer notes that a number of states are implementing the CoreQ survey inside of state incentive programs, including NJ, MA, TN, GA, and others.
- The Committee did not express concerns related to use or usability.

5. Related and Competing Measures

- No competing measures noted.

6. Standing Committee Recommendation for Endorsement: Yes-15; No-2

7. Public and Member Comment

Comments received during public comment were related to intended use and inquiries concerning the relationship with CAHPS measures as well as concerns with exclusions and the testing of CoreQ scoring.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020): Yes-X; No-X

9. Appeals

2615 CoreQ: Long-Stay Resident Measure

Submission | Specifications

Description The measure calculates the percentage of long-stay residents, those living in the facility for 100 days or more, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Long-Stay Resident questionnaire that is a three item questionnaire.

Numerator Statement: The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long -Stay Resident questionnaire.

Denominator Statement: The denominator includes all of the residents that have been in the SNF for 100 days or more regardless of payer status; who received the CoreQ: Long-Stay Resident questionnaire (e.g. people meeting exclusions do not receive the questionnaire), who responded to the questionnaire within the two month time window, who did not have the questionnaire completed by somebody other than the resident, and who did not have more than one item missing.

Exclusions: Exclusions made at the time of sample selection are the following: (1) Residents who have poor cognition defined by the BIMS score; (2) residents receiving hospice; (3) residents with a legal court appointed guardian; and (4) residents who have lived in the SNF for less than 100 days.

Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (two months after the administration date) b) surveys that have more than one questionnaire item missing c) surveys from residents who indicate that someone else answered the questions for the resident. (Note this does not include cases where the resident solely had help such as reading the questions or writing down their responses.)

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post-Acute Care

Type of Measure: Outcome: PRO-PM

Data Source: Instrument-Based Data The collection instrument is the CoreQ: Long-Stay Resident questionnaire and exclusions are from the Resident Assessment Instrument Minimum Data Set (MDS) version 3.0.

Measure Steward: American Health Care Association

STANDING COMMITTEE MEETING 07/09/2020

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Pass-16; No Pass-0**; 1b. Performance Gap: **H-1**; **M-12**; **L-2**; **I-1**;

Rationale:

- The Committee began the discussion of evidence for NQF 2615 by acknowledging that the developer's submission has changed minimally since the 2016 review; the evidence submitted is essentially the same, with the exception of a portion that was added related to the meaningfulness of the measure to patients.
- The Committee noted that drivers for high satisfaction rates include competency of staff, care/concern of staff, and responsiveness of management.
- The Committee also acknowledged that the evidence for this measure was similar to NQF 2614 and elected to carry the vote from the previous with little discussion.
- Developer provided updated statistics broken out by quarter with each quarter representing a rolling 12-month period of data.
- There were no statistically significant differences by race or education, but significant differences by age and gender.
- The performance gap discussion focused on disparities, with the Committee noting that only 6% of the data in the analysis was from Black patients.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-10**; **L-0**; **I-1**; 2b. Validity: **H-1**; **M-11**; **L-3**; **I-0**

Rationale:

2615 CoreQ: Long-Stay Resident Measure

- Developer used the same testing from the 2016 submission.
- Measure developer performed both data element level and score level reliability testing.
- Data element reliability testing included test-retest analysis on a convenience sample of 100 patients.
    - Developer calculated the distribution of responses by question in the original round of surveys, and then again in the follow-up surveys (they should be distributed similarly);
    - Developer subsequently calculated the correlations between the original and follow-up responses by question (they should be highly correlated).
- The stability of the facility-level score was tested using bootstrapping with 10,000 repetitions of the facility score calculation.
    - Developer presented the percent of facility resamples where the facility score is within 1 percentage point, 3 percentage points, 5 percentage points, and 10 percentage points of the original score.
- Data element reliability testing showed very high levels of agreement and no statistically significant difference in the responses to each question between the original and re-test results.
- Validity testing of the CoreQ: Short Stay Discharge questionnaire included both data element level and score level testing.
- Data element level
    - Exploratory factor analysis (EFA) was used to further refine the pilot instrument. This was an iterative process that included using Eigenvalues from the principal factors (unrotated) and correlation analysis of the individual items.
    - Correlation analysis and a factor analysis conducted
    - Face validity evaluated via literature review and review of 12 commonly used satisfaction surveys; also examined face validity of domains and the response scale, using 40 patients in 5 nursing homes. The Flesch-Kinkaid scale was used to determine if patients understood the questions.
    - Also examined correlation between the four items in the measure and all the items on the pilot instrument.
- Measure score level
    - Convergent validity testing was performed.
    - Developers examined correlations between the CoreQ measure scores and i) measures of regulatory compliance and other quality metrics from the Certification and Survey Provider Enhanced Reporting (CASPER) data, ii) several other quality metrics from Nursing Home Compare, iii) risk adjusted discharge to community measure and iv) risk adjusted PointRight® Pro 30™ Rehospitalizations.
- The Committee acknowledged that the reliability testing for this measure was very similar to NQF 2614 and the Committee chose to carry the vote from NQF 2614 with no discussion.
- The Committee acknowledged some differences in the measure exclusions and suggested that risk-factors for poor satisfaction should not be an exclusion for the measure.
    - As such, it was concerning to the Committee that 34% of patients were excluded from the analysis because of poor cognition.
    - The developer noted that the Brief Interview for Mental Status (BIMS) was used to determine cognition levels.
    - It was suggested that BIMS can serve as a proxy for recall of patient satisfaction with care and that patients that are unable to recall their experience may not be able to accurately provide input related to it.
    - The Committee also discussed not having family members serve as a proxy. It was noted that there is significant variance between measures of patient satisfaction and family satisfaction, hence the need for the third measure of family satisfaction (NQF 2616).

3. Feasibility: H-2; M-14; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- This is a satisfaction survey conducted via mailed survey.

2615 CoreQ: Long-Stay Resident Measure

- No fees required to use the measure; the developer did not indicate if there are fees associated with the use of the survey.
- The Committee carried the vote from measure NQF 2614 with no discussion.

4. Usability and Use: <u>The maintenance measure meets the Use subcriterion</u>

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Pass-15; No Pass-1**; 4b. Usability: **H-2; M-14; L-0; I-1**

<u>Rationale</u>:

- The measure is used in several accountability applications.
- Professional Certification or Recognition Program
  - <u>AHCA Quality Initiative</u>
  - <u>AHCA Quality Awards</u>
- Quality Improvement (external benchmarking to organizations)
  - <u>AHCA NCAL Long Term Care Trend Tracker</u>
- Developer notes that a number of states are implementing the CoreQ survey inside of state incentive programs, including NJ, MA, TN, GA, and others.
- The Committee carried the vote from measure NQF 2614 with no discussion.

5. Related and Competing Measures

- No competing measures noted.

6. Standing Committee Recommendation for Endorsement: Yes-14; **No-2**

7. Public and Member Comment

Comments received during public comment were related to intended use and inquiries concerning the relationship with CAHPS measures as well as concerns with exclusions and the testing of CoreQ scoring.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020): Yes-X; No-X

9. Appeals

2616 CoreQ: CoreQ: Long-Stay Family Measure

<u>Submission</u> | <u>Specifications</u>

Description: The measure calculates the percentage of family or designated responsible party for long stay residents (i.e., residents living in the facility for 100 days or more), who are satisfied (see: S.5 for details of the timeframe). This consumer reported outcome measure is based on the CoreQ: Long-Stay Family questionnaire that has three items.

Numerator Statement: The numerator assesses the number of family or designated responsible party for long stay residents that are satisfied. Specifically, the numerator is the sum of the family or designated responsible party members for long stay residents that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long-Stay Family questionnaire.

Denominator Statement The target population is family or designated responsible party members of a resident residing in a SNF for at least 100 days. The denominator includes all of the individuals in the target population who respond to the CoreQ: Long-Stay Family questionnaire within the two month time window (see S.5) who do not meet the exclusion criteria (see S.10).

Exclusions: Please note, the resident representative for each current resident is initially eligible regardless of their being a family member or not. Only one primary contact per resident should be selected.

Exclusions made at the time of sample selection include: (1) family or designated responsible party for residents with hospice; (2) family or designated responsible party for residents with a legal court appointed guardian; (3) representatives of residents who have lived in the SNF for less than 100 days; and (4) representatives who reside in another country.

2616 CoreQ: CoreQ: Long-Stay Family Measure

Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (more than two months after the administration date) and b) surveys that have more than one questionnaire item missing.

Adjustment/Stratification: No risk adjustment or risk stratification
Level of Analysis: Facility
Setting of Care: Post-Acute Care
Type of Measure: Outcome: PRO-PM
Data Source: Instrument-Based Data The collection instrument is the CoreQ: Long-Stay Family questionnaire and for exclusions the Resident Assessment Instrument Minimum Data Set (MDS) version 3.0 is used
Measure Steward: AHCA/NCAL

STANDING COMMITTEE MEETING 07/09/2020

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Pass-16; No Pass-0**; 1b. Performance Gap: **H-1**; **M-12**; **L-2**; **I-1**;

Rationale:

- The Committee began the discussion of evidence for NQF 2616 by acknowledging that the developer's submission has changed minimally since the 2016 review; the evidence that was submitted is essentially the same, with the exception of a portion that was added related to the meaningfulness of the measure to patients.
- Developer provided updated statistics broken out by quarter with each quarter representing a rolling 12-month period of data.
- There were no statistically significant differences by race or education, but significant differences by age and gender.
- The Committee began the discussion on evidence by asking the developer if the items of the survey are sufficiently granular to detect important problems in care delivery. The developer noted that their analysis has indicated that they are sufficiently granular, with no differences in satisfaction scores when detailed questions are used instead of general satisfaction questions.
- The Committee also discussed the differences between satisfaction and experience surveys.
- It was also noted that the family view of the care provided for patients can be very different than the patient perspective, acknowledging the need for this measure as a complement to measure NQF 2615.
- The Committee elected to carry the vote from NQF 2615 on both evidence and performance gap.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-10**; **L-0**; **I-1**; 2b. Validity: **H-1**; **M-11**; **L-3**; **I-0**

Rationale:

- Developer used the same testing from the 2016 submission.
- Measure developer performed both data element level and score level reliability testing.
- Data element reliability testing included test-retest analysis on a convenience sample of 100 patients:
    - Developer calculated the distribution of responses by question in the original round of surveys, and then again in the follow-up surveys (they should be distributed similarly);
    - Developer subsequently calculated the correlations between the original and follow-up responses by question (they should be highly correlated).
- The stability of the facility-level score was tested using bootstrapping with 10,000 repetitions of the facility score calculation.
    - Developer presented the percent of facility resamples where the facility score is within 1 percentage point, 3 percentage points, 5 percentage points, and 10 percentage points of the original score.
- Data element reliability testing showed very high levels of agreement and no statistically significant difference in the responses to each question between the original and re-test results.
- Validity testing of the CoreQ: Short Stay Discharge questionnaire included both data element level and score level testing.

2616 CoreQ: CoreQ: Long-Stay Family Measure

- Data element level
  - Exploratory factor analysis (EFA) was used to further refine the pilot instrument. This was an iterative process that included using Eigenvalues from the principal factors (unrotated) and correlation analysis of the individual items.
  - Correlation analysis and a factor analysis conducted.
  - Face validity evaluated via literature review and review of 12 commonly used satisfaction surveys; also examined face validity of domains and the response scale, using 40 patients in 5 nursing homes. The Flesch-Kinkaid scale was used to determine if patients understood the questions.
  - Also examined correlation between the four items in the measure and all the items on the pilot instrument.
- Measure score level
- Convergent validity testing was performed. Developers examined correlations between the CoreQ measure scores and I) measures of regulatory compliance and other quality metrics from the Certification and Survey Provider Enhanced Reporting (CASPER) data, ii) several other quality metrics from Nursing Home Compare, iii) risk adjusted discharge to community measure and iv) risk adjusted PointRight® Pro 30™ Rehospitalizations.
- The Committee considered the issues related to scientific accepatability to be similar to NQF 2615 and chose to carry the vote for both reliability and validity with no discussion.

3. Feasibility: H-2; M-14; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- This is a satisfaction survey conducted via mailed survey.
- No fees required to use the measure; the developer did not indicate if there are fees associated with the use of the survey.
- The Committee carried the vote from measure NQF 2615 with no discussion.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Pass-15; No Pass-1**; 4b. Usability: **H-2; M-14; L-0; I-1**

Rationale:

- The measure is used in several accountability applications.
- Professional Certification or Recognition Program
  - AHCA Quality Initiative
  - AHCA Quality Awards
- Quality Improvement (external benchmarking to organizations)
  - AHCA NCAL Long Term Care Trend Tracker
- Developer notes that a number of states are implementing the CoreQ survey inside of state incentive programs, including NJ, MA, TN, GA, and others.
- The Committee carried the vote from measure NQF 2615 with no discussion.

5. Related and Competing Measures

- No competing measures noted.

6. Standing Committee Recommendation for Endorsement: Yes-14; **No-2**

7. Public and Member Comment

Comments received during public comment were related to intended use and inquiries concerning the relationship with CAHPS measures.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020): Yes-X; No-X

9. Appeals

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

Submission | Specifications

Description: This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.

Numerator Statement: The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows:

- For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and

- For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR).

SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group.

References:

Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.

Denominator Statement The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.

Exclusions: Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.

Adjustment/Stratification: Statistical risk model. No risk stratification

Level of Analysis: Facility

Setting of Care: Inpatient/Hospital

Type of Measure: Outcome: PRO-PM

Data Source: Claims, Instrument-Based Data The PROM surveys used to define the measure outcome are 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are presently available in English, not yet in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) Global or the Veterans RAND 12 Item Health Survey (VR-12) are used. The PROMIS Global is available in sixteen languages; the VR-12 is available in Spanish, Chinese and German.

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

STANDING COMMITTEE MEETING 06/23/2020, 06/24/2020

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Pass-17; No Pass-0**; 1b. Performance Gap: **H-4**; **M-9**; **L-2**; **I-0**;

Rationale:

- Developer notes that there are many studies indicating how providers can improve outcomes of the patients by addressing aspects of pre-, peri-, and postoperative care (Brown et al., 2012; Choong et al.,

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

2009; Galea et al., 2008; Kim, 2019; McGregor et al., 2004; Moffet et al., 2004; Monticone et al., 2013; Walters, 2016).

- Optimal clinical outcomes may be influenced by:
  - The surgeon performing the procedure
  - Team's efforts in the care of the patient
  - Care coordination across provider groups and specialties
  - Patients' engagement in their own recovery (Feng et al, 2018; Saufl et al, 2007).
- The Committee began the discussion of evidence by noting that the measure is based on two survey instruments, the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR).
  - The Committee noted that for a patient in the denominator to be included in the numerator as well, they need to have a performance score improvement of 20 on the HOOS, JR or 22 on the KOOS, JR.
  - The Committee noted the analysis conducted by Stephen Lyman, et al. for the analytical basis in selecting these thresholds.
- The Committee expressed concern about the ability of patients to interpret the meanings associated with the thresholds, as well as actual implementation by clinicians at the point of care.
- The Committee emphasized the need to have such measures of functional status directly integrated into workflow to allow them to inform decisions related to the direction of care.
- Developer provided disparities data for n=6,734 patients within the Development Dataset, analyzing race, dual eligibility status, and socioeconomic status (SES).
- Chi-square analyses and multivariate analyses did not reveal a statistically significant association between non-White race or SES.
- Dual eligibility was borderline significant (p=0.058) at the bivariate level and statistically significant within the risk model.
- In the discussion on performance gap, the Committee noted the relatively low representation of non-Whites sampled (8.2%) which is not reflective of the general population.
  - The developer noted that elective procedures result in differential access to care and a disproportionately White population that receives full joint replacements electively.
  - They also noted a propensity score weighting used in their analysis as a means to compensate for this phenomenon.
- The Committee also expressed concern for missing differences in care in that patients with English as a second language may be screened out.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: SMP Rating-**High**; Vote to uphold the SMP's rating: **Yes-14 No-1**; 2b. Validity: SMP Rating-**Moderate;** Vote to uphold the SMP's rating: **Yes-10 No-4**

<u>Rationale</u>:

- Reliability testing conducted at the data element and score level.
- Data element reliability testing assessed consistency and test-retest reliability of the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) and Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) instruments.
  - HOOS, JR internal consistency using Person Separation Index (PSI) was 0.86 and 0.87 in the two cohorts tested.
  - HOOS, JR test-retest results produced ICCs between 0.75 and 0.97.
  - KOOS, JR internal consistency using PSI was 0.84 and 0.85 in the two cohorts tested.
  - KOOS, JR test-retest results produced ICCs between 0.75 and 0.93.
- Score level reliability testing consisted of a signal-to-noise analysis. Results from a sample of 123 hospitals yielded a mean of 0.95 and a range from 0.90 to 0.99.
- Validity testing was conducted at both the data element and score levels.

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

- Data element validity testing included responsiveness, external validity, floor and ceiling effects for both HOOS, JR and KOOS, JR. HOOS, JR responsiveness produced standardized response means relative to other PROMs (HOOS domains, The Western Ontario and McMaster University Arthritis Index [WOMAC] domains) measuring post-surgery hip improvement of 2.38 and 2.03 in the two samples.
    - HOOS, JR external validity used Spearman's correlation analysis with the HOOS and WOMAC instruments and produced 0.87 for both samples.
    - HOOS, JR showed floor (0.6%–1.9%) and ceiling (37%–46%) effects and were comparable to or better than HOOS domains and the WOMAC.
    - KOOS, JR responsiveness produced standardized response means relative to other PROMs (KOOS, WOMAC) measuring post-surgery hip improvement of 1.79 and 1.70 in the two samples.
    - KOOS, JR external validity used Spearman's correlation analysis with the KOOS and WOMAC instruments and produced 0.89 and 0.91 for the two samples.
    - KOOS, JR showed floor (0.4%–1.2%) and ceiling (18.8%–21.8%) effects.
- Score level validity testing included empirical comparisons to another quality measure: NQF 1550 Hospital-level risk-standardized complication rate (RSCR) following elective primary THA/TKA. Comparison of THA/TKA PRO-PM RSIRs to RSCR categories indicates an increasing monotonic trend. Those hospitals in the "RSCR Worse than National Average" category have lower median RSIRs (51.87%) than the median RSIR (66.49%) of hospitals in the "RSCR Same as National Average" category, which is lower than that of hospitals in the "RSCR Better than National Average" category (71.13%).
- The Committee reviewed the evaluation of scientific acceptability of the measure by the NQF SMP.
- The Committee expressed no concern on the data element reliability of the measure but cited some concern on the sources of error, noting that the signal to noise analysis conducted using the beta binomial method described by Adams in 2009 only includes one source of provider error but that this measure potentially has several others, such as low response rate.
- It was noted during the validity discussion that risk factors should not be exclusions for the measure and the Committee expressed concern that the exclusions may rule out complications associated with total joint replacement. The developer clarified that this is not the case and that the measure removes second elective procedures.
- The Committee also discussed the 25-patient threshold for public reporting as well as adjusting for social drivers of health.

3. Feasibility: H-0; M-11; L-4; I-1

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- During the feasibility discussion, the Committee noted the burden of paper scoring methodologies for functional status measures and encouraged to developer to explore digital capture.
- The developer noted that the HOOS, JR and KOOS, JR have lower total items than the full scoring tool to reduce burden and that providers do not score the instrument directly.

4. Usability and Use: The maintenance measure meets the Use sub criterion

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Pass-6; No Pass-10** 4b. Usability: **H-0; M-8; L-3; I-5**

Rationale:

- In the discussion on use and usability, the Committee noted that while the measure was commissioned by CMS, they did not provide an explanation related to the intended use of the measure or a plan for its implementation.
- This does not meet the NQF standard and the Committee did not pass the measure on use (not a must-pass criterion for new measures).

5. Related and Competing Measures

3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
• No competing measures noted.
6. Standing Committee Recommendation for Endorsement: Yes-14; No-2
7. Public and Member Comment Comments received during public comment were related to feasibility and implementation concerns regarding cost, administrative burden (providers), survey fatigue (patients), and the process for data collection. Other comments were related to the risk adjustment approach and the multi-step inclusion of social risk factors as well as concerns with the adequacy of the Standing Committee discussion on validity and usability.
8. Consensus Standards Approval Committee (CSAC) Endorsement Decision (November 17-18, 2020): Yes-X; No-X
9. Appeals

Appendix B: Patient Experience and Function Portfolio—Use in Federal Programs[a]

NQF #	Title	Federal Programs: Finalized or Implemented as of June 22, 2020
0005	CAHPS Clinician & Group Surveys (CG-CAHPS)-Adult, Child	Merit-based Incentive Payment System (MIPS) Program (Implemented) Physician Compare (Implemented)
0006	Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey, Version 5.0 (Medicaid and Commercial)	Marketplace Quality Rating System (QRS) (Implemented) Medicare Shared Savings Program (Shared Savings Program) (Implemented)
0166	HCAHPS	Hospital Compare (Implemented) Hospital Inpatient Quality Reporting (IQR) (Implemented) Hospital Value-Based Purchasing (HVBP) (Implemented) Prospective Payment System (PPS)-Exempt Cancer Hospital Quality Reporting (PCHQR) (Implemented)
0258	CAHPS In-Center Hemodialysis Survey	End-Stage Renal Disease Quality Incentive Program (ESRD QIP) (Implemented) Dialysis Facility Compare (DFC) (Implemented)
0291	EMERGENCY TRANSFER COMMUNICATION MEASURE	None
0422	Functional status change for patients with Knee impairments	None
0423	Functional status change for patients with Hip impairments	None
0424	Functional status change for patients with Foot and Ankle impairments	None
0425	Functional status change for patients with lumbar impairments	None
0426	Functional status change for patients with Shoulder impairments	None
0427	Functional status change for patients with elbow, wrist, and hand impairments	None
0428	Functional status change for patients with General orthopedic impairments	None
0517	CAHPS® Home Health Care Survey (experience with care)	None

[a] Per CMS Measures Inventory Tool as of 06/22/2020

NQF #	Title	Federal Programs: Finalized or Implemented as of June 22, 2020
1741	Patient Experience with Surgical Care Based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey	None
2286	Functional Change: Change in Self Care Score	None
2287	Functional Change: Change in Motor Score	None
2321	Functional Change: Change in Mobility Score	None
2483	Gains in Patient Activation (PAM) Scores at 12 Months	None
2548	Child Hospital CAHPS (HCAHPS)	None
2612	CARE: Improvement in Mobility	None
2613	CARE: Improvement in Self Care	None
2614	CoreQ: Short Stay Discharge Measure	None
2615	CoreQ: Long-Stay Resident Measure	None
2616	CoreQ: Long-Stay Family Measure	None
2631	Percent of Long-Term Care Hospital (LTCH) Patients with an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function	Long-Term Care Hospital Quality Reporting (LTCH QRP) (Implemented)
2632	Long-Term Care Hospital (LTCH) Functional Outcome Measure: Change in Mobility Among Patients Requiring Ventilator Support	LTCH QRP (Implemented) Long-Term Care Hospital (LTCH) Compare (Implemented)
2633	Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients	Inpatient Rehabilitation Facility Quality Reporting (IRF QRP) (Implemented) Inpatient Rehabilitation Facility (IRF) Compare (Implemented)
2634	Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients	IRF QRP (Implemented) IRF Compare (Implemented)
2635	Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients	IRF QRP (Implemented) IRF Compare (Implemented)

NQF #	Title	Federal Programs: Finalized or Implemented as of June 22, 2020
2636	Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients	IRF QRP (Implemented) IRF Compare (Implemented)
2643	Average change in functional status following lumbar spine fusion surgery	MIPS Program (Finalized)
2653	Average change in functional status following total knee replacement surgery	MIPS Program (Finalized)
2769	Functional Change: Change in Self Care Score for Skilled Nursing Facilities	None
2774	Functional Change: Change in Mobility Score for Skilled Nursing Facilities	None
2775	Functional Change: Change in Motor Score for Skilled Nursing Facilities	None
2776	Functional Change: Change in Motor Score in Long Term Acute Care Facilities	None
2777	Functional Change: Change in Self Care Score for Long Term Acute Care Facilities	None
2778	Functional Change: Change in Mobility Score for Long Term Acute Care Facilities	None
2958	Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery	None
2962	Shared Decision-Making Process	None
2967	CAHPS® Home- and Community-Based Services Measures	Medicaid (Implemented)
3227	CollaboRATE Shared Decision-Making Score	None
3420	CoreQ: AL Resident Satisfaction Measure	None
3422	CoreQ: AL Family Satisfaction Measure	None
3455	Timely Follow-Up After Acute Exacerbations of Chronic Conditions	None
3461	Functional Status Change for Patients with Neck Impairments	None
3477	Discharge to Community-Post Acute Care Measure for Home Health Agencies	Home Health Value Based Purchasing (HHVBP) (Implemented) Home Health Compare (Implemented)

NQF #	Title	Federal Programs: Finalized or Implemented as of June 22, 2020
3479	Discharge to Community-Post Acute Care Measure for Inpatient Rehabilitation Facilities	IRF QRP (Implemented) IRF Compare (Implemented)
3480	Discharge to Community-Post Acute Care Measure for Long-Term Care Hospitals	LTCH QRP (Implemented) LTCH Compare (Implemented)
3481	Discharge to Community-Post Acute Care Measure for Skilled Nursing Facilities	Skilled Nursing Facility Quality Reporting (SNF QRP) (Implemented)

Appendix C: Patient Experience and Function Standing Committee and NQF Staff

STANDING COMMITTEE

Gerri Lamb, PhD, RN, FAAN (Co-chair)
Associate Professor, Arizona State University
Tucson, AZ

Christopher Stille, MD, MPH, FAAP (Co-chair)
University of Colorado School of Medicine
Aurora, Colorado

Richard Antonelli, MD, MS
Medical Director for Integrated Care, Boston Children's Hospital, Harvard Medical School
Boston, Massachusetts

Adrienne Boissy, MD, MA
Chief Experience Officer, Cleveland Clinic
Cleveland, Ohio

Donald Casey, MD, MPH, MBA, FACP, FAHA, FAAPL, DFACMQ
President, American College of Medical Quality (ACMQ)
Chicago, Illinois

Ariel Cole, MD
Clerkship Director for Geriatrics, Florida State University College of Medicine Orlando campus
Winter Park, Florida

Ryan Coller, MD, MPH
Division Chief, Pediatric Hospital Medicine, University of Wisconsin-Madison
Madison, Wisconsin

Sharon Cross, LISW-S
Patient/Family Centered Care Program Director, The Ohio State University Wexner Medical Center
Columbus, Ohio

Christopher Dezii, MBA, RN, CPHQ
Lead, Healthcare Quality & Performance Measures, Bristol-Myers Squibb Company
Lawrenceville, New Jersey

Shari Erickson, MPH
Vice President, Governmental & Regulatory Affairs, American College of Physicians (ACP)
Washington, District of Columbia

Dawn Hohl, RN, BSN, MS, PhD
Sr. Director of Transitions and Patient Experience, Johns Hopkins Home Care Group
Baltimore, Maryland

Sherrie Kaplan, PhD, MPH
University of California Irvine School of Medicine
Irvine, California

Tracy Kusnir, MBA
Director of Value and Patient Experience, Seattle Cancer Care Alliance
Seattle, Washington

Brenda Leath, MHSA, PMP
President/CEO, Leath & Associates, LLC
Washington, District of Columbia

Brian Lindberg, BSW, MMHS
Executive Director, Consumer Coalition for Quality Health Care
Washington, District of Columbia

Lisa Morrise, MA
Patient & Family Engagement Affinity Group National Partnership for Patients
Salt Lake City, Utah

Randi Oster, MBA
President, Help Me Health
Fairfield, Connecticut

Charissa Pacella, MD
Chief of Emergency Services and Medical Staff, University of Pittsburgh Medical Center (UPMC)
Pittsburgh, Pennsylvania

Lenard Parisi, RN, MA, CPHQ, FNAHQ
Senior Operational Consultant, Strategic Quality Solutions LLC
New York, New York

Debra Saliba, MD, MPH
Professor of Medicine, UCLA/JH Borun Center, VA GRECC, RAND Health
Los Angeles, California

Ellen Schultz, MS
American Institutes for Research
Chicago, Illinois

Lisa Suter, MD
Assistant Professor and Associate Director, Yale School of Medicine, and Yale/CORE
New Haven, Connecticut

Peter Thomas, JD
Principal, Powers, Pyles, Sutter & Verville, P.C.
Washington, District of Columbia

NQF STAFF

Sherri Winsper, RN, MSN, MSHA
Senior Vice President, Quality Measurement

Apryl Clark, MHSA
Acting Vice President, Quality Measurement

Sai Ma, MPA, PhD
Managing Director/Senior Technical Expert, Quality Measurement

Sam Stolpe, PharmD, MPH
Senior Director

Oroma Igwe, MPH
Manager

Yemsrach Kidane, PMP
Project Manager

Udobi Onyeuku, MSHA
Analyst

Appendix D: Measure Specifications

	2614 CoreQ: Short Stay Discharge Measure
Steward	AHCA/NCAL
Description	The measure calculates the percentage of individuals discharged in a six month time period from a SNF, within 100 days of admission, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Short Stay Discharge questionnaire that utilizes four items.
Type	Outcome: PRO-PM
Data Source	Instrument-Based Data The collection instrument is the CoreQ: Short Stay Discharge questionnaire and Resident Assessment Minimum Data Set (MDS) version 3.0.
Level	Facility
Setting	Post-Acute Care
Numerator Statement	The measure assesses the number of patients who are discharged from a SNF, within 100 days of admission, who are satisfied. The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the four questions on the CoreQ: Short Stay Discharge questionnaire.
Numerator Details	The numerator includes all of the patients who were discharged within 100 days of admission and had an average response =>3 on the CoreQ: Short Stay Discharge questionnaire. The calculation of the individual patient's average satisfaction score is done in the following manner: -A numeric score is associated with each response scale option on the CoreQ: Short Stay Discharge questionnaire (that is, Poor=1, Average=2, Good=3, Very Good=4, and Excellent=5). -The following formula is utilized to calculate the individual's average satisfaction score: [Numeric Score Question 1 + Numeric Score Question 2 + Numeric Score Question 3 + Numeric Score Question 4]/4 -The number of respondents whose average satisfaction score >=3 are summed together and function as the numerator. For patients with one missing data point (from the four items included in the questionnaire) imputation is used (representing the average value from the other three available responses). Patients with more than one missing data point, are excluded from the analyses (i.e., no imputation will be used for these patients). Imputation details are described further below (S.22). No risk-adjustment is used (See S.18).
Denominator Statement	The denominator includes all of the patients that are admitted to the SNF, regardless of payor source, for post-acute care, that are discharged within 100 days; who receive the survey (e.g. people meeting exclusions do not receive a questionnaire) and who respond to the CoreQ: Short Stay Discharge questionnaire within the time window.
Denominator Details	The target population includes all of the individuals who respond to the CoreQ: Short Stay Discharge questionnaire within the time window (See: S.5). The data is collected over a maximum 6 month time window. A shorter period can be used if the sample size (125) meets the specifications described below. The questionnaire is administered to discharged patients within 2 weeks of their discharge date. The discharge date is identified from nursing facility records (e.g., MDS, wherein a discharge MDS record is created that includes a discharge date). Note, the questionnaire must be administered after the patient is discharged and not on the day of the discharge. Patients must respond

	2614 CoreQ: Short Stay Discharge Measure
	to the CoreQ: Short Stay Discharge questionnaire within 2 months of receiving the questionnaire.
Exclusions	Exclusions used are made at the time of sample selection and include: (1) Patients who died during their SNF stay; (2) Patients discharged to a hospital, another SNF, psychiatric facility, inpatient rehabilitation facility or long term care hospital; (3) Patients with court appointed legal guardian for all decisions; (4) Patients discharged on hospice; (5) Patients who left the nursing facility against medical advice (AMA); (6) Patients who have dementia impairing their ability to answer the questionnaire defined as having a BIMS score on the MDS as 7 or lower. [Note: we understand that some SNCCs may not have information on cognitive function available to help with sample selection. In that case, we suggest administering the survey to all residents and assume that those with cognitive impairment will not complete the survey or have someone else complete on their behalf which in either case will exclude them from the analysis.] (7) Patients who responded after the two month response period; and (8) Patients whose responses were filled out by someone else.
Exclusion details	Individuals are excluded based on information from the admission Minimum Data Set (MDS) 3.0 assessment. (1) Patients who die: This is recorded in the MDS as Deceased (A2100 = 08). (2) Patients who were discharged to a hospital, another SNCC, psychiatric facility, Inpatient Rehabilitation Facilities (IRF), or MR/DD facility: This is recorded in the MDS as Discharge to hospital (A2100 = 03); another SNCC (A2100 = 02); psychiatric facility (A2100 = 04); Inpatient Rehabilitation Facilities (A2100 = 05); ID/DD facility (A2100 = 06). (3) Patients with Court appointed legal guardian for all decisions as identified from the nursing facility health information system. (4) Patients on hospice: This is recorded in the MDS as Hospice O0100K1 = 1 (“the patient was on hospice in the last 14 days while not a resident”), O0100K2 = 1 (“the patient was on hospice in the last 14 days while a resident”), A1800=07 (“entered from hospice”), or A2100=07 (“discharged to hospice”). (5) Patients who left the nursing facility against medical advice (AMA) as identified from nursing facility health information systems. (6) Patients with a BIMS score on the MDS as 7 or lower. This is recorded in the MDS as C0500 <= 7. (7) Patients who respond after the two month response period. (8) Patients whose responses were filled out by somebody other than him/herself, as identified by the additional questions on the questionnaire. Surveys returned as undeliverable are also excluded from the denominator.
Risk Adjustment	No risk adjustment or risk stratification
Stratification	No stratification is used.
Type Score	Other (specify): Non-weighted score. Score is a percentage. better quality = higher score
Algorithm	1.Identify SNF patients that are discharged within 100 days after admission a.Calculate the duration of the SNF stay [MDS discharge date (A2000) - MDS admission date (A1900)] to determine if it is = 100 days. 2.Take the patients that have a SNF stay of = 100 days and exclude the following: a.Patients who died; patients discharged to a hospital; patients with Court appointed legal guardian for all decisions; patients with hospice; patients who left the nursing facility

	2614 CoreQ: Short Stay Discharge Measure
	against medical advice (AMA), and patients with a BIMS score of less than 7 do not receive that survey as a result of the exclusions (described in detail above). i.Patients who die: This is recorded in the MDS as Die during stay (A2100 = 08) ii.Patients who were discharged to a hospital, another SNCC, psychiatric facility, Inpatient Rehabilitation Facility, or MR/DD facility (A2100 = 06): This is recorded in the MDS as Discharge to hospital (A2100 = 03); another SNCC (A2100 = 02); psychiatric facility (A2100 = 04); Inpatient Rehabilitation Facility (A2100 = 05); MR/DD facility (A2100 = 06). iii.Patients with Court appointed legal guardian for all decisions will be identified from nursing facility health information system. iv.Patients on hospice: This is recorded in the MDS as Hospice O0100K1 = 1 ("the patient was on hospice in the last 14 days while not a resident"), O0100K2 = 1 ("the patient was on hospice in the last 14 days while a resident"), A1800=07 ("entered from hospice"), or A2100=07 ("discharged to hospice"). v.Patients who left the nursing facility against medical advice (AMA) will be identified from nursing facility health information systems. vi.Patients with a BIMS score of 7 or less. This is recorded in the MDS as C0500 <= 7. 3.Administer the CoreQ: Short Stay Discharge questionnaire (See S.25) to these individuals. The questionnaire should be administered to patients discharged within 2 weeks of discharge. Provide individuals 2 months to respond to the survey. a.Create a tracking sheet with the following columns: i.Data Administered ii.Data Response Received iii.Time to Receive Response ([Date Response Received – Date Administered]) b.Exclude any surveys where Time to Receive Response >2 Months 4.Collect data over a maximum 6 month time window or until 125 consecutive usable surveys are received (See S.21). 5.Exclude responses not completed by the intended recipient (e.g. questions were answered by a friend or family members. It is important to note that cases in which the residents had help with reading the questions, or writing down their responses, are included in the measure, because in these cases the residents answer the questions themselves). 6.Exclude surveys that are returned after two months 7.Combine the CoreQ: Short Stay Discharge questionnaire items to calculate a patient level score. Responses for each item should be given the following scores: a.Poor = 1, b.Average = 2, c.Good = 3, d.Very good =4 and e.Excellent = 5. 8.Impute missing data if only one of the four questions are missing data by taking the average of the other questions responses. 9.Exclude any survey with 2 or more survey questions that have missing data. 10.Calculated patient score from usable surveys. Patient score= (Score for Item 1 + Score for Item 2 + Score for Item 3 + Score for Item 4) / 4. a.For example, a patient rates their satisfaction on the CoreQ questions as excellent = 5, very good = 4, very good = 4, and good = 3. The resident's total score will be 5 + 4 + 4 + 3 for a total of 16. The patient's total score (16) will then be divided by the number of

	2614 CoreQ: Short Stay Discharge Measure
	questions (4), which equals 4. Thus the patients average satisfaction rating is 4.0. This individual would be counted in the numerator since their average score is >3.0. 11.Flag those patients with an average score equal to or greater than 3.0 12.Calculate the CoreQ: Short Stay Discharge measure which represents the percent of patients with average scores of 3.0 or above. CoreQ: Short Stay Measure= ([number of valid responses with an average score of =3.0] / [total number of valid responses])*100 13.No risk-adjustment is used. 128727| 142371| 151875
Copyright / Disclaimer	None

	2615 CoreQ: Long-Stay Resident Measure
Steward	American Health Care Association
Description	The measure calculates the percentage of long-stay residents, those living in the facility for 100 days or more, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Long-Stay Resident questionnaire that is a three item questionnaire.
Type	Outcome: PRO-PM
Data Source	Instrument-Based Data The collection instrument is the CoreQ: Long-Stay Resident questionnaire and exclusions are from the Resident Assessment Instrument Minimum Data Set (MDS) version 3.0.
Level	Facility
Setting	Post-Acute Care
Numerator Statement	The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long -Stay Resident questionnaire.
Numerator Details	The numerator includes all of the long-stay residents that had an average response =>3 on the CoreQ: Long Stay Resident questionnaire that do not meet any of the exclusions (see exclusions). The calculation of an individual patient's average satisfaction score is done in the following manner: -Respondents within the appropriate time window (see: S.5) and who do not meet the exclusions (See: S.11) are identified. - A numeric score is associated with each response scale option on the CoreQ: Long-Stay Resident questionnaire (that is, Poor=1, Average=2, Good=3, Very Good=4, and Excellent=5). - The following formula is utilized to calculate the individual's average satisfaction score. [Numeric Score Question 1 + Numeric Score Question 2 + Numeric Score Question 3]/3 -The number of respondents whose average satisfaction score >=3 are summed together and function as the numerator. For residents with one missing data point (from the 3 items included in the questionnaire) imputation is used (representing the average value from the other two available questions). Residents with more than one missing data point, are not counted in the measure (i.e., no imputation is used for these residents since their responses are excluded). Imputation details are described in Section S.22. No risk-adjustment is used (see S.13).

	2615 CoreQ: Long-Stay Resident Measure
Denominator Statement	The denominator includes all of the residents that have been in the SNF for 100 days or more regardless of payer status; who received the CoreQ: Long-Stay Resident questionnaire (e.g. people meeting exclusions do not receive the questionnaire), who responded to the questionnaire within the two month time window, who did not have the questionnaire completed by somebody other than the resident, and who did not have more than one item missing.
Denominator Details	The target population includes all current individuals in the SNF on a given day who have been in the SNF for 100 days or more and respond to the CoreQ: Long-Stay Resident questionnaire and completed the survey within the two month time window (See: S.5). Residents have up to 2 months to complete and return the survey. The length-of-stay is identified from nursing facility records (MDS item A1600 "Entry Date").
Exclusions	Exclusions made at the time of sample selection are the following: (1) Residents who have poor cognition defined by the BIMS score; (2) residents receiving hospice; (3) residents with a legal court appointed guardian; and (4) residents who have lived in the SNF for less than 100 days. Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (two months after the administration date) b) surveys that have more than one questionnaire item missing c) surveys from residents who indicate that someone else answered the questions for the resident. (Note this does not include cases where the resident solely had help such as reading the questions or writing down their responses.)
Exclusion details	Individuals are excluded based on information from the Minimum Data Set (MDS) 3.0 assessment. (1) Residents who have poor cognition: Then the Brief Interview for Mental Status (BIMS), a well validated dementia assessment tool is used. BIMS ranges are 0-7 (lowest); 8-12; and 13-15 (highest). Residents with BIMS scores of equal or less than 7 are excluded. (MDS Section C0200-C0500 items are used) (Saliba, et al., 2012). (2) Patients receiving or having received any hospice. This is recorded in the MDS as Hospice O0100K1 = 1 ("the patient was on hospice in the last 14 days while not a resident"), O0100K2 = 1 ("the patient was on hospice in the last 14 days while a resident"), A1800=07 ("entered from hospice"), or A2100=07 ("discharged to hospice"). (3) Patients with court appointed legal guardian for all decisions will be identified from nursing facility health information system. (4) Residents who have lived in the SNF for less than 100 days will be identified from the MDS. This is recorded in the MDS (Section A1600, Entry Date). (5) Residents that respond after the 2 month response period (see S.18, section 3.a on how this is determined). (6) Residents whose responses were completed by someone other than the resident will be excluded. Identified from an additional question on the CoreQ: Long-Stay Resident questionnaire. (7) Residents without usable data (defined as missing data for 2 or 3 of the survey questions). Saliba D, Buchanan J, Edelen MO, Streim J, Ouslander J, Berlowitz D, Chodosh J. J Am Med Dir Assoc. 2012 Sep;13(7):611-7. doi: 10.1016/j.jamda.2012.06.004. Epub 2012 Jul 15.
Risk Adjustment	No risk adjustment or risk stratification
Stratification	No stratification is used.
Type Score	Other (specify): Non-weighted score. Score is a percent. better quality = higher score

	2615 CoreQ: Long-Stay Resident Measure
Algorithm	1.Identify the residents that have been residing in the SNF for 100 days or more. Length of stay so far is the MDS target date (TRGT_DT) - MDS admission date (A1900). 2. Take the residents that have been residing in the SNF for >=100 days and exclude the following: a. Residents who have poor cognition defined as any residents with BIMS scores of 7 or lower. (MDS Section C0200-C0500 used) (Saliba, et al., 2012). b. Patients receiving or having received any hospice. This is recorded in the MDS as Hospice O0100K1 = 1 ("the patient was on hospice in the last 14 days while not a resident"), O0100K2 = 1 ("the patient was on hospice in the last 14 days while a resident"), A1800=07 ("entered from hospice"), or A2100=07 ("discharged to hospice"). c. Residents with Court appointed legal guardian for all decisions will be identified from nursing facility health information system. 3. Administer the CoreQ: Long-stay Resident questionnaire (See S.25) to these individuals. The questionnaire should be administered to all residents in the SNF after exclusions in step 2 above. Communicate that residents have four weeks to respond to the survey. Note, we will include surveys received up to two months from administration but specify four weeks to help increase response rate and completion within a timely manner. This also allows providers to use follow-up strategy at 4 weeks to get responses by the 8 week cut off. 4.Create a tracking sheet with the following columns: i. Data Administered ii. Data Response Received iii. Time to Receive Response ([Date Response Received – Date Administered]) 5.Exclude any surveys received after 2 months from administration. 6.Exclude responses not completed by the intended recipient (e.g. questions were answered by a friend or family members (Note: this does not include cases where the resident solely had help such as reading the questions or writing down their responses). 7.Exclude responses that are missing data for 2 or 3 of the CoreQ questions. 8.All of the remaining surveys are totaled and become the denominator. 9.Combine the CoreQ: Long-Stay Resident questionnaire items to calculate a resident level score. Responses for each item should be given the following scores: a.Poor = 1, b.Average = 2, c.Good = 3, d.Very Good =4 and e.Excellent = 5. 10.Impute missing data if only one of the three questions are missing data. 11.Calculate resident score from usable surveys. a.Patient score= (Score for Item 1 + Score for Item 2 + Score for Item 3) / 3. i.For example, a resident rates their satisfaction on the three CoreQ questions as excellent = 5, very good = 4, and good = 3. The resident's total score will be 5 + 4 + 3 for a total of 12. The resident total score (12) will then be divided by the number of questions (3), which equals 4.0. Thus the residents average satisfaction rating is 4.0. Since the resident's score is >3.0, this resident will be counted in the numerator. b.Flag those patients with a score equal to or greater than 3.0. These residents will be included in the numerator. 12. Calculate the CoreQ: Long-Stay Resident Measure which represents the percent of residents with average scores of 3.0 or above. CoreQ: Long-Stay Resident Measure=

	2615 CoreQ: Long-Stay Resident Measure
	([number of respondents with an average score of =3.0] / [total number of respondents])*100. 13.No risk-adjustment is used. Saliba, D., Buchanan, J., Edelen, M.O., Streim, J., Ouslander, J., Berlowitz, D, & Chodosh J. (2012). MDS 3.0: brief interview for mental status. Journal of the American Medical Directors Association, 13(7): 611-617. 128727| 142371| 151875
Copyright / Disclaimer	None

	2616 CoreQ: Long-Stay Family Measure
Steward	AHCA/NCAL
Description	The measure calculates the percentage of family or designated responsible party for long stay residents (i.e., residents living in the facility for 100 days or more), who are satisfied (see: S.5 for details of the timeframe). This consumer reported outcome measure is based on the CoreQ: Long-Stay Family questionnaire that has three items.
Type	Outcome: PRO-PM
Data Source	Instrument-Based Data The collection instrument is the CoreQ: Long-Stay Family questionnaire and for exclusions the Resident Assessment Instrument Minimum Data Set (MDS) version 3.0 is used
Level	Facility
Setting	Post-Acute Care
Numerator Statement	The numerator assesses the number of family or designated responsible party for long stay residents that are satisfied. Specifically, the numerator is the sum of the family or designated responsible party members for long stay residents that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long-Stay Family questionnaire.
Numerator Details	The numerator includes all of the family or designated responsible party members for long stay residents that had an average response =>3 on the CoreQ: Long-Stay Family questionnaire. We calculate the average satisfaction score for the individual family or designated responsible party member for long stay residents in the following manner: - Respondents within the appropriate time window (see S.5) and who do not meet the exclusions (see S.11) are identified. - A numeric score is associated with each response scale option on the CoreQ: Long-Stay Family questionnaire (that is, Poor=1, Average=2, Good=3, Very Good=4, and Excellent=5). - The following formula is utilized to calculate the individual's average satisfaction score: [Numeric Score Question 1 + Numeric Score Question 2 + Numeric Score Question 3]/3 - The number of respondents whose average satisfaction score >=3 are summed together and function as the numerator. For respondents with one missing data point (from the 3 items included in the questionnaire) imputation will be used (representing the average value from the other two available questions). For respondents with more than one missing data point, they will be excluded from the analyses (i.e., no imputation will be used for these family members). Imputation details are described further below (S.18). No risk-adjustment is used (see S.13).
Denominator Statement	The target population is family or designated responsible party members of a resident residing in a SNF for at least 100 days. The denominator includes all of the individuals in

	2616 CoreQ: Long-Stay Family Measure
	the target population who respond to the CoreQ: Long-Stay Family questionnaire within the two month time window (see S.5) who do not meet the exclusion criteria (see S.10).
Denominator Details	The denominator includes all of the family or the designated responsible party members for residents that have been in the SNF for 100 days or more regardless of payer status; who received the CoreQ: Long-Stay Family questionnaire (e.g. people meeting exclusions do not receive the questionnaire), and who responded to the questionnaire within the two month time window. The length-of-stay (of the resident of the family member or designated responsible party) will be identified from MDS nursing facility records (MDS item A1600 "Entry Date").
Exclusions	Please note, the resident representative for each current resident is initially eligible regardless of their being a family member or not. Only one primary contact per resident should be selected. Exclusions made at the time of sample selection include: (1) family or designated responsible party for residents with hospice; (2) family or designated responsible party for residents with a legal court appointed guardian; (3) representatives of residents who have lived in the SNF for less than 100 days; and (4) representatives who reside in another country. Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (more than two months after the administration date) and b) surveys that have more than one questionnaire item missing.
Exclusion details	Exclusions will be based on information from the Minimum Data Set (MDS) 3.0 assessment. Representatives of residents with the following criteria will be excluded: (1) Residents on hospice. This is recorded in the MDS as Hospice O0100K1 = 1 ("the patient was on hospice in the last 14 days while not a resident"), O0100K2 = 1 ("the patient was on hospice in the last 14 days while a resident"), A1800=07 ("entered from hospice"), or A2100=07 ("discharged to hospice"). (2) Residents with court appointed legal guardian for all decisions will be identified from nursing facility health information system. (3) Residents who have lived in the SNF for less than 100 days will be identified from the MDS. This is recorded in the MDS (item A1600 "Entry Date"). (4) Respondents who reside in another country, to be identified from nursing facility health information system. (5) Respondents who have two or more missing data point are excluded from the analysis. (6) Respondents that respond after the two month response period will be excluded.
Risk Adjustment	No risk adjustment or risk stratification
Stratification	No stratification is used.
Type Score	Other (specify): Non-weighted score. Score is a percent. better quality = higher score
Algorithm	1. Identify the representatives of residents that have been residing in the SNF for 100 days or more. Length of stay so far is the MDS target date (TRGT_DT) - MDS admission date (A1900). 2. Take the representatives of residents that have been residing in the SNF for >=100 days and exclude the following: a. Representatives of residents on hospice. This is recorded in the MDS as Hospice O0100K1 = 1 ("the patient was on hospice in the last 14 days while not a resident"), O0100K2 = 1 ("the patient was on hospice in the last 14 days while a resident"), A1800=07 ("entered from hospice"), or A2100=07 ("discharged to hospice"). b. Residents with Court appointed legal guardian for all decisions as identified from nursing facility health information system. 3. Exclude representatives of residents who reside in another country.

	2616 CoreQ: Long-Stay Family Measure
	4. Administer the CoreQ: Long-Stay Family questionnaire (See S.25) to the representatives that do not meet these exclusion criteria. Provide the family or designated responsible party member for the resident two months to respond to the survey. a. Create a tracking sheet with the following columns: i. Date Administered ii. Date Response Received iii. Time to Receive Response: ([Date Response Received – Date Administered]) b. Exclude any surveys where Time to Receive Response >60 days (2 months) 5.Combine the CoreQ: Long-Stay Family questionnaire items to calculate a resident' representative satisfaction score. Responses for each item should be given the following scores: a.Poor = 1, b.Average = 2, c.Good = 3, d.Very good =4 and e.Excellent = 5. 6.Impute missing data if only one of the three questions are missing data. Drop all survey response if 2 or more survey questions have missing data. 7.Calculate resident's representative score from usable surveys. a.Representative average score = (Score for Item 1 + Score for Item 2 + Score for Item 3) / 3. b.Flag those representatives with a score equal to or greater than 3.0 i.For example, a representative of a resident rates their satisfaction on the three CoreQ questions as excellent = 5, very good = 4, and good = 3. The family member's total score will be 5 + 4 + 3 for a total of 12. The representative of the long-stay resident total score (12) will then be divided by the number of questions (3), which equals 4.0. Thus the representative's average satisfaction rating is 4.0. Since this person's average response is >3.0 they would be counted in the numerator. If it was <3.0 they would not be counted. 8.Calculate the facility's CoreQ: Long-Stay Family Measure which represents the percent of respondents with average scores of 3.0 or above. a.CoreQ: Long-Stay Family Measure = ([number of respondents with an average score of =3.0] / [total number of valid responses])*100 9.No risk-adjustment is used. 128727| 142371| 151875
Copyright / Disclaimer	None

	3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
Steward	Centers for Medicare & Medicaid Services (CMS)
Description	This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.
Type	Outcome: PRO-PM

	3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
Data Source	Claims, Instrument-Based Data The PROM surveys used to define the measure outcome are 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are presently available in English, not yet in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) Global or the Veterans RAND 12 Item Health Survey (VR-12) are used. The PROMIS Global is available in sixteen languages; the VR-12 is available in Spanish, Chinese and German.
Level	Facility
Setting	Inpatient/Hospital
Numerator Statement	The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows: - For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and - For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR). SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group. References: Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.
Numerator Details	This is a patient-reported outcome-based performance measure (PRO-PM). Two joint-specific Patient Reported Outcome Measure (PROM) surveys are used to collect the data for calculating the numerator: 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These PROM data and specific risk variable data will be collected 90 to 0 days prior to surgery, and PROM data will be collected again 270 to 365 days following surgery. Data elements used to define the numerator and for risk adjustment that are collected with PROM data include: - HOOS, JR or KOOS, JR - Date of Birth - Single-Item Literacy Screening (SILS2) Questionnaire - Body Mass Index (BMI) or Weight (kg) and Height (cm) - Chronic (>90 Day) Narcotic Use - Total Painful Joint Count (Patient-Reported in Non-Operative Lower Extremity Joint) - Quantified Spinal Pain (Patient-Reported Back Pain, Oswestry Index Question) - PROMIS Global Mental Health Score (calculated with data from the Patient-Reported Outcomes Measurement Information Systems (PROMIS) Global or Veteran's Rand 12-Item Health Survey (VR-12); data from VR-12 is translated to PROMIS Global Mental Health scores using a crosswalk created by Cella et. al for PROsetta® Stone)

	3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
	(Please note: Data elements listed above are detailed in the Data Dictionary accompanying this NQF submission; see Tabs: Risk Variables with PRO Data; HOOS, JR; KOOS, JR; PROMIS Global; VR-12) Center for Medicare and Medicaid Services (CMS) administrative data is used to identify eligible THA/ TKA procedures for the measure cohort (denominator) and additional risk variables, including patient demographics and clinical comorbidities (ICD-10 codes for eligible THA/TKA procedures identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.) The numerator is the risk-adjusted proportion of patients undergoing an elective primary THA/TKA that meet or exceed a SCB improvement on the HOOS, JR or KOOS, JR from preoperative to postoperative assessment. SCB improvement is defined as: - For THA patients, an increase of 22 points or more on the HOOS, JR - For TKA patients, an increase of 20 points or more on the KOOS, JR SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, TEP, and Technical Advisory Group. Further, the measure numerator was defined with extensive patient and clinician input. Among the numerator definitions considered by stakeholders during measure development included: - Change in PROM score from preoperative to postoperative assessment reported as an average for a hospital's patients; - Postoperative PROM score reported as an average for a hospital's patients; - A threshold change in PROM score from preoperative to postoperative assessment reported as a proportion of a hospital's patients meeting the threshold; - A threshold postoperative PROM score reported as a proportion of a hospital's patients meeting the threshold; and - A combination of both a minimum threshold change in PROM score from preoperative to postoperative assessment and a minimum threshold for postoperative PROM score. Clinical experts and patients supported a numerator definition that assessed change in PROM score from preoperative to postoperative assessment over a numerator definition that focused on postoperative PROM score. TEP members and patients noted that patients want to see improvement, and that the numerator definition should reflect change following surgery. Comments against using a numerator definition focusing on the postoperative PROM score included concern that it does not reflect degree of improvement, and may incentivize surgery on patients with less severe disease who have better preoperative scores. This concern about assessment of the postoperative PROM score also led to dislike of the last option noted above, a numerator definition combining threshold change and threshold postoperative PROM score. Stakeholders also strongly supported a numerator definition assessing a threshold change in PROM score over averaging patient change in PROM scores for hospital reporting. They noted that measurement of a threshold change will highlight lower performing patients, will protect at-risk patients, and is easy to understand as a performance measure. Comments against a reported average change included concern that a hospital whose patients all achieve average results could have a reported average change result that would be very similar to a hospital whose patients achieve either very good or very poor

	3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
	results; an average change numerator could show similar results for hospitals with very different patient outcomes). The numerator definition of SCB threshold change, supported by patients and clinical experts, provides an easy to understand metric that patients found intuitive. Using a SCB threshold avoids the potential for misleading consumers and patients by averaging patient change scores across a hospital when individual patient outcomes within hospitals may vary considerably (as noted above). Using a SCB incentivizes providers to perform surgery on patients with worse baseline scores, a group that might otherwise not be offered surgery, as patients with poorer baseline PRO scores have more room to improve and thus a greater opportunity to achieve SCB. It also encourages providers to not perform THA/TKA procedures on patients with minimal symptoms, who will not benefit at all from surgery. And, since the SCB was defined with close input from patients and clinicians, it does set a minimum improvement threshold, but not one so large as to cause surgeons to avoid performing THA/TKA procedures on patients who would benefit. References: Cella D, Schalet BD, Kallen M, Lai JS, Cook KF, Rutsohn J, Choi SW. PROsetta® Stone Analysis Report Volume 2: A Rosetta Stone for Patient Reported Outcomes, PROMIS Global Health – Mental Component and VR-12 – Mental Component (Algorithmic Scores). http://www.prosettastone.org/LinkingTables1/GlobalHealth/Pages/default.aspx, 2018. Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.
Denominator Statement	The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.
Denominator Details	The cohort for this measure is Medicare FFS patients 65 years of age and older undergoing an elective primary THA/TKA procedure at a non-federal short-term acute care hospital. Inclusion criteria includes patients: - Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission - Discharged alive from a non-federal short-term acute care hospital - Undergoing only elective primary THA/TKA procedures (patients with fractures and revision procedures or with bone metastases are not included) - Inclusion criteria are harmonized with CMS's existing measure cohort for the hospital-level 90- day risk-standardized THA/TKA complication measure Center for Medicare and Medicaid Services (CMS) administrative data is used to identify qualifying THA/TKA procedures for the measure cohort. (ICD-10 codes for eligible THA/TKA procedures are identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.)
Exclusions	Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.
Exclusion details	Patients with staged procedures in the measure period are excluded. A staged procedure is identified if a patient has more than one hospitalization with an eligible, elective primary THA or TKA procedure during the measurement period. ICD-10 codes for eligible, elective primary THA/TKA procedures (listed in the Data Dictionary on "ICD-10 2017-2018" tab) are

	3559 Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)
	used to identify all eligible procedures during the measurement period; patients with an ICD-10 code for an eligible elective primary THA or TKA procedure in two or more hospital admissions during the measurement period are identified as having a staged procedure, and the patient, including all procedures, is removed from the measure cohort.
Risk Adjustment	Statistical risk model
Stratification	N/A
Type Score	Rate/proportion better quality = higher score
Algorithm	Target population: Medicare FFS patients 65 years and older undergoing an elective primary THA or TKA in a non-federal short-term acute care hospital. To create the denominator: Step 1. If the patient is a Medicare FFS patient, go to Step 2. If not, do not include in the denominator. Step 2. If the patient is identified in CMS administrative claims data as having undergone an eligible elective primary THA or TKA during the measurement period, go to Step 3. If not, do not include in the denominator. Step 3. If the patient is 65 years of age or older, go to Step 4. If not, do not include in the denominator. Step 4. If the patient was enrolled in Medicare FFS Part A and Part B for the 12 months prior to index admission, and enrolled in Part A during the index admission, then go to Step 5. If not, do not include in the denominator. Step 5. If the patient was discharged alive from the hospital, include in the denominator. If not, do not include in the denominator. Step 6. If the patient experienced only one elective primary THA/TKA during the measurement period, or if the patient experience more than one elective primary THA/TKA during a singular hospitalization during the measurement period, + in the denominator. If the patient experienced two elective primary THA/TKA procedures during the measurement period performed during distinct hospitalizations, do not include in the denominator. To create the numerator: If the patient has complete PRO data collected during the prescribed preoperative and postoperative time windows, and meets or exceeds the SCB improvement threshold on the joint-specific PROM between the preoperative and postoperative assessment: - for THA patients, an increase of 22 points on the HOOS, JR - for TKA patients, an increase of 20 points on the KOOS, JR then include in the numerator. If not, then do not include in the numerator. The hospital-level measure result is calculated by aggregating all patient-level results across the hospital. For calculation of measure results, we recommend that hospitals should have a minimum case-volume of 25 elective primary THA/TKA patients with complete patient-reported outcomes and risk variable data collected 90 – 0 days preoperatively and complete patient-reported outcomes data collected 270 – 365 days postoperatively. Hospital-specific risk-standardized improvement rates (RSIRs) are calculated as the ratio of a hospital's "predicted" improvement to "expected" improvement multiplied by the overall observed improvement rate. Both predicted improvement and expected improvement are derived based on the output of a hierarchical logistic regression model that adjusts for patient case-mix and applies stabilized inverse probability weighting (IPW) to address potential non-response bias. 146637
Copyright / Disclaimer	None

Appendix E: Related and Competing Measures

Comparison of NQF 3559 and NQF 0422, 0423, 0424, 0425, 0426

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
Steward	Centers for Medicare & Medicaid Services (CMS)	Focus on Therapeutic Outcomes, Inc	Focus on Therapeutic Outcomes, Inc	Focus on Therapeutic Outcomes, Inc	Focus on Therapeutic Outcomes, Inc	Focus on Therapeutic Outcomes, Inc
Description	This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.	A self-report measure of change in functional status for patients 14 year+ with knee impairments. The change in functional status assessed using FOTO's (knee) PROM is adjusted to patient characteristics known to be associated with functional status outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician, and at the clinic level to assess quality.	A self-report measure of change in functional status for patients 14 years+ with hip impairments. The change in functional status assessed using FOTO's (hip) PROM is adjusted to patient characteristics known to be associated with functional status outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician, and at the clinic level to assess quality.	A self-report measure of change in functional status for patients 14 years+ with foot and ankle impairments. The change in functional status assessed using FOTO's (foot and ankle) PROM is adjusted to patient characteristics known to be associated with functional status outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician, and at the clinic level to assess quality.	This is a patient-reported outcome performance measure (PRO-PM) consisting of an item response theory-based patient-reported outcome measure (PROM) of risk-adjusted change in functional status (FS) for patients aged 14 years and older with low back impairments. The change in FS is assessed using the Low Back FS PROM. The measure is adjusted to patient characteristics known to be associated with FS outcomes (risk adjusted) and used as a performance measure at the patient, individual clinician, and clinic levels to assess quality. Scores are reported on a 0 to 100 continuous scale with higher scores indicating better FS. The Low Back FS PROM maps to the Mobility and Self-care constructs within the Activities and Participation domain of the International Classification of Functioning, Disability and Health.	A self-report outcome measure of change in functional status for patients 14 years+ with shoulder impairments. The change in functional status assess using FOTO's (shoulder) PROM is adjusted to patient characteristics known to be associated with functional status outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician, and at the clinic level to assess quality.
Type	Outcome: PRO-PM	Outcome: PRO-PM	Outcome: PRO-PM	Outcome: PRO-PM	Outcome: PRO-PM	Outcome: PRO-PM
Data Source	Claims, Instrument-Based Data The PROM surveys used to define the measure outcome are 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are	Electronic Health Data, Instrument-Based Data, Paper Medical Records Focus On Therapeutic Outcomes, Inc maintains the database. Information on the instrument, risk-adjustment procedures etc. is available at http://www.fotoinc.com/science-of-foto/NQF0425.html Available at measure-specific web page URL identified in S.1 Attachment	Electronic Health Data, Instrument-Based Data, Paper Medical Records Focus On Therapeutic Outcomes, Inc maintains the database. Information on the instrument, risk-adjustment procedures etc. is available at http://www.fotoinc.com/science-of-foto/NQF0425.html Available at measure-specific web page URL identified in S.1 Attachment Hip_data_dictionary_2017Dec.xlsx	Electronic Health Data, Instrument-Based Data, Paper Medical Records Focus On Therapeutic Outcomes, Inc maintains the database. Information on the instrument, risk-adjustment procedures etc. is available at http://www.fotoinc.com/science-of-foto/NQF0425.html Available at measure-specific web page URL identified in S.1 Attachment	Instrument-Based Data The data source is the Focus on Therapeutic Outcomes measurement and reporting system. The instruments are the Low Back FS PROM and risk adjustment questions (as described in the measure Testing Form). A patient completes the FS PROM and respond to risk adjustment questions at the start of an episode of care. The patient again responds to the FS PROM, at	Electronic Health Data, Instrument-Based Data, Paper Medical Records Focus On Therapeutic Outcomes, Inc. maintains the database. Information on the instrument, risk-adjustment procedures etc. is available at http://www.fotoinc.com/science-of-foto/NQF0425.html Available at measure-specific web page URL identified in S.1 Attachment

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	presently available in English, not yet in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) Global or the Veterans RAND 12 Item Health Survey (VR-12) are used. The PROMIS Global is available in sixteen languages; the VR-12 is available in Spanish, Chinese and German. Available in attached appendix at A.1 Attachment Del4-8bHBPNQF3559HipKneePROPMDataDict_For_Submission030520.xlsx	Knee_data_dictionary_2017Dec.xlsx		AnkleFoot_data_dictionary_2017Dec-636501410120533703.xlsx	a minimum, at or near the time of discharge from the episode of care. The Low Back FS PROM may be administered via computer adaptive testing (CAT) or a 10-item short form (static/paper-pencil). CAT administration is preferred as it reduces patient response burden by administrating the minimum number of items needed to achieve the targeted measurement accuracy. The components needed to complete NQF 0425 are publicly available on the FOTO website at no charge. Proxy and Recorder modes of administration are described above in section S.15. Sampling. Available at measure-specific web page URL identified in S.1 Attachment Low_Back_Data_Dictionary_-_RA_Coefficients_NQF2019July-637001594271537751.xlsx	Shoulder_data_dictionary_2017Dec.xlsx
Level	Facility	Facility, Clinician : Group/Practice, Clinician : Individual	Facility, Clinician : Group/Practice, Clinician : Individual	Facility, Clinician : Group/Practice, Clinician : Individual	Clinician : Group/Practice, Clinician : Individual	Facility, Clinician : Group/Practice, Clinician : Individual
Setting	Inpatient/Hospital	Other, Outpatient Services, Post-Acute Care Hospital Outpatient	Other, Outpatient Services, Post-Acute Care Hospital Outpatient	Other, Outpatient Services, Post-Acute Care Hospital Outpatient	Outpatient Services	Home Care, Other, Outpatient Services, Post-Acute Care Hospital Outpatient
Numerator Statement	The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows: - For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and	Patient Level: The residual functional status score for the individual patient (residual scores are the actual change scores - predicted change after risk adjustment. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for knee impairment. Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month time period for knee impairment.	Patient Level: The residual functional status score for the individual patient (residual scores are the actual change scores - predicted change after risk adjustment. Individual Clinician Level: The average residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for hip impairment. Clinic Level: The average residuals in functional status scores in patients who were treated by a clinic in a 12 month time period for hip impairment.	Patient Level: The residual functional status score for the individual patient (residual scores are the actual change scores - predicted change after risk adjustment) Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for foot and or ankle impairment. Clinic Level: The average of residuals in patients who were treated by a clinic in a 12 month time period for foot and or ankle impairment.	The numerator is based on residual scores (actual change scores - predicted change after risk adjustment) of patients receiving care for Low Back impairments and who completed the Low Back PRO-PM. The numerator, as it applies to the 3 levels, is defined as follows: Patient Level: The residual functional status score for the individual patient with a low back impairment. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12-month	Patient Level: The residual functional status score for the individual patient (residual scores are the actual change scores - predicted change after risk adjustment. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for shoulder impairment. Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month time period for shoulder impairment.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	- For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR). SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group. References: Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.				time period for a low back impairment. Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12-month time period for a low back impairment.	
Numerator Details	This is a patient-reported outcome-based performance measure (PRO-PM). Two joint-specific Patient Reported Outcome Measure (PROM) surveys are used to collect the data for calculating the numerator: 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These PROM data and specific risk variable data will be collected 90 to 0 days prior to surgery, and PROM data will be collected again 270 to 365 days following surgery. Data elements used to define the numerator and for risk adjustment that are collected with PROM data include: - HOOS, JR or KOOS, JR - Date of Birth - Single-Item Literacy Screening (SILS2) Questionnaire - Body Mass Index (BMI) or Weight (kg) and Height (cm)	Patient Level: The residual score for the individual patients with knee impairments is derived by applying the statistical risk adjustment model described in S.14 and S.15 and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of this application. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for knee impairment. Average scores are calculated for all clinicians, however performance is evaluated only for those clinicians that had a minimum of 10 patients in the previous 12 months. To maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinician regardless of clinic size, but has recently changed its procedure to enable participation by clinicians that do	Patient Level: The residual score for the individual patients with hip impairments is derived by applying the statistical risk adjustment model described in S.14 and S.15 and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of this application. Individual Clinician Level: The average residuals in scores in patients who were treated by a clinician in a 12 month time period for hip impairment. Average scores are calculated for all clinicians, however, performance is evaluated only for those clinicians that had a minimum of 10 patients in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinician regardless of clinic size, but has recently changed its procedure to enable participation by clinicians that do not have a sufficient volume of patients. The score is	Patient Level: The residual score for the individual patients with foot and ankle impairments is derived by applying the statistical risk adjustment model described in S.14 and S.15 and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of this application. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for foot and ankle impairment. Average scores are calculated for all clinicians, however only those clinicians that had a minimum of 10 patients in the previous 12 months. are included in the comparative benchmarked report. In 2011-2013, FOTO used a standard threshold of 40 patients/clinician, but has recently changed its procedure to enable participation by clinicians that do not have a sufficient volume of patients. The	Patient Level: The residual score for the individual patients with low back impairments is derived by applying the statistical risk adjustment model described in S.10 and applying steps 1-5 as described in S.14. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12-month time period for low back impairment. Average scores are calculated for all clinicians, but performance is evaluated only for those clinicians that had a minimum of 10 patients in the previous 12 months to maximize stability of the benchmarking estimates. The score is derived by applying steps 1-6 as described in S.14. Clinic Level: The average of residuals in functional status scores in patients who were treated within a clinic in a 12-month time period for lumbar impairments. Average scores are calculated for all clinics, but performance is evaluated only for	Patient Level: The residual score for the individual patients with shoulder impairments is derived by applying the statistical risk adjustment model described in S.14 and S.15 and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of this application. Individual Clinician Level: The average residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for shoulder impairment. Average scores are calculated for all clinicians, however performance is evaluated only for those clinicians that had a minimum of 10 patients in the previous 12 months. To maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinician regardless of clinic size, but has recently changed its procedure to enable participation by clinicians that do

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	- Chronic (>90 Day) Narcotic Use - Total Painful Joint Count (Patient-Reported in Non-Operative Lower Extremity Joint) - Quantified Spinal Pain (Patient-Reported Back Pain, Oswestry Index Question) - PROMIS Global Mental Health Score (calculated with data from the Patient-Reported Outcomes Measurement Information Systems (PROMIS) Global or Veteran's Rand 12-Item Health Survey (VR-12); data from VR-12 is translated to PROMIS Global Mental Health scores using a crosswalk created by Cella et. al for PROsetta® Stone) (Please note: Data elements listed above are detailed in the Data Dictionary accompanying this NQF submission; see Tabs: Risk Variables with PRO Data; HOOS, JR; KOOS, JR; PROMIS Global; VR-12) Center for Medicare and Medicaid Services (CMS) administrative data is used to identify eligible THA/ TKA procedures for the measure cohort (denominator) and additional risk variables, including patient demographics and clinical comorbidities (ICD-10 codes for eligible THA/TKA procedures identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.) The numerator is the risk-adjusted proportion of patients undergoing an elective primary THA/TKA that meet or exceed a SCB improvement on the HOOS, JR or	not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Clinic Level: The average of residuals in functional status scores in patients who were treated within a clinic in a 12 month time period for knee impairment. Average scores are calculated for all clinics, however performance is evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinics regardless of clinic size, but has recently changed its procedure to enable participation by smaller clinics that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Both comparative benchmark reports (i.e., scale ranging from 0 to 100 with higher scores meaning higher functional abilities) at the clinician or clinic level include patients with knee impairments, who were treated in therapy, had their functional status assessed at admission and at the end of their episode of therapy and were discharged from therapy.	derived by applying steps 1-6 as described in S.18 Clinic Level: The average residuals functional status scores in patients who were treated within a clinic in a 12 month time period for hip impairment. Average scores are calculated for all clinics, however performance is evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinics regardless of clinic size, but has recently changed its procedure to enable participation by smaller clinics that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Both comparative benchmark reports (i.e., scale ranging from 0 to 100 with higher scores meaning higher functional abilities) at the clinician or clinic level include patients with hip impairments, who were treated in therapy and had their functional status assessed at admission and at the end of their episode of therapy and were discharged from therapy.	score is derived by applying steps 1-6 as described in S.18 Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month time period for foot and ankle impairment. Average scores are calculated for all clinics, however performance is evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinics regardless of clinic size, but has recently changed its procedure to enable participation by smaller clinics that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Both comparative benchmark reports (i.e., scale ranging from 0 to 100 with higher scores meaning higher functional abilities) at the clinician or clinic level include patients with foot and ankle impairments, who were treated in therapy and had their functional status assessed at admission and at the end of their episode of therapy and were discharged from therapy.	large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. The score is derived by applying steps 1-6 as described in S.14. Items and response options are provided in the attachment in section S.2c. above.	not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18. Clinic Level: The average of residuals in functional status scores in patients who were treated within a clinic in a 12 month time period for shoulder impairment. Average scores are calculated for all clinics, however performance is evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinics regardless of clinic size, but has recently changed its procedure to enable participation by smaller clinics that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18. Both comparative benchmark reports (i.e., scale ranging from 0 to 100 with higher scores meaning higher functional abilities) at the clinician or clinic level include patients with shoulder impairments, who were treated in therapy and had their functional status assessed at admission and at the end of their episode of therapy and were discharged from therapy.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	KOOS, JR from preoperative to postoperative assessment. SCB improvement is defined as: - For THA patients, an increase of 22 points or more on the HOOS, JR - For TKA patients, an increase of 20 points or more on the KOOS, JR SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, TEP, and Technical Advisory Group. Further, the measure numerator was defined with extensive patient and clinician input. Among the numerator definitions considered by stakeholders during measure development included: - Change in PROM score from preoperative to postoperative assessment reported as an average for a hospital's patients; - Postoperative PROM score reported as an average for a hospital's patients; - A threshold change in PROM score from preoperative to postoperative assessment reported as a proportion of a hospital's patients meeting the threshold; - A threshold postoperative PROM score reported as a proportion of a hospital's patients meeting the threshold; and - A combination of both a minimum threshold change in PROM score from preoperative to postoperative assessment and a minimum threshold for postoperative PROM score. Clinical experts and patients supported a numerator definition that assessed change in PROM score from preoperative to postoperative assessment over a					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	numerator definition that focused on postoperative PROM score. TEP members and patients noted that patients want to see improvement, and that the numerator definition should reflect change following surgery. Comments against using a numerator definition focusing on the postoperative PROM score included concern that it does not reflect degree of improvement, and may incentivize surgery on patients with less severe disease who have better preoperative scores. This concern about assessment of the postoperative PROM score also led to dislike of the last option noted above, a numerator definition combining threshold change and threshold postoperative PROM score. Stakeholders also strongly supported a numerator definition assessing a threshold change in PROM score over averaging patient change in PROM scores for hospital reporting. They noted that measurement of a threshold change will highlight lower performing patients, will protect at-risk patients, and is easy to understand as a performance measure. Comments against a reported average change included concern that a hospital whose patients all achieve average results could have a reported average change result that would be very similar to a hospital whose patients achieve either very good or very poor results; an average change numerator could show similar results for hospitals with very different patient outcomes). The numerator definition of SCB threshold change, supported by patients and clinical experts, provides an easy to understand					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	metric that patients found intuitive. Using a SCB threshold avoids the potential for misleading consumers and patients by averaging patient change scores across a hospital when individual patient outcomes within hospitals may vary considerably (as noted above). Using a SCB incentivizes providers to perform surgery on patients with worse baseline scores, a group that might otherwise not be offered surgery, as patients with poorer baseline PRO scores have more room to improve and thus a greater opportunity to achieve SCB. It also encourages providers to not perform THA/TKA procedures on patients with minimal symptoms, who will not benefit at all from surgery. And, since the SCB was defined with close input from patients and clinicians, it does set a minimum improvement threshold, but not one so large as to cause surgeons to avoid performing THA/TKA procedures on patients who would benefit. References: Cella D, Schalet BD, Kallen M, Lai JS, Cook KF, Rutsohn J, Choi SW. PROsetta® Stone Analysis Report Volume 2: A Rosetta Stone for Patient Reported Outcomes, PROMIS Global Health – Mental Component and VR-12 – Mental Component (Algorithmic Scores). http://www.prosettastone.org/LinkingTables1/GlobalHealth/Pages/default.aspx, 2018. Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
Denominator Statement	The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.	All patients 14 years and older with knee impairments who have initiated rehabilitation treatment and completed the FOTO knee FS PROM at admission and discharge.	All patients 14 years and older with hip impairments who have initiated rehabilitation treatment and complete the FOTO hip FS PROM at admission and discharge.	All patients 14 years and older with foot or ankle impairments who have initiated rehabilitation treatment and completed the FOTO foot and ankle PROM at admission and discharge	The target population is all patients 14 years and older with a Low Back impairment who have initiated an episode of care and completed the Low Back FS PROM.	All patients 14 years and older with shoulder impairments who have initiated rehabilitation treatment and completed the FOTO shoulder FS outcome instrument at admission and discharge.
Denominator Details	The cohort for this measure is Medicare FFS patients 65 years of age and older undergoing an elective primary THA/TKA procedure at a non-federal short-term acute care hospital. Inclusion criteria includes patients: - Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission - Discharged alive from a non-federal short-term acute care hospital - Undergoing only elective primary THA/TKA procedures (patients with fractures and revision procedures or with bone metastases are not included) - Inclusion criteria are harmonized with CMS's existing measure cohort for the hospital-level 90-day risk-standardized THA/TKA complication measure Center for Medicare and Medicaid Services (CMS) administrative data is used to identify qualifying THA/TKA procedures for the measure cohort. (ICD-10 codes for eligible THA/TKA procedures are identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.)	The established ICD-9-CM codes for the knee include: Diagnoses specific to the knee: 682.6, 711.06, 711.16, 711.26, 711.36, 711.46, 711.56, 711.76, 711.86, 711.96, 712.16, 712.26, 712.36, 712.86, 715.16, 715.26, 715.36, 715.86, 715.96, 716.06, 716.16, 716.26, 716.36, 716.46, 716.56, 716.66, 716.86, 716.96, *717, 718.26, 718.36, 718.46, 718.56, 718.76, 718.86, 719.06, 719.16, 719.26, 719.36, 719.46, 719.56, 719.66, 719.86, 719.96, *726.6, 726.90, 727.51, 727.65, 727.66, 729.31, 730.06, 730.16, 730.26, 730.36, 730.76, 730.86, 732.4, 736.4, 736.5, 736.6, 739.6, 755.64, *822, *836, *844, 928.10, 924.11, 959.7, V43.65 * Use of an asterisk is to include all codes in the category Or Diagnoses not specific to the knee, but affect the function of the knee: 337.22, 355.2, 355.3, 355.4, 355.71, 355.79, 355.8, 355.9, 710.4, 710.8, 710.9, 711.09, 711.19, 711.29, 711.39, 711.59, 711.69, 711.79, 711.89, 711.99, 712.29, 712.39, 712.89, 714.0, 714.30, 714.4, 714.89, 714.9, 715.09, 715.18, 715.28, 715.38, 715.89, 715.98, 716.09, 716.19, 719.29, 716.39, 716.49, 716.59, 719.89, 716.99, 718.29, 718.39, 718.49, 718.59, 718.89, 719.09, 719.19, 719.29, 719.39, 719.49, 719.59, 719.69, 719.7, 719.89, 719.99, 726.90, 727.00, 727.02,	The established ICD-9-CM codes for the hip include: Diagnoses specific to the hip: 715.05, 715.15, 715.25, 715.35, 715.95, 716.05, 716.15, 716.25, 716.35, 716.45, 716.55, 716.65, 716.85, 716.95, 718.05, 718.15, 718.25, 718.35, 718.45, 718.55, 718.65, 718.75, 718.85, 718.95, 719.05, 719.15, 719.25, 719.35, 719.45, 719.55, 719.65, 719.75, 719.85, 719.95, 726.5, 730.05, 730.15, 730.25, 730.35, 730.75, 730.85, 730.95, 733.98, *736.3, 738.6, 739.4, 739.5, *754.3, 755.63, *808, *821, *835, *843, *846, 847.3, 847.4, 848.5, *924.0, *928.0, V54.13, V54.23, V57.1. * Use of an asterisk is to include all codes in the category Or Diagnoses not specific to the hip, but affect the function of the hip: *355, 710.05, 710.3, 710.4, 710.8710.9, 711.99, 714.31, 714.4, 715.09, 715.80, 715.89, 715.98, 716.09, 716.19, 719.29, 716.39, 716.49, 716.59, 716.89, 716.99, 718.09, 718.19, 718.29, 718.39, 718.49, 718.59, 718.79, 718.89, 718.99, 719.09, 719.19, 719.29, 719.39, 719.49, 719.59, 719.69, 719.89, 719.99, 725, 726.8, 726.90, 727.09, 727.3, *724.4, 727.50, 728.3, 718.4, 728.0, *728.1, 728.2, 728.5, *728.8, 728.9, 729.0, 729.1, 729.2, 729.5, *729.8, *729.9, 730.09, 730.19, 730.29, 730.39, 730.79, 730.89, 730.99, 732.1, 732.2, 732.9, *733.0, 733.15,	The established ICD-9-CM codes for the the FOTO foot and ankle measure are: soft tissue disorders of muscle, synovium, tendon, bursa, plantar fasciitis, or enthesopathies (ICD-9 codes 725-729); sprains and strains of the ankle or foot (ICD-9 codes 844-845 including unspecified sprain or strain); fractures (ICD-9 823-826 including ankle, tarsal, metatarsal bones, or phalanges of foot); arthropathies (ICD-9 codes 710-719, including osteoarthoses, rheumatoid arthritis); disorders of the bone and cartilage (ICD-9 codes 730-739); uncomplicated post-surgical (CPT codes 29894-29899, including arthroscopy of the ankle); and gait abnormality (ICD-9 code 781.2). The ICD 10 Crosswalk is provided on the measure specific webpage provided in S.1.	The ICD-10 codes relevant for this measure are: G54.1, G54.4, G57.0, M43.06, M43.07, M43.08, M43.16, M43.17, M43.18, M43.26, M43.27, M43.28, M43.5X6 , M43.5X7, M43.5X8, M43.8X6, M43.8X7, M43.8X8, M45.6, M45.7, M45.8 M46.1, M46.46, M46.47, M46.48, M47.16, M47.26, M47.27, M47.28, M47.816, M47.817, M47.896, M47.897, M47.898, M48.06, M48.07, M51.06, M51.16, M51.17, M51.26, M51.27, M51.36, M51.37, M51.46, M51.47, M51.86, M51.87, M51.9, M53.2X6, M53.2X7, M53.2X8, M53.88, M54.16, M54.17, M54.18, M54.3, M54.4, M54.5, M99.73, S32.0, S32.1, S32.2, S33.0, S33.1, S33.2, S33.3, S33.5, S33.10, S33.11, S33.12, S33.13, S39.002, S39.012	The established ICD-9-CM codes for the shoulder include: soft tissue disorders of muscle, synovium, tendon, bursa, or enthesopathies (ICD-9 codes 725-729); sprains and strains of the shoulder (ICD-9 code 840 including unspecified sprain or strain); fractures (ICD-9 codes 810-819 including clavicle, scapula, humerus); arthropathies (ICD-9 codes 710-719, including osteoarthoses, rheumatoid arthritis); disorders of the bone and cartilage (ICD-9 codes 730-739); dislocations of shoulder (ICD-9 codes 831); post-surgical (CPT codes including 23107 arthrotomy, 23405 tenotomy). The ICD 10 Crosswalk is provided on the measure specific webpage provided in S.1.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
		727.09, 727.2, 727.3, 727.40, *727.8, 727.9, 728.2, 728.3, 728.4, 728.5, 728.87, 728.89, 728.9, 729.0, 729.1, 729.2, 729.4, 729.5, 729.81, 729.89, 729.90, 730.09, 730.19, 730.29, 730.39, 730.79, 730.89, 732.9, *733.0, 733.49, 736.81, 780.79, 781.2, 781.3, 827.0, 827.1, 848.8, *891, 897.0, 897.1, 924.4, 996.77, 996.78, V49.75, V54.81, V57.1, V57.81, V57.89, V58.78 * Use of an asterisk is to include all codes in the category The ICD 10 Crosswalk is provided on the measure specific webpage provided in S.1.	733.19, 733.29, 733.40, 733.42, *733.8, 733.90, 733.91, 733.96, 733.97, 780.79, 780.96, 781.2, 792.3, *827, 848.8, 848.9, *897, 922.31, 922.32, 928.8, *956, 959.6, *996.4, 996.70, 996.77, V54.81, V54.89, V58.78 * Use of an asterisk is to include all codes in the category The ICD 10 Crosswalk is provided on the measure specific webpage provided in S.1.			
Exclusions	Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.	•Patients who are not being treated for a Knee impairment •<14 years of age	•Patients who are not being treated for a Hip impairment •<14 years of age	•Patients who are not being treated for a foot and ankle impairment •<14 years of age	Patients who are not being treated for a Low Back impairment. Patients who are less than 14 years of age.	•Patients who are not being treated for a Shoulder impairment •<14 years of age
Exclusion Details	Patients with staged procedures in the measure period are excluded. A staged procedure is identified if a patient has more than one hospitalization with an eligible, elective primary THA or TKA procedure during the measurement period. ICD-10 codes for eligible, elective primary THA/TKA procedures (listed in the Data Dictionary on "ICD-10 2017-2018" tab) are used to identify all eligible procedures during the measurement period; patients with an ICD-10 code for an eligible elective primary THA or TKA procedure in two or more hospital admissions during the measurement period are identified as having a staged procedure, and the patient, including all	• Patients who are not being treated for a knee impairment • Age under 14 years old.	• Patients who are not being treated for a hip impairment • <14 years of age	• Patients who are not being treated for a foot and ankle impairment conditions • Age under 14 years old.	NA	• Patients who are not being treated for a shoulder impairment • Age under 14 years old.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	procedures, is removed from the measure cohort.					
Risk Adjustment	Statistical risk model	Statistical risk model	Statistical risk model	Statistical risk model	Statistical risk model	Statistical risk model
Stratification	N/A	Risk adjusted - not stratified	Risk adjusted - not stratified	Risk adjusted - not stratified	This measure is risk-adjusted, not risk-stratified. The methods used to develop the FOTO risk-adjustment Low Back model were the same as the methods described in detail in a recent publication by Deutscher et at, 2018 [Deutscher, D., Werneke, M. W., Hayes, D., Mioduski, J. E., Cook, K. F., Fritz, J. M., et al. (2018). Impact of Risk Adjustment on Provider Ranking for Patients with Low Back Pain Receiving Physical Therapy. J Orthop Sports Phys Ther, 48(8), 637-648] Briefly, we used data from adult patients with Low Back pain treated in outpatient rehabilitation clinics during 2014-2016, that had complete outcomes data at admission and discharge, to develop the risk-adjustment model. The data included the following patient factors that could be evaluated for inclusion in a model for risk-adjustment: FS at admission (continuous); age (continuous); sex (male/female); acuity as number of days from onset of the treated condition (6 categories); type of payer (10 categories); number of related surgeries (4 categories); exercise history (3 categories); use of medication at intake for the treatment of LBP (yes/no); previous treatment for LBP (yes/no); treatment post-surgery (low back fusion, laminectomy or other); and 31 comorbidities. For further details, please see Measure Testing Form section 2b3. Risk Adjustment/Stratification for	Risk adjusted - not stratified

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
					Outcome or Resource Use Measures. The model variables and coefficients are contained in the document attached above in section S.2b. Data Dictionary, Code Table, or Value Sets.	
Type Score	Rate/proportion better quality = higher score	Continuous variable, e.g. average better quality = higher score	Continuous variable, e.g. average better quality = higher score	Continuous variable, e.g. average better quality = higher score	Continuous variable, e.g. average better quality = higher score	Continuous variable, e.g. average better quality = higher score
Algorithm	Target population: Medicare FFS patients 65 years and older undergoing an elective primary THA or TKA in a non-federal short-term acute care hospital. To create the denominator: Step 1. If the patient is a Medicare FFS patient, go to Step 2. If not, do not include in the denominator. Step 2. If the patient is identified in CMS administrative claims data as having undergone an eligible elective primary THA or TKA during the measurement period, go to Step 3. If not, do not include in the denominator. Step 3. If the patient is 65 years of age or older, go to Step 4. If not, do not include in the denominator. Step 4. If the patient was enrolled in Medicare FFS Part A and Part B for the 12 months prior to index admission, and enrolled in Part A during the index admission, then go to Step 5. If not, do not include in the denominator. Step 5. If the patient was discharged alive from the hospital, include in the denominator. If not, do not include in the denominator. Step 6. If the patient experienced only one elective primary THA/TKA during the measurement period, or if the patient experience more than one elective primary THA/TKA during a singular hospitalization during the measurement period, + in the denominator. If the patient experienced two elective primary	STEPS TAKEN TO PRODUCE THIS MEASURE: Definitions: Patient's Functional Status Score. A functional status score is produced when the patient completes the FOTO (knee) PROM administered by internet or paper and pencil survey. The functional status score is continuous and linear. Scores range from 0 (low function) to 100 (high function). The survey is standardized, and the scores are validated for the measurement of function for this population. Patient's Functional Status Change Score. A functional status change score is calculated by subtracting the Patient's Functional Status Score at Admission from the Patient's Functional Status Score at Discharge. Predicted Functional Status Change Score. Functional Status Change Scores for patients are risk adjusted using multiple linear regression methods that accounted for the following independent variables: Patient's Functional Status Score at Admission, patient age, gender, symptom acuity, surgical history, specific co morbidities, payer type, use of medication for the knee impairment at Intake, previous treatment for the condition, exercise history, and post-surgical category if applicable. The Patient's Functional Status Change Score is the dependent variable.	STEPS TAKEN TO PRODUCE THIS MEASURE: Definitions: Patient's Functional Status Score. A functional status score is produced when the patient completes the FOTO (hip) PROM administered by internet or paper and pencil. The functional status score is continuous and linear. Scores range from 0 (low function) to 100 (high function). The survey is standardized, and the scores are validated for the measurement of function for this population. Patient's Functional Status Change Score. A functional status change score is calculated by subtracting the Patient's Functional Status Score at Admission from the Patient's Functional Status Score at Discharge. Predicted Functional Status Change Score. Functional Status Change Scores for patients are risk adjusted using multiple linear regression methods that accounted for the following independent variables: Patient's Functional Status Score at Admission, patient age, symptom acuity, surgical history, gender, specific co morbidities, payer type, use of medication for the hip impairment at Intake, previous treatment for the hip impairment, exercise history, and post-surgical category if applicable.	STEPS TAKEN TO PRODUCE THIS MEASURE: Definitions: Patient's Functional Status Score. A functional status score is produced when the patient completes the FOTO PROM administered by internet or paper and pencil. The functional status score is continuous and linear. Scores range from 0 (low function) to 100 (high function). The survey is standardized, and the scores are validated for the measurement of function for this population. Patient's Functional Status Change Score. A functional status change score is calculated by subtracting the Patient's Functional Status Score at Admission from the Patient's Functional Status Score at Discharge. Predicted Functional Status Change Score. Functional Status Change Scores for patients are risk adjusted using multiple linear regression methods that accounted for the following independent variables: Patient's Functional Status Score at Admission, patient age, symptom acuity, surgical history, gender, specific co morbidities, payer type, use of medication for the ankle/foot impairment at Intake, previous treatment for the condition, exercise history, and post-surgical category if applicable. The Patient's Functional Status Change Score is the dependent	DEFINITIONS: Patient's Functional Status Score. A Functional Status (FS) Score is produced when the patient completes the FOTO Low Back FS PROM administered via computer adaptive testing or short form. Patient's FS Change Score. An FS Change Score is calculated by subtracting the Patient's FS Score at the Initial Evaluation (i.e., the start of the care episode) from the Patient's FS Score at Discharge (i.e., the end of the care episode). Predicted FS Change Score. FS Change Scores for patients are risk adjusted with a model developed using multiple linear regression methods that account for the following independent variables: Patient's FS Score at Initial Evaluation, patient age, symptom acuity, surgical history, gender, specific co-morbidities, payer type, use of medication for the low back impairment at Initial Evaluation, previous treatment for the low back impairment, exercise history, and post-surgical category if applicable. The Patient's FS Change Score is the dependent variable. The statistical regression method provides a set of coefficients that accounts ("adjusts") for the association of each variable with the FS outcome as it applies to each patient, resulting in a risk-adjusted Predicted FS Change Score.	STEPS TAKEN TO PRODUCE THIS MEASURE: Definitions: Patient's Functional Status Score. A functional status score is produced when the patient completes the FOTO (shoulder) PROM administered by internet or paper and pencil. The functional status score is continuous and linear. Scores range from 0 (low function) to 100 (high function). The survey is standardized, and the scores are validated for the measurement of function for this population. Patient's Functional Status Change Score. A functional status change score is calculated by subtracting the Patient's Functional Status Score at Admission from the Patient's Functional Status Score at Discharge. Predicted Functional Status Change Score. Functional Status Change Scores for patients are risk adjusted using multiple linear regression methods that accounted for the following independent variables: Patient's Functional Status Score at Admission, patient age, symptom acuity, surgical history, gender, specific co morbidities, payer type, use of medication for the shoulder impairment at Intake, previous treatment for the shoulder condition, exercise history, and post-surgical category if applicable.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	THA/TKA procedures during the measurement period performed during distinct hospitalizations, do not include in the denominator. To create the numerator: If the patient has complete PRO data collected during the prescribed preoperative and postoperative time windows, and meets or exceeds the SCB improvement threshold on the joint-specific PROM between the preoperative and postoperative assessment: - for THA patients, an increase of 22 points on the HOOS, JR - for TKA patients, an increase of 20 points on the KOOS, JR then include in the numerator. If not, then do not include in the numerator. The hospital-level measure result is calculated by aggregating all patient-level results across the hospital. For calculation of measure results, we recommend that hospitals should have a minimum case-volume of 25 elective primary THA/TKA patients with complete patient-reported outcomes and risk variable data collected 90 – 0 days preoperatively and complete patient-reported outcomes data collected 270 – 365 days postoperatively. Hospital-specific risk-standardized improvement rates (RSIRs) are calculated as the ratio of a hospital's "predicted" improvement to "expected" improvement multiplied by the overall observed improvement rate. Both predicted improvement and expected improvement are derived based on the output of a hierarchical logistic regression model that adjusts for patient case-mix and applies stabilized	The statistical regression produces a Risk-Adjusted Predicted Functional Status Change Score. Risk-adjusted Functional Status Change Residual Score. The difference between the actual change and the predicted change scores (after risk adjustment) is the residual score and should be interpreted as the unit of functional status change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted residual change scores of zero (0) or greater (>0) should be interpreted as functional status change scores that were predicted or better than predicted given the risk-adjustment variables of the patient, and risk-adjusted residual change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated risk-adjusted residual scores: The average of residual scores of functional status (actual change - predicted change after risk adjustment) from a provider (clinician or clinic). The aggregated scores are used to make comparisons between clinicians or clinics. STEPS: First, the patient completes FOTO's functional status survey for the Knee at Admission, which generates the Patient's Functional Status Score at Admission	The Patient's Functional Status Change Score is the dependent variable. The statistical regression produces a Risk-Adjusted Predicted Functional Status Change Score. Risk-adjusted Functional Status Change Residual Score. The difference between the actual change and the predicted change scores (after risk adjustment) is the residual score and should be interpreted as the unit of functional status change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted residual change scores of zero (0) or greater (>0) should be interpreted as functional status change scores that were predicted or better than predicted, respectively, given the risk-adjustment variables of the patient. Risk-adjusted residual change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated risk-adjusted residual scores: The average of residual scores of functional status (actual change - predicted change after risk adjustment) from a provider (clinician or clinic). The aggregated scores are used to make comparisons between clinicians or clinics. STEPS: First, the patient completes FOTO's functional status survey for the Hip at Admission, which generates the	variable. The statistical regression produces a Risk-Adjusted Predicted Functional Status Change Score. Risk-adjusted Functional Status Change Residual Score. The difference between the actual change and the predicted change scores (after risk adjustment) is the residual score and should be interpreted as the unit of functional status change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted residual change scores of zero (0) or greater (>0) should be interpreted as functional status change scores that were predicted or better than predicted given the risk-adjustment variables of the patient. Risk-adjusted residual change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated risk-adjusted residual scores: The average of residual scores of functional status (actual change - predicted change after risk adjustment) from a provider (clinician or clinic). The aggregated scores are used to make comparisons between clinicians or clinics. STEPS: Patient, level measures use steps 1-5 Clinician and clinic level measures use steps 1-6. 1) the patient completes FOTO's functional status survey for the foot and ankle impairment at	Residual Score: The Residual Score is calculated as the difference between the actual change and risk-adjusted predicted change scores and should be interpreted as the unit of FS change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted Residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted Residual change scores of zero (0) or greater (>0) should be interpreted as functional status change scores that were predicted or better than predicted given the risk-adjustment variables of the patient. Risk-adjusted residual change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated Residual Scores: The average of Residual scores of FS (actual change - predicted change after risk adjustment) from a provider (clinician or clinic). The aggregated scores are used to make comparisons between clinicians or clinics. STEPS TO CALCULATE THE PRO-PM SCORE, APPLYING THE ABOVE DEFINITIONS: Patient level measures use steps 1-5. Clinician and clinic level measures use steps 1-6. 1) The patient is identified as age 14 or older and presenting for an episode of care for a low back impairment and completing the FOTO Low Back FS PROM which generates the Patient's FS Score at Initial Evaluation.	The Patient's Functional Status Change Score is the dependent variable. The statistical regression produces a Risk-Adjusted Predicted Functional Status Change Score. Risk-adjusted Functional Status Change Residual Score. The difference between the actual change and the predicted change scores (after risk adjustment) is the residual score and should be interpreted as the unit of functional status change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted residual change scores of zero (0) or greater (>0) should be interpreted as functional status change scores that were predicted or better than predicted given the risk-adjustment variables of the patient, and risk-adjusted residual change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated risk-adjusted residual scores: The average of residual scores of functional status (actual change - predicted change after risk adjustment) from a provider (clinician or clinic). The aggregated scores are used to make comparisons between clinicians or clinics. STEPS: First, the patient completes FOTO's functional status survey for the Shoulder at Admission, which generates the Patient's Functional Status Score at Admission

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	inverse probability weighting (IPW) to address potential non-response bias.	Second, patient completes FOTO's functional status survey at or near Discharge, which generates the Patient's Functional Status Score at Discharge Third, the Patient's Functional Status Change Score (raw, non-risk-adjusted) is generated Fourth, a Risk-adjusted Predicted Functional Status Change Score is generated using a regression equation Fifth, a Risk-adjusted Functional Status Change Residual Score is generated for each patient. Sixth, the average residual scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (5 or more clinicians) in order to obtain stable estimates of provider performance.	Patient's Functional Status Score at Admission Second, patient completes FOTO's functional status survey at or near Discharge, which generates the Patient's Functional Status Score at Discharge Third, the Patient's Functional Status Change Score (raw, non-risk-adjusted) is generated Fourth, a Risk-adjusted Predicted Functional Status Change Score is generated using a regression equation Fifth, a Risk-adjusted Functional Status Change Residual Score is generated for each patient. Sixth, the average residual scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (5 or more clinicians) in order to obtain stable estimates of provider performance.	Admission, which generates the Patient's Functional Status Score at Admission 2) the patient completes FOTO's functional status survey at or near Discharge, which generates the Patient's Functional Status Score at Discharge 3) the Patient's Functional Status Change Score (raw, non-risk-adjusted) is generated 4) a Risk-adjusted Predicted Functional Status Change Score is generated using a regression equation 5) a Risk-adjusted Functional Status Change Residual Score is generated for each patient. 6.) the average residual scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (5 or more clinicians) in order to obtain stable estimates of provider performance.	2) The patient completes the FOTO Low Back FS PROM at or near Discharge, which generates the Patient's FS Score at Discharge. 3) The Patient's FS Change Score (raw, non-risk-adjusted) is generated. 4) A Predicted FS Change Score is generated for the patient using the risk-adjustment model. 5) A Residual Score is generated for the patient. 6) The average Residual Scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (5 or more clinicians) in order to obtain stable estimates of provider performance.	Second, patient completes FOTO's functional status survey at or near Discharge, which generates the Patient's Functional Status Score at Discharge Third, the Patient's Functional Status Change Score (raw, non-risk-adjusted) is generated Fourth, a Risk-adjusted Predicted Functional Status Change Score is generated using a regression equation Fifth, a Risk-adjusted Functional Status Change Residual Score is generated for each patient. Sixth, the average residual scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (5 or more clinicians) in order to obtain stable estimates of provider performance.
Submission items	5.1 Identified measures: 1550 : Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)	5.1 Identified measures: 5a.1 Are specs completely harmonized?	5.1 Identified measures: 5a.1 Are specs completely harmonized? No	5.1 Identified measures: 5a.1 Are specs completely harmonized?	5.1 Identified measures: 5a.1 Are specs completely harmonized? No	5.1 Identified measures: 5a.1 Are specs completely harmonized? No

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	1551 : Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) 2958 : Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery 0422 : Functional status change for patients with Knee impairments 0423 : Functional status change for patients with Hip impairments 0424 : Functional status change for patients with Foot and Ankle impairments 0425 : Functional Status Change for Patients with Low Back Impairments 0426 : Functional status change for patients with Shoulder impairments 0427 : Functional status change for patients with elbow, wrist and hand impairments 0428 : Functional status change for patients with General orthopaedic impairments 2643 : Average change in functional status following lumbar spine fusion surgery 5a.1 Are specs completely harmonized? Yes 5a.2 If not completely harmonized, identify difference, rationale, impact: To the extent feasible, we have harmonized with existing, related measures. However, we have prioritized the goal of the measure to assess substantial clinical benefit (SCB) improvement in patient-reported outcomes for elective primary THA/TKA patients with minimal patient and provider	5a.2 If not completely harmonized, identify difference, rationale, impact: NA 5b.1 If competing, why superior or rationale for additive value: NA	5a.2 If not completely harmonized, identify difference, rationale, impact: NA 5b.1 If competing, why superior or rationale for additive value: NA	5a.2 If not completely harmonized, identify difference, rationale, impact: 5b.1 If competing, why superior or rationale for additive value:	5a.2 If not completely harmonized, identify difference, rationale, impact: N/A 5b.1 If competing, why superior or rationale for additive value: N/A	5a.2 If not completely harmonized, identify difference, rationale, impact: NA 5b.1 If competing, why superior or rationale for additive value: NA

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	burden over harmonization if discrepancies occur. 5b.1 If competing, why superior or rationale for additive value: NQF # 2653: Average change in functional status following total knee replacement surgery. This PRO-PM measure differs from NQF #2653 in attribution, cohort, outcome, and risk adjustment. Attribution: This PRO-PM is a hospital-level quality measure, whereas NQF #2653 is a clinician-level measure. Cohort: This PRO-PM includes both THA and TKA procedures, as clinical experts agree that hospital-level processes are shared across these procedures, and includes only primary, not revision, procedures, based upon clinical input that revision procedures are more complicated to perform and patient-reported outcomes may be influenced by the initial surgery. The target population is Medicare FFS beneficiaries 65 years of age and older. NQF #2653 includes only TKA procedures, includes knee replacement revisions as well as primary procedures, and includes all adults 18 years of age and older. Outcome: This PRO-PM collects PROs with the HOOS, JR for THA patients and the KOOS, JR for TKA patients. Timing of PRO data collection is 90 – 0 days prior to and 270 – 365 days following surgery. The numerator measures SCB improvement for each patient from preoperative to postoperative assessment with a binary outcome (Yes/No), and the measure produces a risk-standardized improvement rate that elucidates for hospitals the					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	risk-adjusted proportion of patients with improvement and those without improvement. In contrast, NQF #2653 collects PRO data with the Oxford Knee Score three months prior to and 9 – 15 months following surgery, and measures average change in knee function score. The outcome definition of SCB, with a defined threshold for change in PROM score, allows patients with poorer baseline PRO scores more room to improve and thus a greater opportunity to achieve SCB. This was identified by our TEP members as a specific benefit of measuring SCB versus average change; measuring SCB incentivizes providers to offer and perform THA/TKA procedures on even those with poor PRO scores. Further stated TEP and Patient Working Group concerns with measuring an average change score included the fact that hospitals with all average outcomes would look similar to hospitals whose patients either did very well or very poorly (bimodal distributed outcomes), thus providing potentially misleading information to consumers and patients. Risk Adjustment: This risk model for this PRO-PM includes important risk variables supported by technical expert panel (TEP) and other expert clinical consultants including health literacy, other musculoskeletal pain and chronic narcotic use which are not included in NQF #2653; these risk variables were identified and tested based upon input from orthopedic professional societies, including AAHKS and AAOS, through public comment (Centers for Medicare & Medicaid Services ,					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0422: Functional status change for patients with Knee impairments	0423: Functional status change for patients with Hip impairments	0424: Functional status change for patients with Foot and Ankle impairments	0425: Functional Status Change for Patients with Low Back Impairments	0426: Functional status change for patients with Shoulder impairments
	CJR Final Rule 2015, Section III.D.3.A). This PRO-PM is superior to NQF #2653: 1) it more appropriately provides a signal of hospital quality which reflects outcomes for both THA and TKA recipients since within hospitals, care for patients undergoing THA/TKA procedures is provided by the same providers and hospital staff; 2) it assesses SCB improvement with a binary outcome that elucidates for hospitals and patients the risk-adjusted proportion of patients with and without improvement (a clear, understandable metric that patients support); 3) it uses a more robust and stakeholder-driven risk model, anticipated to produce a measure with greater face validity with stakeholders; and 4) it is harmonized with related measures including NQF #1550 Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA) and 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 (MUC19-28). References: Comprehensive Care for Joint Replacement (CJR) Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services Final Rule, 80 C.F.R. 73273 (Nov 24, 2015).					

Comparison of NQF 3559 and NQF 0427, 0428, 1550, 1551, 2643

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
Steward	Centers for Medicare & Medicaid Services (CMS)	Focus on Therapeutic Outcomes, Inc	Focus on Therapeutic Outcomes, Inc	Centers for Medicare & Medicaid Services	Centers for Medicare & Medicaid Services	MN Community Measurement
Description	This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.	A self-report outcome measure of functional status for patients 14 years+ with elbow, wrist, hand impairments. The change in functional status assessed using FOTO (elbow, wrist and hand) PROM is adjusted to patient characteristics known to be associated with functional status outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician, and at the clinic level to assess quality.	A self-report outcome measure of functional status for patients 14 years+ with general orthopaedic impairments. The change in functional status assessed using FOTO (general orthopedic) PROM is adjusted to patient characteristics known to be associated with functional status outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician, and at the clinic level by to assess quality.	The measure estimates a hospital-level risk-standardized complication rate (RSCR) associated with elective primary THA and TKA in Medicare Fee-For-Service beneficiaries who are 65 years and older. The outcome (complication) is defined as any one of the specified complications occurring from the date of index admission to 90 days post date of the index admission (the admission included in the measure cohort). The target population is patients 18 and over. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal acute-care hospitals.	This measure estimates a hospital-level, 30-day RSRR following elective primary THA and/or TKA. The outcome is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission. A specified set of planned readmissions do not count as readmissions. The Centers for Medicare & Medicaid Services (CMS) annually reports the measure for patients who are 65 years and older and are Medicare fee-for-service (FFS) beneficiaries hospitalized in non-federal hospitals.	For patients age 18 and older undergoing lumbar spine fusion surgery, the average change from pre-operative functional status to one year (nine to fifteen months) post-operative functional status using the Oswestry Disability Index (ODI version 2.1a) patient reported outcome tool.
Type	Outcome: PRO-PM	Outcome: PRO-PM	Outcome: PRO-PM	Outcome	Outcome	Outcome: PRO-PM
Data Source	Claims, Instrument-Based Data The PROM surveys used to define the measure outcome are 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are presently available in English, not yet in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) Global or the Veterans RAND 12 Item Health Survey (VR-	Electronic Health Data, Instrument-Based Data, Paper Medical Records Focus On Therapeutic Outcomes, Inc maintains the database. Information on the instrument, risk-adjustment procedures etc. is available at http://www.fotoinc.com/science-of-foto/NQF0425.html Available at measure-specific web page URL identified in S.1 Attachment ElbowWristHand_data_dictionary_2017Dec.xlsx	Electronic Health Data, Instrument-Based Data, Paper Medical Records Focus On Therapeutic Outcomes, Inc maintains the database. Information on the instrument, risk-adjustment procedures etc. is available at http://www.fotoinc.com/science-of-foto/NQF0425.html Available at measure-specific web page URL identified in S.1 Attachment General_Ortho_data_dictionary_2017Dec-636501409496471203.xlsx	Claims Data sources: The currently publically reported measure is specified and has been tested using: 1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission. 2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such	Claims Data sources: The currently publically reported measure is specified and has been testing using: 1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission. 2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such	Instrument-Based Data, Other, Paper Medical Records Oswestry Disability Index (ODI) version 2.1a A ten item self-administered questionnaire with a six point Likert response scale. Items are scored on a 0 to 5 scale with 0 indicating no limitation of function due to pain and 5 indicating major functional disability due to back pain. Time for patient completion is 3 to 5 minutes. Languages available are English and Spanish. The tool is available for use in clinical practice at no cost and can be obtained by completing a user agreement with MAPI Trust, Inc. The ODI is a valid, reliable, and responsive condition-specific assessment tool that is suited for use in clinical practice. It is easy to administer and score, objectifies

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	12) are used. The PROMIS Global is available in sixteen languages; the VR-12 is available in Spanish, Chinese and German. Available in attached appendix at A.1 Attachment Del4-8bHBPNQF3559HipKneePROPMDataDict_For_Submission030520.xlsx			as Medicare status on admission as well as vital status at discharge. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992). During original measure development we validated the administrative claims-based definition of THA/TKA complication (original model specification) against a medical record data. 3. Data abstracted from medical records from eight participating hospitals (approximately 96 records per hospital; 644 total records) for Medicare beneficiaries over the age of 65 years who had a qualifying THA/TKA procedure between January 1 2007 and December 31, 2008. The measure was also specified and testing using an all-payer claims dataset although it is only publically reported using the data sources listed above 4. California Patient Discharge Data is a large, linked database of patient hospital admissions in the state of California. Using all-payer data from California, we performed analyses to determine whether the THA/TKA complication measure can be applied to all adult patients, including not only FFS Medicare patients aged 65 years or over, but also non-FFS Medicare patients aged 18-64 years at the time of admission. Additional Data source used for analysis of the impact of SES variables on the measure's risk model. Note, the variables derived from these data are not included in the measure as specified 5. The American Community Survey (2009-2013): The American	as Medicare status on admission as well as vital status at discharge. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992). The measure was also specified and testing using an all-payer claims dataset although it is only publically reported using the data sources listed above: 3. California Patient Discharge Data in addition to CMS Medicare FFS data for patients in California hospitals. Using all-payer data from California, we performed analyses to determine whether the THA/TKA readmission measure can be applied to all adult patients, including not only FFS Medicare patients aged 65 years or over, but also non-FFS Medicare patients aged 18-64 years at the time of admission. Additional data source used for the analysis of the impact of SES variables on the measure's risk model. Note that the variables derived from these data are not included in the measure as specified 4. The American Community Survey (2009-2013): The American Community Survey data is collected annually and an aggregated 5-years data was used to calculate the AHRQ socioeconomic status (SES) composite index score. Reference: 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. Dorsey K, Grady J, Desai N, et al. 2016 Procedure-Specific Measures	client's complaints, and monitors effects of therapy. The ODI shows good construct validity; internal consistency is rated as acceptable; test-retest reliability and responsiveness have been shown to be high; and burden of administration is low. Internal consistency with Cronbach's alpha in ranges from .17 to .87 with test re-test reliability ranges of r = 0.83 to 0.99 and intraclass correlation coefficient values from 0.84 to 0.94. (Vinanin Psychometric properties and clinical usefulness of the Oswestry Disability Index Journal of Chiropractic Medicine 2008). Available in attached appendix at A.1 Attachment MNCM_Data_Dictionary_Lumbar_Spine-635490746124015022-635733283337486734.xlsx

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
				Community Survey data is collected annually and an aggregated 5-years data was used to calculate the AHRQ socioeconomic status (SES) composite index score. Reference: 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. Suter LG, Parzynski CS, Grady JN, et al. 2014 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 3.0). 2014 No data collection instrument provided Attachment Del20mHOP5NQF1550HKComplicationDataDictionary022519.xlsx	Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures: Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) & Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (Version 5.0). 2016 No data collection instrument provided Attachment Del20nHOP5NQF1551HKReadmissionDataDictionary022519.xls	
Level	Facility	Facility, Clinician : Group/Practice, Clinician : Individual	Facility, Clinician : Group/Practice, Clinician : Individual	Facility	Facility	Clinician : Group/Practice
Setting	Inpatient/Hospital	Other, Outpatient Services, Post-Acute Care Hospital Outpatient	Other, Outpatient Services, Post-Acute Care Hospital Outpatient	Inpatient/Hospital	Inpatient/Hospital, Other Hospital: Acute Care Facility	Outpatient Services
Numerator Statement	The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows: - For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis	Patient Level: The residual functional status score for the individual patient (residual scores are the actual change scores - predicted change after risk adjustment). Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for elbow, wrist and hand impairment. Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month	Patient Level: The residual functional status score for the individual patient (residual scores are the actual change scores - predicted change after risk adjustment). Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for general orthopaedic impairment. Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month	The outcome for this measure is any complication occurring during the index admission (not coded present on arrival) to 90 days post-date of the index admission. Complications are counted in the measure only if they occur during the index hospital admission or during a readmission. The complication outcome is a dichotomous (yes/no) outcome. If a patient experiences one or more of these complications in the applicable time period, the complication outcome for that	The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge of the index hospitalization. If a patient has more than one unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an	There is not a traditional numerator for this measure; the measure is calculating the average change in functional status score from pre-operative to post-operative functional status score. The measure is NOT aiming for a numerator target value for a post-operative ODI score. For example: The average change in low back function was an increase in 17.2 points one year post-operatively on a 100 point scale.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	Outcome Score for Joint Replacement (HOOS, JR); and - For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR). SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group. References: Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.	time period for elbow, wrist and hand impairments.	time period for general orthopaedic impairment.	patient is counted in the measure as a “yes”. Additional details are provided in S.5 Numerator Details.	unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission, because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission. Additional details are provided in S.5 Numerator Details.	
Numerator Details	This is a patient-reported outcome-based performance measure (PRO-PM). Two joint-specific Patient Reported Outcome Measure (PROM) surveys are used to collect the data for calculating the numerator: 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These PROM data and specific risk variable data will be collected 90 to 0 days prior to surgery, and PROM data will be collected again 270 to 365 days following surgery. Data elements used to define the numerator and for risk adjustment that are collected with PROM data include: - HOOS, JR or KOOS, JR - Date of Birth - Single-Item Literacy Screening (SILS2) Questionnaire	Patient Level: The residual score for the individual patients with elbow, wrist and hand impairments is derived by applying the statistical risk adjustment model described in S.14 and S.15 and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of this application. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for elbow, wrist and hand impairment. Average scores are calculated using data from all clinicians, however performance is evaluated only for those clinicians that had a minimum of 10 patients in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40	Patient Level: The residual score for the individual patients with general orthopaedic impairments is derived by applying the statistical risk adjustment model described in S.14 and S.15 and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of this application. Individual Clinician Level: The average of residuals in functional status scores in patients who were treated by a clinician in a 12 month time period for general orthopaedic impairment. Average scores are calculated using data from all clinicians, however performance is evaluated only for those clinicians that had a minimum of 10 patients in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinician regardless of	Outcome Definition The composite complication is a dichotomous outcome (yes for any complication(s); no for no complications). Therefore, if a patient experiences one or more complications, the outcome variable will get coded as a "yes". Complications are counted in the measure only if they occur during the index hospital admission (and are not present on admission) or during a readmission. The complications captured in the numerator are identified during the index admission OR associated with a readmission up to 90 days post-date of index admission, depending on the complication. The follow-up period for complications from date of index admission is as follows: • The follow-up period for AMI, pneumonia, and sepsis/septicemia/shock is seven days from the date of index admission because these	Outcome Definition The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index THA and/or TKA hospitalization, excluding planned readmissions as defined below. Rationale: Planned readmissions are generally not a signal of quality of care. Including planned readmissions in a readmission measure could create a disincentive to provide appropriate care to patients who are scheduled for elective or necessary procedures within 30 days of discharge. From a patient perspective, an unplanned readmission from any cause is an adverse event. Outcomes occurring within 30 days of discharge can be influenced by hospital care and the early transition to the non-acute care setting. The 30-day time frame is a clinically meaningful period for hospitals to collaborate with their	There is not a traditional numerator for this measure; the measure is calculating the average change in functional status score from pre-operative to post-operative functional status score. The measure is NOT aiming for a numerator target value for a post-operative ODI score. The average change is calculated as follows: Change is first calculated for each patient and then changed scores are summed and then an average is determined. Measure calculation takes into account those patients that have an improvement and those patients whose function decreases post-operatively. Example below: Patient Pre-op ODI Post-op ODI Change in ODI Patient A 47 18 29 Patient B 45 52 -7

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	- Body Mass Index (BMI) or Weight (kg) and Height (cm) - Chronic (>90 Day) Narcotic Use - Total Painful Joint Count (Patient-Reported in Non-Operative Lower Extremity Joint) - Quantified Spinal Pain (Patient-Reported Back Pain, Oswestry Index Question) - PROMIS Global Mental Health Score (calculated with data from the Patient-Reported Outcomes Measurement Information Systems (PROMIS) Global or Veteran's Rand 12-Item Health Survey (VR-12); data from VR-12 is translated to PROMIS Global Mental Health scores using a crosswalk created by Cella et. al for PROsetta® Stone) (Please note: Data elements listed above are detailed in the Data Dictionary accompanying this NQF submission; see Tabs: Risk Variables with PRO Data; HOOS, JR; KOOS, JR; PROMIS Global; VR-12) Center for Medicare and Medicaid Services (CMS) administrative data is used to identify eligible THA/ TKA procedures for the measure cohort (denominator) and additional risk variables, including patient demographics and clinical comorbidities (ICD-10 codes for eligible THA/TKA procedures identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.) The numerator is the risk-adjusted proportion of patients undergoing	patients/clinician regardless of clinic size, but has recently changed its procedure to enable participation by clinicians that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month time period for elbow, wrist and hand impairment. Average scores are calculated using data from all clinics, however performance is evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinics regardless of clinic size, but has recently changed its procedure to enable participation by smaller clinics that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Both comparative benchmark reports (clinician or clinic level) include patients with elbow, wrist and hand impairments, who were treated in therapy and had their functional status assessed at admission and at the end of their episode of therapy and were discharged from therapy.	clinic size, but has recently changed its procedure to enable participation by clinicians that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Clinic Level: The average of residuals in functional status scores in patients who were treated by a clinic in a 12 month time period for general orthopaedic impairment. Average scores are calculated using data from all clinics, however performance is evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months to maximize stability of the benchmarking estimates. In 2011-2013, FOTO used a standard threshold of 40 patients/clinics regardless of clinic size, but has recently changed its procedure to enable participation by smaller clinics that do not have a sufficient volume of patients. The score is derived by applying steps 1-6 as described in S.18 Both comparative benchmark reports (i.e., scale ranging from 0 to 100 with higher scores meaning higher functional abilities) at the clinician or clinic level include patients with general orthopaedic impairments, who were treated in therapy and had their functional status assessed at the end of their episode of therapy and were discharged from therapy.	conditions are more likely to be attributable to the procedure if they occur within the first week after the procedure. Additionally, analyses indicated a sharp decrease in the rate of these complications after seven days. • The follow up period for death, surgical site bleeding, and pulmonary embolism is 30 days following admission because clinical experts agree these complications are still likely attributable to the hospital performing the procedure during this period and rates for these complications remained elevated until roughly 30 days post admission. • The measure follow-up period is 90 days after admission for mechanical complications and periprosthetic joint infection/wound infection. Experts agree that mechanical complications and periprosthetic joint infection/wound infections due to the index THA/TKA occur up to 90 days following THA/TKA. The measure counts all complications occurring during the index admission regardless of when they occur. For example, if a patient experiences an AMI on day 10 of the index admission, the measure will count the AMI as a complication, although the specified follow-up period for AMI is seven days. Clinical experts agree with this approach, as such complications likely represent the quality of care provided during the index admission. As of 2014 reporting, the measure does not count complications in the complications outcome that are coded as POA during the index admission; this prevents	communities to reduce readmissions. Planned Readmission Algorithm (Version 4.0) The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital. The Planned Readmission Algorithm has three fundamental principles: 1. A few specific, limited types of care are always considered planned (transplant surgery, maintenance chemotherapy/immunotherapy, rehabilitation); 2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and 3. Admissions for acute illness or for complications of care are never planned. The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort. THA/TKA readmission measures make a few modifications to the	Patient C 56 12 44 Patient D 62 25 37 Patient E 42 57 -15 Patient F 51 10 41 Patient G 62 25 37 Patient H 43 20 23 Patient I 74 35 39 Patient J 59 23 36 Average change in ODI one year post-op 26.4 points on a 100 point scale

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	an elective primary THA/TKA that meet or exceed a SCB improvement on the HOOS, JR or KOOS, JR from preoperative to postoperative assessment. SCB improvement is defined as: - For THA patients, an increase of 22 points or more on the HOOS, JR - For TKA patients, an increase of 20 points or more on the KOOS, JR SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, TEP, and Technical Advisory Group. Further, the measure numerator was defined with extensive patient and clinician input. Among the numerator definitions considered by stakeholders during measure development included: - Change in PROM score from preoperative to postoperative assessment reported as an average for a hospital's patients; - Postoperative PROM score reported as an average for a hospital's patients; - A threshold change in PROM score from preoperative to postoperative assessment reported as a proportion of a hospital's patients meeting the threshold; - A threshold postoperative PROM score reported as a proportion of a hospital's patients meeting the threshold; and - A combination of both a minimum threshold change in PROM score from preoperative to postoperative assessment and a minimum threshold for postoperative PROM score. Clinical experts and patients supported a numerator definition			identifying a condition as a complication of care if it was present on admission for the THA/TKA procedure. For full list of ICD-10 codes defining complications, see the Data Dictionary attached.	planned readmission algorithm which can be found in the attached data dictionary. For more details on the Planned Readmission Algorithm, please see the report titled "2018 Procedure-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures, Version 7.0" posted in the webpage provided in data field S.1.	

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	that assessed change in PROM score from preoperative to postoperative assessment over a numerator definition that focused on postoperative PROM score. TEP members and patients noted that patients want to see improvement, and that the numerator definition should reflect change following surgery. Comments against using a numerator definition focusing on the postoperative PROM score included concern that it does not reflect degree of improvement, and may incentivize surgery on patients with less severe disease who have better preoperative scores. This concern about assessment of the postoperative PROM score also led to dislike of the last option noted above, a numerator definition combining threshold change and threshold postoperative PROM score. Stakeholders also strongly supported a numerator definition assessing a threshold change in PROM score over averaging patient change in PROM scores for hospital reporting. They noted that measurement of a threshold change will highlight lower performing patients, will protect at-risk patients, and is easy to understand as a performance measure. Comments against a reported average change included concern that a hospital whose patients all achieve average results could have a reported average change result that would be very similar to a hospital whose patients achieve either very good or very poor results; an average change numerator could show similar results for hospitals with very different patient outcomes).					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	The numerator definition of SCB threshold change, supported by patients and clinical experts, provides an easy to understand metric that patients found intuitive. Using a SCB threshold avoids the potential for misleading consumers and patients by averaging patient change scores across a hospital when individual patient outcomes within hospitals may vary considerably (as noted above). Using a SCB incentivizes providers to perform surgery on patients with worse baseline scores, a group that might otherwise not be offered surgery, as patients with poorer baseline PRO scores have more room to improve and thus a greater opportunity to achieve SCB. It also encourages providers to not perform THA/TKA procedures on patients with minimal symptoms, who will not benefit at all from surgery. And, since the SCB was defined with close input from patients and clinicians, it does set a minimum improvement threshold, but not one so large as to cause surgeons to avoid performing THA/TKA procedures on patients who would benefit. References: Cella D, Schalet BD, Kallen M, Lai JS, Cook KF, Rutsohn J, Choi SW. PROsetta® Stone Analysis Report Volume 2: A Rosetta Stone for Patient Reported Outcomes, PROMIS Global Health – Mental Component and VR-12 – Mental Component (Algorithmic Scores). http://www.prosettastone.org/LinkingTables1/GlobalHealth/Pages/default.aspx, 2018. Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.					
Denominator Statement	The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.	All patients 14 years and older with elbow, wrist or hand impairments who have initiated rehabilitation treatment and completed the FOTO (elbow, wrist and hand) PROM.	All patients 14 years and older with general orthopaedic impairments who have initiated rehabilitation treatment and completed the FOTO (general orthopaedic) PROM.	The target population for the publicly reported measure includes admissions for Medicare FFS beneficiaries who are at least 65 years of age undergoing elective primary THA and/or TKA procedures. Additional details are provided in S.9 Denominator Details.	The target population for the publicly reported measure includes admissions for Medicare FFS beneficiaries who are at least 65 years of age undergoing elective primary THA and/or TKA procedures. Additional details are provided in S.7 Denominator Details.	Adult patients age and older (no upper age limit) who undergo a lumbar spine fusion procedure during a calendar year performance period (e.g. dates of procedure occurring between 1/1/2013 and 12/31/2013) AND have a completed pre-operative and post-operati
Denominator Details	The cohort for this measure is Medicare FFS patients 65 years of age and older undergoing an elective primary THA/TKA procedure at a non-federal short-term acute care hospital. Inclusion criteria includes patients: - Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission - Discharged alive from a non-federal short-term acute care hospital - Undergoing only elective primary THA/TKA procedures (patients with fractures and revision procedures or with bone metastases are not included) - Inclusion criteria are harmonized with CMS's existing measure cohort for the hospital-level 90-day risk-standardized THA/TKA complication measure Center for Medicare and Medicaid Services (CMS) administrative data is used to identify qualifying THA/TKA procedures for the measure cohort. (ICD-10 codes for eligible THA/TKA procedures are identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.)	The established ICD-9-CM codes for measure include soft tissue disorders of muscle, synovium, tendon, bursa, or enthesopathies (ICD-9 codes 725-729); sprains and strains of the elbow, wrist or hand (ICD-9 codes 841-842 including unspecified sprain or strain); fractures (ICD-9 813-819 including humerus, ulna, radius, carpal bones, metacarpals); arthropathies (ICD-9 codes 710-719, including osteoarthoses, rheumatoid arthritis); disorders of the bone and cartilage (ICD-9 codes 730-739); dislocations of elbow, wrist or fingers (ICD-9 codes 832-834); post-surgical (CPT codes including 24301 elbow muscle or tendon transfer, 64721 carpal tunnel decompression). Please refer to the Letter of Intent submitted to NQF under separate cover (email) to complete ICD10 Mapping for this measure by end of February 2015.	The established ICD-9-CM codes for the neck, cranium, mandible, thoracic spine, ribs or other general orthopedic impairment include: Diagnosis specific to the cervical spine: 333.83, 353.2, 716.58, 718.88, 718.98, 719.08, 719.18, 719.48, 719.58, 719.68, 721.0, 721.1, 722.0, 722.4, 722.71, 722.81, 722.91, *723, 730.08, 730.09, 730.18, 739.1, 741.01, 741.91, 754.1, *805.0, *805.1, *806.0, *806.1, 847.0, *952.0, 953.0 * Use of an asterisk is to include all codes in the category Diagnosis specific to the thoracic spine: 353.3, 721.2, 721.41, 722.11, 722.31, 722.51, 722.72, 722.82, 722.92, 724.01, 724.1, 724.4, 724.5, 730.08, 730.09, 730.18, 739.2, 741.02, 741.92, 805.2, 805.3, *806.2, *806.3, 847.1, *952.1, 953.1 * Use of an asterisk is to include all codes in the category Diagnosis specific to the Cranium and Mandible 307.81, *346, *350.2, *351, *524.6, 754.0, 784.0, *830, 848.1 * Use of an asterisk is to include all codes in the category	To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria: 1. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of admission; and enrolled in Part A during the index admission; 2. Aged 65 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 in the principal or secondary discharge diagnosis fields of the index admission; • A concurrent partial hip or knee arthroplasty procedure, in discharges on or after October 1, 2015; a concurrent partial hip arthroplasty procedure, in discharges prior to October 1, 2015; • A concurrent revision, resurfacing, or implanted device/prosthesis removal procedure; • Mechanical complication coded in the principal discharge diagnosis field;	To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria: 1. Enrolled in Medicare FFS Part A and Part B Medicare for the 12 months prior to the date of admission; and enrolled in Part A during the index admission; 2. Aged 65 or over; 3. Discharged alive from a non-federal acute care hospital; and, 4. Have 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 in the principal or secondary discharge diagnosis fields of the index admission; •A concurrent partial hip or knee arthroplasty procedure, in discharges on or after October 1, 2015; a concurrent partial hip arthroplasty procedure, in discharges prior to October 1, 2015; •A concurrent revision, resurfacing, or implanted device/prosthesis removal procedure;	The initial patient population is adult patients age 18 and older (no upper age limit) who undergo a lumbar spine fusion procedure during a calendar year performance period (e.g. dates of procedure occurring between 1/1/2013 and 12/31/2013). CPT procedure codes: 22533, 22534, 22558, 22586 22612, 22630, and 22633. If any portion of the lumbar spine is fused (L1 to L5), the patient is to be included. If the fusion of the lumbar spine also incorporates thoracic vertebrae, the patient is to be included. Inclusion in the denominator that measures the average change between pre-operative and post-operative functional status requires completion of a patient reported outcome assessment tool (ODI) BOTH pre-operatively (within three months prior to the procedure) AND one year post-operatively (nine to fifteen months after the procedure) The denominator for calculating the average change in function at a practice level is those patients included in the initial patient population who have both a completed pre-operative and post-operative Oswestry Disability

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
			Diagnosis specific to the Ribs 733.6, 739.8, 756.2, 756.3, *786.5, *807.0, *807.1, 807.2, 807.3, 839.61, 848.3, *848.4, 922.1, 922.3 * Use of an asterisk is to include all codes in the category Diagnosis not specific to the cervical or thoracic spine, cranium/mandible or ribs, but effect the function of the cervical or thoracic spine, cranium/mandible, ribs or other general impairment: 338.29, 353.0, 353.8, 710.0, 711.98, 714.0, 715.09, 715.18, 715.19, 715.28, 715.38, 715.88, 715.89, 715.98, 716.98, 716.99, 716.59, 716.98, 716.99, 718.08, 718.09, 718.19, 718.28, 718.29, 718.38, 718.39, 719.49, 719.59, 718.89, 718.99, 720.0, 720.9, *721.9, 722.2, 722.6, 724.00, 724.09, 724.08, 724.5, 724.9, 728.2, 728.85, 728.87, 730.19, 732.0, *733.0, 733.13, 733.90, *737, 754.2, 756.19, 759.79, 781.92, 847.9, 952.8, V54.17, V54.89, V57.1, V59.49, V67.0 * Use of an asterisk is to include all codes in the category The ICD 10 Crosswalk is provided on the measure specific webpage provided in S.1.	• 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; or, • Transfer from another acute care facility for the THA/TKA ICD-9 and ICD-10 cohort codes are included in the attached Data Dictionary.	•Mechanical complication coded in the principal discharge diagnosis field; •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; or, •Transfer from another acute care facility for the THA/TKA ICD-9 and ICD-10 cohort codes are included in the attached Data Dictionary.	Index patient reported outcome tool (ODI version 2.1a)
Exclusions	Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.	•Patients who are not being treated for an elbow, wrist and/or hand impairment •<14 years of age	•Patients who are not being treated for a General orthopaedic impairment •<14 years of age	This measure excludes index admissions for patients: 1. Without at least 90 days post-discharge enrollment in FFS Medicare; 2. Who were discharged against medical advice (AMA); or, 3. Who had more than two THA/TKA procedure codes during the index hospitalization. After exclusions #1-3 are applied, the measure randomly selects one index admission per patient per	This Hip/knee readmission measure excludes admissions for patients: 1. Without at least 30 days post-discharge enrollment in Medicare FFS; 2. Discharged against medical advice; 3. Admitted for the index procedure and subsequently transferred to another acute care facility;	Exclusions are for patients with spine related cancer, fracture and infection and idiopathic or congenital scoliosis.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
				year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. Additional admissions within that year are excluded. For the three-year combined data, when index admissions occur during the transition between two years within the measurement period (that is, March and April-June 2015 or March and April-June 2016), and both are randomly selected for inclusion in the measure, the measure includes both admissions, but a complication that falls within the defined timeframe for both admissions would only be captured in the complication outcome for the first admission.	4. Who had more than two THA/TKA procedure codes during the index hospitalization; or 5. Who had THA/TKA admissions within 30 days of a prior THA/TKA index admission	
Exclusion Details	Patients with staged procedures in the measure period are excluded. A staged procedure is identified if a patient has more than one hospitalization with an eligible, elective primary THA or TKA procedure during the measurement period. ICD-10 codes for eligible, elective primary THA/TKA procedures (listed in the Data Dictionary on "ICD-10 2017-2018" tab) are used to identify all eligible procedures during the measurement period; patients with an ICD-10 code for an eligible elective primary THA or TKA procedure in two or more hospital admissions during the measurement period are identified as having a staged procedure, and the patient, including all procedures, is removed from the measure cohort.	• Patients who are not being treated for an elbow, wrist and/or hand impairment • Age under 14 years old.	• Patients who are not being treated for a general orthopaedic impairment • Age under 14 years old.	This measure excludes index admissions for patients: 1. Without at least 90 days post-discharge enrollment in Medicare FFS, which is identified is by examining the Medicare Enrollment Database. 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), which are identified using the discharge disposition indicator in claims data.; or, 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 admission, which is identified by examining procedure codes in the claims data.	This measure excludes index admissions for patients: 1. Without at least 30 days of post-discharge enrollment in Medicare FFS, which is identified is by examining the Medicare Enrollment Database. Rationale: The 30-day readmission outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted. 2. Discharged against medical advice (AMA), which are identified using the discharge disposition indicator in claims data. Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge. 3. Admitted for the index procedure and subsequently transferred to another acute care facility, as identified in claims data, which are defined as when a patient with an inpatient hospital admission (with at least one qualifying THA/TKA procedure) is	Patients who are undergoing a lumbar spine fusion procedure for an acute fracture (trauma), metastatic or bone cancer, infection or scoliosis are not included in this patient population because their expected course of care and outcomes could be significantly different from the population of patients undergoing the procedure for relief of back and/or leg pain (degenerative disc disease, disc herniation, stenosis or spondylolisthesis). ICD-9/ ICD-10 diagnosis codes for exclusions are provided in the data dictionary at S.2.b

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
				Rationale: 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.	discharged from an acute care hospital and admitted to another acute care hospital on the same or next day. Rationale: Patients admitted for the index procedure and subsequently transferred to another acute care facility are excluded, as determining to which hospital the readmission outcome should be attributed is difficult. 4. With more than two THA/TKA procedure codes during the index admission, which is identified by examining procedure codes in the claims data. Rationale: 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. 5. THA/TKA admissions within 30 days of discharge from a prior THA/TKA index admission, which are identified by comparing the discharge date from the index admission with the readmission date. Rationale: Additional THA/TKA admissions within 30 days are excluded as index admissions because they are part of the outcome. A single admission is not considered both an index admission and a readmission for another index admission.	
Risk Adjustment	Statistical risk model	Statistical risk model	Statistical risk model	Statistical risk model	Statistical risk model	Statistical risk model
Stratification	N/A	Risk adjusted - not stratified	Risk adjusted - not stratified	N/A	N/A	Clinical Condition Reason for Procedure field is collected for purposes of stratification (potential) or use in a risk adjustment model (more likely). The choices for this variable are: 1 = Degenerative Disc Disease, 2 =

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
						Disc Herniation, 3 = Spinal Stenosis, 4 = Spondylolisthesis. These conditions are definable by ICD-9/ ICD-10 codes and are provided in the data dictionary at S.2.b. The use of this variable for stratification of outcomes is dependent on procedure volume at the practice level; it has been our experience so far that the volumes at a practice level do not support reliable stratification by four variables as they may result in volumes that do not meet our standards for public reporting at the practice level. These variables, however, are important for several reasons. The may prove appropriate for inclusion in a future risk adjustment model. They also serve analytical purposes for further understanding of the patient reported outcome rates as some of the conditions represent an area of controversy in terms of appropriateness of procedures and successful outcomes for the patient.
Type Score	Rate/proportion better quality = higher score	Continuous variable, e.g. average better quality = higher score	Continuous variable, e.g. average better quality = higher score	Rate/proportion better quality = lower score	Rate/proportion better quality = lower score	Continuous variable, e.g. average better quality = higher score
Algorithm	Target population: Medicare FFS patients 65 years and older undergoing an elective primary THA or TKA in a non-federal short-term acute care hospital. To create the denominator: Step 1. If the patient is a Medicare FFS patient, go to Step 2. If not, do not include in the denominator. Step 2. If the patient is identified in CMS administrative claims data as having undergone an eligible elective primary THA or TKA during the measurement period, go to Step 3. If not, do not include in the denominator.	STEPS TAKEN TO PRODUCE THIS MEASURE: Definitions: Patient's Functional Status Score. A functional status score is produced when the patient completes the FOTO (elbow, wrist and hand) PROM administered by internet or a paper and pencil survey. The functional status score is continuous and linear. Scores range from 0 (low function) to 100 (high function). The survey is standardized, and the scores are validated for the measurement of function for this population.	STEPS TAKEN TO PRODUCE THIS MEASURE: Definitions: Patient's Functional Status Score. A functional status score is produced when the patient completes the FOTO (general orthopaedic) PROM administered by internet or a paper and pencil survey. The functional status score is continuous and linear. Scores range from 0 (low function) to 100 (high function). The survey is standardized, and the scores are validated for the measurement of function for this population.	The measure estimates hospital-level RSCRs following elective primary THA/TKA using hierarchical logistic regression models. 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 90 days of the index admission using age, sex, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it	The measure estimates hospital-level 30-day all-cause RSRRs following each procedure using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of hospital readmission within 30 days of discharge using age, sex, selected clinical covariates, and a hospital-specific effect. At the hospital level, the approach models the	Please also refer to measure flow logic in the data dictionary in S.2.b and flow chart in Appendix A-1 Initial patient population: Was the patient born on or prior to 01/01/xxxx? Did the patient undergo a lumbar fusion (any portion of the lumbar spine) procedure between 01/01/2013 to 12/31/2013? Patients who had fusion of the lumbar spine which incorporate the thoracic vertebrae are included. Does the patient have one of the following CPT codes?

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	Step 3. If the patient is 65 years of age or older, go to Step 4. If not, do not include in the denominator. Step 4. If the patient was enrolled in Medicare FFS Part A and Part B for the 12 months prior to index admission, and enrolled in Part A during the index admission, then go to Step 5. If not, do not include in the denominator. Step 5. If the patient was discharged alive from the hospital, include in the denominator. If not, do not include in the denominator. Step 6. If the patient experienced only one elective primary THA/TKA during the measurement period, or if the patient experience more than one elective primary THA/TKA during a singular hospitalization during the measurement period, + in the denominator. If the patient experienced two elective primary THA/TKA procedures during the measurement period performed during distinct hospitalizations, do not include in the denominator. To create the numerator: If the patient has complete PRO data collected during the prescribed preoperative and postoperative time windows, and meets or exceeds the SCB improvement threshold on the joint-specific PROM between the preoperative and postoperative assessment: - for THA patients, an increase of 22 points on the HOOS, JR - for TKA patients, an increase of 20 points on the KOOS, JR then include in the numerator. If not, then do not include in the numerator. The hospital-level measure result is calculated by aggregating all	Patient's Functional Status Change Score. A functional status change score is calculated by subtracting the Patient's Functional Status Score at Admission from the Patient's Functional Status Score at Discharge. Predicted Functional Status Change Score. Functional Status Change Scores for patients are risk adjusted using multiple linear regression methods that accounted for the following independent variables: Patient's Functional Status Score at Admission, patient age, symptom acuity, surgical history, gender, specific co morbidities, payer type, use of medication for the elbow/wrist/hand impairment at Intake, previous treatment for the impairment, exercise history, and post-surgical category if applicable. The Patient's Functional Status Change Score is the dependent variable. The statistical regression produces a Risk-Adjusted Predicted Functional Status Change Score. Risk-adjusted Functional Status Change Residual Score. The difference between the actual change and predicted change scores (after risk adjustment) is the residual score and should be interpreted as the unit of functional status change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted residual change scores of zero (0) or greater (>0) should be interpreted as functional status	Patient's Functional Status Change Score. A functional status change score is calculated by subtracting the Patient's Functional Status Score at Admission from the Patient's Functional Status Score at Discharge. Predicted Functional Status Change Score. Functional Status Change Scores for patients are risk adjusted using multiple linear regression methods that accounted for the following independent variables: Patient's Functional Status Score at Admission, patient age, symptom acuity, surgical history, gender, specific co morbidities, payer type, exercise history, use of medication for the condition, and previous treatment for the condition. The Patient's Functional Status Change Score is the dependent variable. The statistical regression produces a Risk-Adjusted Predicted Functional Status Change Score. Risk-adjusted Functional Status Change Residual Score. The difference between the actual change and predicted change scores (after risk adjustment) is the residual score and should be interpreted as the unit of functional status change different than predicted given the risk-adjustment variables of the patient being treated. As such, the risk-adjusted residual change score represents risk-adjusted change corrected for patient characteristics. Risk-adjusted residual change scores of zero (0) or greater (>0) should be interpreted as functional status change scores that were predicted or better than predicted given the risk-adjustment variables of the patient, and risk-adjusted residual	models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a complication at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals. The RSCR is calculated as the ratio of the number of "predicted" to the number of "expected" admissions 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 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 admissions with a complication (the numerator) is calculated by	hospital-specific effects as arising from a normal distribution. The hospital effect represents the underlying risk of a readmission 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 RSRR is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital, multiplied by the national observed readmission rate. For each hospital, the numerator of the ratio is the number of readmissions within 30 days predicted based on the hospital's performance with its observed case mix, and the denominator is the number of readmissions 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 a particular hospital's performance, given its case mix, to be compared to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected readmission rates or better quality, while a higher ratio indicates higher-than-expected readmission rates or worse quality. The "predicted" number of readmissions (the numerator) is calculated using the coefficients	22533, 22534, 22558, 22586, 22612, 22630, 22633 Inclusion in Denominator (has pre-op and post-op ODI) Valid date in the Pre-op ODI Date field? No = remove from denominator; Yes continue Is the Pre-op ODI Date field within 3 months prior to the procedure? No = remove from denominator; Yes continue Is there a value in the Pre-op ODI Summary Score field? Yes = Pre-op ODI Hold this score for calculation if postop score is present, if No evaluate if individual responses submitted for score calculation. Are there at least 8 completed value (valid 0 to 5) responses in the following fields? Pre-op ODI, Pain Pre-op ODI Care, Pre-op ODI Lifting, Pre-op ODI Walking, Pre-op ODI Sitting, Pre-op ODI Standing, Pre-op ODI Sleeping, Pre-op ODI Sex, Pre-op ODI Social, Pre-op ODI Travelling. If Yes = Pre-op ODI Hold this score for calculation if postop score is present, if No remove from the denominator. Is the 1 Yr Post-op ODI Date field within nine to fifteen months after the Date of Procedure? No = remove from denominator; Yes continue. Is there a value in the 1 Yr Post-op ODI Summary Score field ? If Yes 1 Yr Post-op ODI Hold this score for calculation, if No evaluate if individual responses submitted for score calculation. Are there at least 8 completed value (valid 0 to 5) responses in the following fields? 1 Yr Post-op ODI Pain, 1 Yr Post-op Care, 1 Yr Post-op Lifting, 1 Yr Post-op Walking, 1 Yr Post-op Sitting, 1 Yr

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	patient-level results across the hospital. For calculation of measure results, we recommend that hospitals should have a minimum case-volume of 25 elective primary THA/TKA patients with complete patient-reported outcomes and risk variable data collected 90 – 0 days preoperatively and complete patient-reported outcomes data collected 270 – 365 days postoperatively. Hospital-specific risk-standardized improvement rates (RSIRs) are calculated as the ratio of a hospital's "predicted" improvement to "expected" improvement multiplied by the overall observed improvement rate. Both predicted improvement and expected improvement are derived based on the output of a hierarchical logistic regression model that adjusts for patient case-mix and applies stabilized inverse probability weighting (IPW) to address potential non-response bias.	change scores that were predicted or better than predicted given the risk-adjustment variables of the patient, and risk-adjusted residual change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated risk-adjusted residual scores allow meaningful comparisons amongst clinicians or clinics. Aggregated risk-adjusted residual scores: The average of residual scores of functional status (actual change - predicted change after risk adjustment) from a provider (clinician or clinic). The aggregated scores are used to make comparisons between clinicians or clinics. STEPS: First, the patient completes FOTO (elbow, wrist and hand) PROM at Admission, which generates the Patient's Functional Status Score at Admission Second, patient completes FOTO (elbow, wrist and hand) PROM at or near Discharge, which generates the Patient's Functional Status Score at Discharge Third, the Patient's Functional Status Change Score (raw, non-risk-adjusted) is generated Fourth, a Risk-adjusted Predicted Functional Status Change Score is generated using a regression equation Fifth, a Risk-adjusted Functional Status Change Residual Score is generated for each patient.	change scores less than zero (<0) should be interpreted as functional status change scores that were less than predicted given the risk-adjustment variables of the patient. Aggregated risk-adjusted residual scores allow meaningful comparisons amongst clinicians or clinics. STEPS: First, the patient completes FOTO (general orthopaedic) PROM at Admission, which generates the Patient's Functional Status Score at Admission. Second, patient completes FOTO FOTO (general orthopaedic) PROM at or near Discharge, which generates the Patient's Functional Status Score at Discharge Third, the Patient's Functional Status Change Score (raw, non-risk-adjusted) is generated Fourth, a Risk-adjusted Predicted Functional Status Change Score is generated using a regression equation Fifth, a Functional Status Change Residual Score after risk adjustment is generated for each patient. Sixth, the average residual scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and	using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of having an admission with a complication. The estimated hospital-specific intercept 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 admissions with a complication (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 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 re-estimate the model coefficients using the years of data in that period. Multiplying the predicted over expected ratio by the national observed complication rate transforms the ratio into a rate that can be compared to the national observed complication rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2012). References: 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. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of	estimated by regressing the risk factors and the hospital-specific effect on the risk of readmission. The estimated hospital-specific effect is added to the sum of the estimated regression coefficients multiplied by patient characteristics. The results are log transformed and summed over all patients attributed to a hospital to calculate a predicted value. The "expected" number of readmissions (the denominator) is obtained in the same manner, except that 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 attributed to a hospital to calculate an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period. Multiplying the predicted over expected ratio by the national observed readmission rate transforms the ratio into a rate that can be compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2012). References: 1. Grosso L, Curtis J, Geary L, et al. Hospital-level 30-Day All-Cause Risk-Standardized Readmission Rate Following Elective Primary Total Hip Arthroplasty (THA) And/Or Total Knee Arthroplasty (TKA) Measure Methodology Report. 2012; https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=	Post-op Standing, 1 Yr Post-op Sleeping, 1 Yr Post-op ODI Sex, 1 Yr Post-op ODI Social, 1 Yr Post-op ODI Travelling. If Yes = Hold this score for calculation, if No remove from denominator. For each patient remaining in the denominator calculate the change in function by taking the pre-op ODI score and subtracting the one year post-op ODI score. Save this change score. To calculate the rate of average change in functional status for the practice; average the change in function score. Example: Patient ODI Pre-op ODI Post-op ODI Change in ODI Patient A 47 18 29 Patient B 45 52 -7 Patient C 56 12 44 Patient D 62 25 37 Patient E 42 57 -15 Patient F 51 10 41 Patient G 62 25 37 Patient H 43 20 23 Patient I 74 35 39 Patient J 59 23 36 Average change in ODI one year post-op 26.4

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
		Sixth, the average residual scores per clinician and/or clinic are calculated, and scores for all clinicians/clinics in the database are ranked. The quality score is the percentile of the clinician and/or clinic ranking. The quality scores and its 95% CI can be compared to the benchmark (a score of zero) to determine if the performance is below, at, or above the predicted average. FOTO recommends that clinicians have a minimum of 10 patients/year and clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (4 or more clinicians) in order to obtain stable estimates of provider performance.	clinics have a minimum of 10 patients/therapist per year for small clinics or 40 patients per year for larger clinics (5 or more clinicians) in order to obtain stable estimates of provider performance.	Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.	QnetPublic%2FPage%2FQnetTier4&cid=1219069855841. 2. Normand S-L, Shahian D. Statistical and clinical aspects of hospital outcomes profiling. Statistical Science. 2007;22(2):206-226.	
Submission items	5.1 Identified 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) 2958 : Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery 0422 : Functional status change for patients with Knee impairments 0423 : Functional status change for patients with Hip impairments 0424 : Functional status change for patients with Foot and Ankle impairments 0425 : Functional Status Change for Patients with Low Back Impairments	5.1 Identified measures: 5a.1 Are specs completely harmonized? 5a.2 If not completely harmonized, identify difference, rationale, impact: 5b.1 If competing, why superior or rationale for additive value: NA	5.1 Identified measures: 5a.1 Are specs completely harmonized? 5a.2 If not completely harmonized, identify difference, rationale, impact: 5b.1 If competing, why superior or rationale for additive value:	5.1 Identified measures: 0534 : Hospital specific risk-adjusted measure of mortality or one or more major complications within 30 days of a lower extremity bypass (LEB). 0564 : Cataracts: Complications within 30 Days Following Cataract Surgery Requiring Additional Surgical Procedures 1551 : Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) 2052 : Reduction of Complications through the use of Cystoscopy during Surgery for Stress Urinary Incontinence 5a.1 Are specs completely harmonized? Yes 5a.2 If not completely harmonized, identify difference, rationale, impact: We did not include in our	5.1 Identified measures: 0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization 0505 : Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization. 0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization 1550 : Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) 1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR) 1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization	5.1 Identified measures: 0425 : Functional status change for patients with lumbar impairments 5a.1 Are specs completely harmonized? No 5a.2 If not completely harmonized, identify difference, rationale, impact: Significant differences in these two measures; related but not competing. Only commonality is the desire to measure change in functional status. Target populations, settings of care and provider types are completely different as are the mechanisms for measuring change. # 0425 targets physical therapy settings and providers, for a population of patients with low back pain, and uses a proprietary (monthly/ per provider fee based) web-based CAT tool. Our measure focuses on patients undergoing lumbar fusion procedures, focus on orthopedic and neurosurgery providers in the

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	0426 : Functional status change for patients with Shoulder impairments 0427 : Functional status change for patients with elbow, wrist and hand impairments 0428 : Functional status change for patients with General orthopaedic impairments 2643 : Average change in functional status following lumbar spine fusion surgery 5a.1 Are specs completely harmonized? Yes 5a.2 If not completely harmonized, identify difference, rationale, impact: To the extent feasible, we have harmonized with existing, related measures. However, we have prioritized the goal of the measure to assess substantial clinical benefit (SCB) improvement in patient-reported outcomes for elective primary THA/TKA patients with minimal patient and provider burden over harmonization if discrepancies occur. 5b.1 If competing, why superior or rationale for additive value: NQF # 2653: Average change in functional status following total knee replacement surgery. This PRO-PM measure differs from NQF #2653 in attribution, cohort, outcome, and risk adjustment. Attribution: This PRO-PM is a hospital-level quality measure, whereas NQF #2653 is a clinician-level measure. Cohort: This PRO-PM includes both THA and TKA procedures, as clinical experts agree that hospital-level processes are shared across these procedures, and includes			list of related measures any non-outcome measures (for example, process measures) with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure). 5b.1 If competing, why superior or rationale for additive value: N/A	5a.1 Are specs completely harmonized? Yes 5a.2 If not completely harmonized, identify difference, rationale, impact: We did not include in our list of related measures any non-outcome measures (for example, process measures) with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure). 5b.1 If competing, why superior or rationale for additive value: N/A	ambulatory setting (pre and post procedure) and utilizes a valid, free PRO tool with strong psychometric properties that is easy to administer (3 to 5 min to complete) and easy to score. 5b.1 If competing, why superior or rationale for additive value: Measures do not address the same target population, providers or setting of care. They are related but not competing.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	only primary, not revision, procedures, based upon clinical input that revision procedures are more complicated to perform and patient-reported outcomes may be influenced by the initial surgery. The target population is Medicare FFS beneficiaries 65 years of age and older. NQF #2653 includes only TKA procedures, includes knee replacement revisions as well as primary procedures, and includes all adults 18 years of age and older. Outcome: This PRO-PM collects PROs with the HOOS, JR for THA patients and the KOOS, JR for TKA patients. Timing of PRO data collection is 90 – 0 days prior to and 270 – 365 days following surgery. The numerator measures SCB improvement for each patient from preoperative to postoperative assessment with a binary outcome (Yes/No), and the measure produces a risk-standardized improvement rate that elucidates for hospitals the risk-adjusted proportion of patients with improvement and those without improvement. In contrast, NQF #2653 collects PRO data with the Oxford Knee Score three months prior to and 9 – 15 months following surgery, and measures average change in knee function score. The outcome definition of SCB, with a defined threshold for change in PROM score, allows patients with poorer baseline PRO scores more room to improve and thus a greater opportunity to achieve SCB. This was identified by our TEP members as a specific benefit of measuring SCB versus average change; measuring SCB incentivizes providers to offer and perform THA/TKA procedures on					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	even those with poor PRO scores. Further stated TEP and Patient Working Group concerns with measuring an average change score included the fact that hospitals with all average outcomes would look similar to hospitals whose patients either did very well or very poorly (bimodal distributed outcomes), thus providing potentially misleading information to consumers and patients. Risk Adjustment: This risk model for this PRO-PM includes important risk variables supported by technical expert panel (TEP) and other expert clinical consultants including health literacy, other musculoskeletal pain and chronic narcotic use which are not included in NQF #2653; these risk variables were identified and tested based upon input from orthopedic professional societies, including AAHKS and AAOS, through public comment (Centers for Medicare & Medicaid Services , CJR Final Rule 2015, Section III.D.3.A). This PRO-PM is superior to NQF #2653: 1) it more appropriately provides a signal of hospital quality which reflects outcomes for both THA and TKA recipients since within hospitals, care for patients undergoing THA/TKA procedures is provided by the same providers and hospital staff; 2) it assesses SCB improvement with a binary outcome that elucidates for hospitals and patients the risk-adjusted proportion of patients with and without improvement (a clear, understandable metric that patients support); 3) it uses a more robust and stakeholder-driven risk					

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	0427: Functional status change for patients with elbow, wrist and hand impairments	0428: Functional status change for patients with General orthopaedic impairments	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)	2643: Average change in functional status following lumbar spine fusion surgery
	model, anticipated to produce a measure with greater face validity with stakeholders; and 4) it is harmonized with related measures including NQF #1550 Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA) and 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 (MUC19-28). References: Comprehensive Care for Joint Replacement (CJR) Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services Final Rule, 80 C.F.R. 73273 (Nov 24, 2015).					

Comparison of NQF 3559 and NQF 2958

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
Steward	Centers for Medicare & Medicaid Services (CMS)	Massachusetts General Hospital
Description	This patient-reported outcome-based performance measure will estimate a hospital-level, risk-standardized improvement rate (RSIR) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age and older. Improvement will be calculated with patient-reported outcome data collected prior to and following the elective procedure. The preoperative data collection timeframe will be 90 to 0 days before surgery and the postoperative data collection timeframe will be 270 to 365 days following surgery.	The measure is derived from patient responses to the Hip or Knee Decision Quality Instruments. Participants who have a passing knowledge score (60% or higher) and a clear preference for surgery are considered to have met the criteria for an informed, patient-centered decision. The target population is adult patients who had a primary hip or knee replacement surgery for treatment of hip or knee osteoarthritis.
Type	Outcome: PRO-PM	Outcome: PRO-PM
Data Source	Claims, Instrument-Based Data The PROM surveys used to define the measure outcome are 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients. These instruments can be administered in paper or electronic form, filled out in person or over the phone. The HOOS, JR and KOOS, JR are presently available in English, not yet in other languages. For measurement of global mental health for risk adjustment, the Patient-Reported Outcomes Measurement Information System (PROMIS) Global or the Veterans RAND 12 Item Health Survey (VR-12) are used. The PROMIS Global is available in sixteen languages; the VR-12 is available in Spanish, Chinese and German. Available in attached appendix at A.1 Attachment Del4-8bHBPNQF3559HipKneePROPMDataDict_For_Submission030520.xlsx	Instrument-Based Data The measure is derived from responses to the Hip and Knee Decision Quality Instruments. These patient reported surveys have been administered by mail, phone, and online for patients. The method we have used most often is mail with a postage paid return envelope. A combination of mail, email, and phone reminders are often needed to achieve adequate response rates. A third party vendor may also be used to administer the survey. We have used these questions in English and Spanish. Available in attached appendix at A.1 Attachment NQF_IPC_Hip_Knee_Replacement_Measure_ICD10CPTcodes.xlsx
Level	Facility	Clinician : Group/Practice
Setting	Inpatient/Hospital	Outpatient Services
Numerator Statement	The numerator is the risk-standardized proportion of patients undergoing an elective primary THA or TKA who meet or exceed an a priori, patient-defined substantial clinical benefit (SCB) threshold of improvement between preoperative and postoperative assessments on joint-specific patient-reported outcome measure (PROM) surveys. SCB improvement is defined as follows: - For THA patients, an increase of 22 points or more on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR); and - For TKA patients, an increase of 20 points or more on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR). SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, Technical Expert Panel (TEP) and Technical Advisory Group. References: Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.	The numerator is the number of respondents who have an adequate knowledge score (60% or greater) and a clear preference for surgery.
Numerator Details	This is a patient-reported outcome-based performance measure (PRO-PM). Two joint-specific Patient Reported Outcome Measure (PROM) surveys are used to collect the data for calculating the numerator: 1) the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) for THA patients, and 2) the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) for TKA patients.	The numerator is the number of respondents who have a positive decision quality assessment. The numerator is calculated based on patient responses to 6 questions from the Hip or Knee Decision Quality Instruments (these items are listed below in S.18 and included as an appendix): five multiple choice knowledge items and one preference item. One point is awarded for each correct knowledge item and then a total knowledge score is calculated and scaled from (0-100%).

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
	These PROM data and specific risk variable data will be collected 90 to 0 days prior to surgery, and PROM data will be collected again 270 to 365 days following surgery. Data elements used to define the numerator and for risk adjustment that are collected with PROM data include: - HOOS, JR or KOOS, JR - Date of Birth - Single-Item Literacy Screening (SILS2) Questionnaire - Body Mass Index (BMI) or Weight (kg) and Height (cm) - Chronic (>90 Day) Narcotic Use - Total Painful Joint Count (Patient-Reported in Non-Operative Lower Extremity Joint) - Quantified Spinal Pain (Patient-Reported Back Pain, Oswestry Index Question) - PROMIS Global Mental Health Score (calculated with data from the Patient-Reported Outcomes Measurement Information Systems (PROMIS) Global or Veteran's Rand 12-Item Health Survey (VR-12); data from VR-12 is translated to PROMIS Global Mental Health scores using a crosswalk created by Cella et. al for PROsetta® Stone) (Please note: Data elements listed above are detailed in the Data Dictionary accompanying this NQF submission; see Tabs: Risk Variables with PRO Data; HOOS, JR; KOOS, JR; PROMIS Global; VR-12) Center for Medicare and Medicaid Services (CMS) administrative data is used to identify eligible THA/ TKA procedures for the measure cohort (denominator) and additional risk variables, including patient demographics and clinical comorbidities (ICD-10 codes for eligible THA/TKA procedures identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.) The numerator is the risk-adjusted proportion of patients undergoing an elective primary THA/TKA that meet or exceed a SCB improvement on the HOOS, JR or KOOS, JR from preoperative to postoperative assessment. SCB improvement is defined as: - For THA patients, an increase of 22 points or more on the HOOS, JR - For TKA patients, an increase of 20 points or more on the KOOS, JR SCB thresholds were defined using published literature (Lyman and Lee, 2018) and vetted by our Patient Working Group, TEP, and Technical Advisory Group. Further, the measure numerator was defined with extensive patient and clinician input. Among the numerator definitions considered by stakeholders during measure development included: - Change in PROM score from preoperative to postoperative assessment reported as an average for a hospital's patients; - Postoperative PROM score reported as an average for a hospital's patients;	Respondents who score 60% or higher on knowledge and who indicate a clear preference for surgery have a positive decision quality assessment and are counted in the numerator. Those who score less than 60% and/or who are either unclear or prefer nonsurgical options have a negative decision quality assessment, and are not counted in the numerator.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
	- A threshold change in PROM score from preoperative to postoperative assessment reported as a proportion of a hospital's patients meeting the threshold; - A threshold postoperative PROM score reported as a proportion of a hospital's patients meeting the threshold; and - A combination of both a minimum threshold change in PROM score from preoperative to postoperative assessment and a minimum threshold for postoperative PROM score. Clinical experts and patients supported a numerator definition that assessed change in PROM score from preoperative to postoperative assessment over a numerator definition that focused on postoperative PROM score. TEP members and patients noted that patients want to see improvement, and that the numerator definition should reflect change following surgery. Comments against using a numerator definition focusing on the postoperative PROM score included concern that it does not reflect degree of improvement, and may incentivize surgery on patients with less severe disease who have better preoperative scores. This concern about assessment of the postoperative PROM score also led to dislike of the last option noted above, a numerator definition combining threshold change and threshold postoperative PROM score. Stakeholders also strongly supported a numerator definition assessing a threshold change in PROM score over averaging patient change in PROM scores for hospital reporting. They noted that measurement of a threshold change will highlight lower performing patients, will protect at-risk patients, and is easy to understand as a performance measure. Comments against a reported average change included concern that a hospital whose patients all achieve average results could have a reported average change result that would be very similar to a hospital whose patients achieve either very good or very poor results; an average change numerator could show similar results for hospitals with very different patient outcomes). The numerator definition of SCB threshold change, supported by patients and clinical experts, provides an easy to understand metric that patients found intuitive. Using a SCB threshold avoids the potential for misleading consumers and patients by averaging patient change scores across a hospital when individual patient outcomes within hospitals may vary considerably (as noted above). Using a SCB incentivizes providers to perform surgery on patients with worse baseline scores, a group that might otherwise not be offered surgery, as patients with poorer baseline PRO scores have more room to improve and thus a greater opportunity to achieve SCB. It also encourages providers to not perform THA/TKA procedures on patients with minimal symptoms, who will not benefit at all from surgery. And, since the SCB was defined with close input from patients and clinicians, it does set a minimum improvement threshold, but not one so large as to cause surgeons to avoid performing THA/TKA procedures on patients who would benefit. References: Cella D, Schalet BD, Kallen M, Lai JS, Cook KF, Rutsohn J, Choi SW. PROsetta® Stone Analysis Report Volume 2: A Rosetta Stone for Patient Reported Outcomes, PROMIS Global Health – Mental Component and VR-12 – Mental Component (Algorithmic Scores).	

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
	http://www.prosettastone.org/LinkingTables1/GlobalHealth/Pages/default.aspx , 2018. Lyman S and Lee YY. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res, 467(12):2432-2441.	
Denominator Statement	The cohort (target population) includes, Medicare fee-for-service (FFS) patients 65 years of age and older undergoing elective primary THA/TKA procedures, excluding patients with hip fractures, pelvic fractures and revision THAs/TKAs.	The denominator includes the number of respondents from the target population who have undergone primary knee or hip replacement surgery for treatment of knee or hip osteoarthritis.
Denominator Details	The cohort for this measure is Medicare FFS patients 65 years of age and older undergoing an elective primary THA/TKA procedure at a non-federal short-term acute care hospital. Inclusion criteria includes patients: - Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission - Discharged alive from a non-federal short-term acute care hospital - Undergoing only elective primary THA/TKA procedures (patients with fractures and revision procedures or with bone metastases are not included) - Inclusion criteria are harmonized with CMS's existing measure cohort for the hospital-level 90- day risk-standardized THA/TKA complication measure Center for Medicare and Medicaid Services (CMS) administrative data is used to identify qualifying THA/TKA procedures for the measure cohort. (ICD-10 codes for eligible THA/TKA procedures are identified in the Data Dictionary accompanying this NQF submission; see Tab ICD-10 2017-2018.)	The denominator is all adult patients who had a primary hip or knee replacement surgery for treatment of osteoarthritis and responded to the Hip or Knee Decision Quality Instrument.There is an attached sheet with ICD 10 and CPT codes needed to identify eligible patients to be surveyed for inclusion in the measure.
Exclusions	Patients with staged procedures, defined as more than one elective primary THA or TKA performed on the same patient during distinct hospitalizations during the measurement period, are excluded. All THA/TKA procedures for patients with staged procedures during the measurement period are removed.	Respondents who are missing 3 or more knowledge items do not get a total knowledge score and are excluded. Similarly, respondents who do not indicate a preferred treatment are excluded. No other exclusions as long as the respondent has the procedure for the designated condition.
Exclusion Details	Patients with staged procedures in the measure period are excluded. A staged procedure is identified if a patient has more than one hospitalization with an eligible, elective primary THA or TKA procedure during the measurement period. ICD-10 codes for eligible, elective primary THA/TKA procedures (listed in the Data Dictionary on "ICD-10 2017-2018" tab) are used to identify all eligible procedures during the measurement period; patients with an ICD-10 code for an eligible elective primary THA or TKA procedure in two or more hospital admissions during the measurement period are identified as having a staged procedure, and the patient, including all procedures, is removed from the measure cohort.	Respondents missing 3, 4, or 5 knowledge responses. Respondents missing a response to the preference item.
Risk Adjustment	Statistical risk model	No risk adjustment or risk stratification
Stratification	N/A	
Type Score	Rate/proportion better quality = higher score	Categorical, e.g., yes/no passing score defines better quality
Algorithm	Target population: Medicare FFS patients 65 years and older undergoing an elective primary THA or TKA in a non-federal short-term acute care hospital. To create the denominator:	The following steps need to be taken to calculate the measure: (1) identify eligible patients (2) administer the Hip or Knee Decision Quality Instrument (3) collect and code responses (4) calculate total knowledge scores and exclude those with 3 or more knowledge items missing (5) calculate the numerator (informed and clear preference for surgery or not) for each individual, excluding

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
	Step 1. If the patient is a Medicare FFS patient, go to Step 2. If not, do not include in the denominator. Step 2. If the patient is identified in CMS administrative claims data as having undergone an eligible elective primary THA or TKA during the measurement period, go to Step 3. If not, do not include in the denominator. Step 3. If the patient is 65 years of age or older, go to Step 4. If not, do not include in the denominator. Step 4. If the patient was enrolled in Medicare FFS Part A and Part B for the 12 months prior to index admission, and enrolled in Part A during the index admission, then go to Step 5. If not, do not include in the denominator. Step 5. If the patient was discharged alive from the hospital, include in the denominator. If not, do not include in the denominator. Step 6. If the patient experienced only one elective primary THA/TKA during the measurement period, or if the patient experience more than one elective primary THA/TKA during a singular hospitalization during the measurement period, + in the denominator. If the patient experienced two elective primary THA/TKA procedures during the measurement period performed during distinct hospitalizations, do not include in the denominator. To create the numerator: If the patient has complete PRO data collected during the prescribed preoperative and postoperative time windows, and meets or exceeds the SCB improvement threshold on the joint-specific PROM between the preoperative and postoperative assessment: - for THA patients, an increase of 22 points on the HOOS, JR - for TKA patients, an increase of 20 points on the KOOS, JR then include in the numerator. If not, then do not include in the numerator. The hospital-level measure result is calculated by aggregating all patient-level results across the hospital. For calculation of measure results, we recommend that hospitals should have a minimum case-volume of 25 elective primary THA/TKA patients with complete patient-reported outcomes and risk variable data collected 90 – 0 days preoperatively and complete patient-reported outcomes data collected 270 – 365 days postoperatively. Hospital-specific risk-standardized improvement rates (RSIRs) are calculated as the ratio of a hospital's "predicted" improvement to "expected" improvement multiplied by the overall observed improvement rate. Both predicted improvement and expected improvement are derived based on the output of a hierarchical logistic regression model that adjusts for patient case-mix and applies stabilized inverse probability weighting (IPW) to address potential non-response bias.	those with no knowledge score and/or no preference item and (6) aggregate the measure into a rate over the center or practice. Responses to five knowledge questions and one preference item from the Hip or Knee Decision Quality Instrument are needed to calculate the Informed, Patient Centered (IPC) surgery measure and are coded and scored as indicated below. Scoring of Knee Items used to generate the measure 1. Which treatment is most likely to provide relief from knee pain caused by osteoarthritis? Surgery (Coded- 1) Non-surgical treatments (coded =0) Both are about the same (coded= 0) Multiple responses = 0 Missing response = 0.33 2. After knee replacement surgery, about how many months does it take most people to get back to doing their usual activities? Less than 2 months (coded= 0) 2 to 6 months (coded = 1) 7 to 12 months (coded= 0) More than 12 months (coded= 0) Multiple responses = 0 Missing response = 0.25 3.If 100 people have knee replacement surgery, about how many will have less knee pain after the surgery? 20 (coded= 0) 40 (coded= 0) 60 (coded= 0) 80 (coded = 1) Multiple response = 0 Missing response = 0.25 4.If 100 people have knee replacement surgery, about how many will have a serious complication within 3 months after surgery? 4 (Coded=1) 10 (coded= 0) 14 (coded= 0) 20 (coded= 0) Multiple responses = 0 Missing response = 0.25 5. If 100 people have knee replacement surgery, about how many will need to have the same knee replaced again in less than 15 years? More than half (coded= 0) About half (coded= 0) Less than half (coded =1) Multiple responses = 0 Missing = 0.33 Scoring of Preference Item for Knee:

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
		6. Which treatment did you want to have to treat your knee osteoarthritis? Surgery (coded=1) Non-surgical treatments (coded= 0) Not sure (coded= 0) Multiple responses (coded=0) Scoring of Hip Items used to generate the measure: 1. Which treatment is most likely to provide relief from hip pain caused by osteoarthritis? Surgery (Coded- 1) Non-surgical treatments (coded =0) Both are about the same (coded= 0) Multiple responses = 0 Missing response = 0.33 2. After hip replacement surgery, about how many months does it take most people to get back to doing their usual activities? Less than 2 months (coded= 0) 2 to 6 months (coded = 1) 7 to 12 months (coded= 0) More than 12 months (coded= 0) Multiple responses = 0 Missing response = 0.25 3. If 100 people have hip replacement surgery, about how many will have less hip pain after the surgery? 30 (coded= 0) 50 (coded= 0) 70 (coded= 0) 90 (coded = 1) Multiple response = 0 Missing response = 0.25 4. If 100 people have hip replacement surgery, about how many will have a serious complication within 3 months after surgery? 4 (Coded=1) 10 (coded= 0) 14 (coded= 0) 20 (coded= 0) Multiple responses = 0 Missing response = 0.25 5. If 100 people have hip replacement surgery, about how many will need to have the same hip replaced again in less than 20 years? More than half (coded= 0) About half (coded= 0) Less than half (coded =1) Multiple responses = 0 Missing = 0.33 Scoring of Preference Item for Hip:

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
		6. Which treatment did you want to have to treat your hip osteoarthritis? Surgery (coded=1) Non-surgical treatments (coded= 0) Not sure (coded= 0) Multiple responses (coded=0) Knowledge: The responses are coded as indicated above. A total knowledge score is calculated by summing the five items, dividing by 5 and converting to percentage to get scores 0-100%. Missing answers are imputed with 1/k where k is the number of possible responses (essentially equivalent to guessing). Multiple responses (e.g. on paper survey) are considered incorrect and coded as 0. A total knowledge score is calculated for all surveys that have three or more knowledge items completed. Preference item: Respondents who mark surgery are considered to indicate a clear preference for surgery. Respondents that mark either non surgical treatments or not sure, are not considered to have a clear preference for surgery. Missing responses are not counted. Multiple responses (e.g. on a paper survey) are considered "not sure" and coded as 0. A positive assessment "yes" for decision quality requires a knowledge score of 60% or higher and a clear preference for surgery. Otherwise, decision quality is "no."
Submission items	5.1 Identified 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) 2958 : Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery 0422 : Functional status change for patients with Knee impairments 0423 : Functional status change for patients with Hip impairments 0424 : Functional status change for patients with Foot and Ankle impairments 0425 : Functional Status Change for Patients with Low Back Impairments 0426 : Functional status change for patients with Shoulder impairments 0427 : Functional status change for patients with elbow, wrist and hand impairments 0428 : Functional status change for patients with General orthopaedic impairments 2643 : Average change in functional status following lumbar spine fusion surgery 5a.1 Are specs completely harmonized? Yes 5a.2 If not completely harmonized, identify difference, rationale, impact: To the extent feasible, we have harmonized with existing, related measures. However, we have prioritized the goal of the measure to assess substantial clinical benefit (SCB) improvement in patient-reported outcomes for elective primary THA/TKA patients with minimal patient and provider burden over harmonization if discrepancies occur.	5.1 Identified measures: 5a.1 Are specs completely harmonized? 5a.2 If not completely harmonized, identify difference, rationale, impact: Not applicable. 5b.1 If competing, why superior or rationale for additive value: Not applicable.

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
	5b.1 If competing, why superior or rationale for additive value: NQF # 2653: Average change in functional status following total knee replacement surgery. This PRO-PM measure differs from NQF #2653 in attribution, cohort, outcome, and risk adjustment. Attribution: This PRO-PM is a hospital-level quality measure, whereas NQF #2653 is a clinician-level measure. Cohort: This PRO-PM includes both THA and TKA procedures, as clinical experts agree that hospital-level processes are shared across these procedures, and includes only primary, not revision, procedures, based upon clinical input that revision procedures are more complicated to perform and patient-reported outcomes may be influenced by the initial surgery. The target population is Medicare FFS beneficiaries 65 years of age and older. NQF #2653 includes only TKA procedures, includes knee replacement revisions as well as primary procedures, and includes all adults 18 years of age and older. Outcome: This PRO-PM collects PROs with the HOOS, JR for THA patients and the KOOS, JR for TKA patients. Timing of PRO data collection is 90 – 0 days prior to and 270 – 365 days following surgery. The numerator measures SCB improvement for each patient from preoperative to postoperative assessment with a binary outcome (Yes/No), and the measure produces a risk-standardized improvement rate that elucidates for hospitals the risk-adjusted proportion of patients with improvement and those without improvement. In contrast, NQF #2653 collects PRO data with the Oxford Knee Score three months prior to and 9 – 15 months following surgery, and measures average change in knee function score. The outcome definition of SCB, with a defined threshold for change in PROM score, allows patients with poorer baseline PRO scores more room to improve and thus a greater opportunity to achieve SCB. This was identified by our TEP members as a specific benefit of measuring SCB versus average change; measuring SCB incentivizes providers to offer and perform THA/TKA procedures on even those with poor PRO scores. Further stated TEP and Patient Working Group concerns with measuring an average change score included the fact that hospitals with all average outcomes would look similar to hospitals whose patients either did very well or very poorly (bimodal distributed outcomes), thus providing potentially misleading information to consumers and patients. Risk Adjustment: This risk model for this PRO-PM includes important risk variables supported by technical expert panel (TEP) and other expert clinical consultants including health literacy, other musculoskeletal pain and chronic narcotic use which are not included in NQF #2653; these risk variables were identified and tested based upon input from orthopedic professional societies, including AAHKS and AAOS, through public comment (Centers for Medicare & Medicaid Services , CJR Final Rule 2015, Section III.D.3.A). This PRO-PM is superior to NQF #2653: 1) it more appropriately provides a signal of hospital quality which reflects outcomes for both THA and TKA recipients since within hospitals, care for patients undergoing THA/TKA procedures is provided by the same providers and hospital staff; 2) it assesses SCB improvement with a binary outcome that elucidates for hospitals and patients the risk-adjusted proportion of patients with and without improvement (a clear, understandable metric that patients support); 3) it uses a more robust and stakeholder-driven risk model, anticipated to produce a measure with greater face validity with stakeholders; and 4) it is harmonized with related measures including NQF #1550 Hospital-level risk-standardized complication	

	3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	2958: Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery
	rate (RSCR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA) and 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 (MUC19-28). References: Comprehensive Care for Joint Replacement (CJR) Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services Final Rule, 80 C.F.R. 73273 (Nov 24, 2015).	

Appendix F: Pre-Evaluation Comments

Comments received as of June 12, 2020

Topic	Commenter	Comment
3559: Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	Submitted by The Federation of American Hospitals (FAH)	The Federation of American Hospitals (FAH) appreciates the opportunity to comment on measure #3559, Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA), prior to the Standing Committee's evaluation. The FAH supports the development and implementation of patient-reported outcomes performance measures (PRO-PMs), but we also believe that additional questions and work remain before their widespread use. For instance, the degree to which multiple PRO-PMs could lead to survey fatigue for patients, the potential impact additional PRO-PMs may have on the reporting of well-established measures such as HCAHPs, and what level of data collection burden for an individual PRO-PM is acceptable for a hospital or other healthcare provider. Specifically, on review of the measure specifications, the FAH notes that multiple data points beyond the typical clinical variables are required to ensure that the measure results are adequately risk adjusted. The FAH supports the inclusion of these data points, but we are concerned that the developer has not provided sufficient information on how these data are collected and what additional workload and time will be required. For example, several of the data elements needed for risk adjustment are derived from patient-reported surveys, which must be collected within 0-90 days pre-operative. No information was provided on the processes used by the hospitals such as whether it required coordination with orthopedic practices or if the burden of the additional data collection was placed on hospital staff on the day of surgery. To what extent did these requirements impact clinical workflows and were additional staff resources required? What additional costs might an individual hospital encounter as a result of implementation of this PRO-PM? Alternatively, from the patient's perspective, did the additional questions seem relevant and was the point in time during which these additional data were collected appropriate? It would also be useful to understand whether there is a potential for individuals

Topic	Commenter	Comment
		to prioritize the completion of one survey over another and therefore lead to negative unintended consequences on response rates for other PRO-PMs such as HCAHPS? The FAH believes that these questions should have been addressed during the development of this PRO-PM and this detail should have been provided within the measure submission rather than the generalized statements that we see in the responses under the feasibility criterion. In addition, while the FAH strongly supports the inclusion of health literacy in the risk adjustment model, we believe that the risk adjustment approach used by many developers considers the identification and testing of social risk factors as supplementary to clinical risk factors. This approach was identified as a concern by the NQF Disparities Standing Committee. Given that this was a new measure, it provided an opportunity for the measure developer to include these factors within the testing of the model rather than the previous approach of "adding on" factors after the model is developed. This type of approach would assist hospitals and others in understanding how their inclusion could impact the model and provide additional information for groups examining this issue such as the NQF and Office of the Assistant Secretary for Planning and Evaluation. As a result, the FAH believes that this measure lacks sufficient information on the potential impact these social risk variables have on the risk adjustment model. The FAH requests that the Standing Committee consider these important issues during the evaluation of this measure. Thank you for the opportunity to comment.

National Quality Forum
1099 14th Street NW, Suite 500
Washington, DC 20005
http://www.qualityforum.org