Cost and Efficiency, Spring 2019 Review Cycle: CDP Report

TECHNICAL REPORT

February 21, 2020

NATIONAL QUALITY FORUM

This report is funded by the Department of Health and Human Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001.

Contents

Cost and Efficiency, Spring 2019 Review Cycle

TECHNICAL REPORT

Executive Summary

It is estimated that healthcare spending in the United States will increase the health share of GDP from 17.9 percent in 2017 to 19.4 percent by 2027.[1] This level of healthcare spending and growth has the potential to increase federal deficits and debt further, or crowd out spending for other important national priorities.[2] These economic realities require performance measures that can accurately capture spending, particularly spending that results from inefficient or poor-quality care.

Reducing wasteful spending requires the coordination of multiple providers and care settings to ensure efficient, high-quality patient transitions. Thus, cost and quality measures used together can help to assess efficiency and value of care delivered and drive improvement in the U.S. healthcare system.

The Cost and Efficiency Standing Committee oversees NQF's portfolio of seven cost and efficiency measures. For Spring 2019 cycle, the Standing Committee evaluated three newly submitted measures against NQF's cost and resource use evaluation criteria. The Standing Committee recommended 3509 *Routine Cataract Removal with Intraocular Lens (IOL) Implantation,* 3510 *Screening/Surveillance Colonoscopy,* and 3512 *Knee Arthroplasty* for endorsement. During the Standing Committee's deliberations, issues regarding the opportunity for impromvent and the adjustment for social factors emerged. The Standing Committee underscored the importance of better undstanding the source of performance variation, noting that variation in cost for these measures is narrow and is likely due to complications. Additionally, the Committee encouraged developers to examine the complex role of social risk factors and their role in the clinical episodes measured. These issues were factored into the Committee's evaluation and recommendations for all three measures.

Brief summaries of the measures are included in the body of the report; detailed summaries of the Committee's discussion and ratings of the criteria for the measures can be found in Appendix A.

Introduction

Healthcare spending in the United States is projected to grow 4.8 percent in 2019, reaching $2.8 trillion dollars.[1] Forecasts from 2018 to 2027 estimate that healthcare spending will outpace gross domestic product (GDP) growth by 0.8 percent. This increase will raise the health share of GDP from 17.9 percent in 2017 to 19.4 percent by 2027.[1] These concerning trends can be attributed to many causes, including high costs for drugs, procedures, and administrative services, as well as poor coordination and overutilization of unnecessary health services. This level of healthcare spending and growth has the potential to increase federal deficits and debt further, or crowd out spending for other important national priorities.[2] Given this trend, healthcare cost measurement continues to be a critical component to assess and improve the efficiency of the US healthcare system.

Improving U.S. health system efficiency has the potential to reduce cost growth and improve the quality of care provided, simultaneously. Cost measures are the building blocks to efficiency and value. When NQF launched its first effort to endorse cost and resource use measures in 2009, one of the foundational principles was that cost and resource use measures should be used in the context of and reported with quality measures. NQF, with the guidance and support of the Cost and Efficiency Standing Committee, continues to explore approaches and best practices for evaluating efficiency constructs.

As part of NQF's redesign of the Consensus Development Process in 2017, the Cost and Resource Use Standing Committee expanded its charge to assess efficiency more broadly, including measures assessing the efficiency of healthcare delivery. The Cost and Resource Use Standing Committee was renamed the Cost and Efficiency Standing Committee. While there are currently no efficiency measures in the portfolio, the new scope allows the Committee to take a more holistic view of drivers of healthcare spending and identify sources of inefficiency and waste across the system.

This cycle, the Cost and Efficiency Committee's evaluation was informed by inputs from the NQF Scientific Methods Panel (SMP), several NQF clinically-focused Technical Expert Panels (TEPs), as well as stakeholder comments. A total of eight measures were submitted to the project for consideration and all were reviewed by the SMP. However, only three of those eight measures were ultimately passed on to the committee for consideration and voting due to non-passing votes for reliability and/or validity by the SMP (see Table 1). The Committee reviewed three cost measures: 3509 *Routine Cataract Removal with Intraocular Lens (IOL) Implantation,* 3510 *Screening/Surveillance Colonoscopy,* and 3512 *Knee Arthroplasty*; the Committee recommended all three measures for endorsement.

Table 1. Cost and Efficiency Measures Submitted to Standing Committee for Evaluation after SMP Review

	Maintenance	New	Total
Measures under endorsement consideration	0	3	3
Endorsed Measures	0	3	3
Measures not endorsed	0	0	0

NQF Portfolio of Performance Measures for Cost and Efficiency

The Cost and Efficiency Standing Committee (see Appendix C) oversees NQF's portfolio of cost and efficiency measures (see Appendix B). This portfolio contains seven cost and efficiency measures (see Table 2).

Table 2. NQF Cost and Efficiency Portfolio of Measures

NQF #	Title	Category
1598	Total Resource Use Population-Based PMPM Index	Noncondition-specific per capita resource use measure
1604	Total Cost of Care Population-Based PMPM Index	Noncondition-specific per capita cost measure
2431	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Acute Myocardial Infarction (AMI)	Condition-specific, episode-based cost measure
2436	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Heart Failure	Condition-specific, episode-based cost measure
2579	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Pneumonia	Condition-specific, episode-based cost measure
2158	Medicare Spending Per Beneficiary	Noncondition-specific, episode-based cost measure
3474	Hospital-Level, Risk-Standardized Payment Associated with a 90-Day Episode of Care for Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA).	Condition-specific, episode-based cost measure

Cost and Efficiency Measure Evaluation

After review by the NQF Scientific Methods Panel, three NQF-convened, clinically-focused, Technical Expert Panels (TEP) met between April 8 and 10, 2019 to review the clinical aspects of the measure specifications. The NQF TEPs were charged with providing the Cost and Efficiency Standing Committee a qualitative assessment of each measure's clinical specifications. On June 27, 2019, the Cost and Efficiency Standing Committee considered all of the inputs and evaluated the measures against NQF's Cost and Resource Use Evaluation Criteria.

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 8, 2019 and closed on June 19, 2019. As of June 19, one comment was submitted and shared with the Committee prior to the measure evaluation meetings (Appendix F). The comment was provided to the Committee prior to its deliberations at the measure evaluation meeting.

Comments Received After Committee Evaluation

The continuous 16-week public commenting period with NQF member support closed on September 6, 2019. Following the Committee's evaluation of the measures under consideration, NQF received six comments pertaining to the draft report and to the measures under consideration. All comments for 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 member provided their expression of support for measure *3510 Screening/Surveillance Colonoscopy*.

Overarching Issues

During the Standing Committee's discussion of the measures, several overarching issues emerged that were factored into the Committee's ratings and recommendations for all three measures and are not repeated in detail with each individual measure.

Opportunity for Improvement

Measures submitted included a range of cost performance between the accountable units. However, the Standing Committee noted across the measures that the interquartile range of the performance was narrow. The procedures captured by these measures are fairly routinized and the variation in cost is typically due to complications. The measure developers argued that while the variation in cost among procedures is narrow, given the frequency that these common procedures take place, there is an opportunity to reduce costs in aggregate by reducing complications and costs across all episodes. The Standing Committee noted that it is critical to understand the source of performance variation, especially given the potentially small gains that can be made within an individual episode, and how the accountable entity can meaningfully influence that variation. The measure steward and implementers should monitor for unintended negative consequences to patients by potentially reducing utilization of necessary healthcare services to achieve improvements.

Adjustment for Social Factors

Cost/resource measures are influenced by both the care received in a healthcare setting and patient social risk factors, since they typically measure the cost or resource use over multiple providers, settings, and across time. While the developer did analysis of several social factors for the risk adjustment model, none were included in the three measures under review. The Standing Committee encouraged developers to understand and examine the complex role of social risk factors and their role in the clinical episodes measured. The Committee noted the need to ensure that providers serving people with social risk factors are not penalized unfairly. While the Committee noted that it is important to maximize the predictive value of a risk-adjustment model, understanding the role that social risk factors play in a clinical cost episode is critical. The impact of social risk factors in cost and efficiency measures is unique in that these factors may ultimately increase overall costs through poor transitions and hand-offs, or potentially lower resource use because of access-to-care challenges. Each cost and efficiency measure should be examined on a case-by-case basis to understand the role of patient social risk factors in the measure.

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.

3509 Routine Cataract Removal with Intraocular Lens (IOL) Implantation

Description: The Routine Cataract Removal with Intraocular Lens (IOL) Implantation cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 60 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Routine Cataract Removal with IOL Implantation measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period. **Measure Type:** Cost/Resource Use; **Level of Analysis:** Clinician: Group/Practice, Clinician: Individual; **Data Source:** Claims, Enrollment Data, Other

This measure calculates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive routine cataract removal with intraocular lens (IOL) implantation. The developer provided data demonstrating that routine cataract surgery is the most common surgical procedure in the United States with a range of cost performance at the clinician group and the clinician levels. The Committee agreed that it was important to measure but noted that the there was little variation because of the narrow interquartile range. The Committee noted that much of the variation appears to be driven by complications from the procedure, and several Committee members noted that a quality measure of complications might be a more direct method of assessing provider performance. The developer explained that although the interquartile difference appears small, there is a high volume of these procedures and savings across all episodes can add up to a large cost reduction.

The Scientific Methods Panel (SMP) reviewed the scientific acceptability of the measure. The SMP passed the measure on the reliability criteria but did not reach consensus on validity. Overall, the Committee agreed that the reliability testing was appropriate and voted to uphold the SMP rating of moderate. For the evaluation of validity, the Committee members and SMP questioned the appropriateness of the exclusions. The developer replied that the exclusions are consistent with two other NQF-endorsed measures related to cataract surgery outcomes and that they wanted consistency among measures to better align cost to quality. The Committee discussed the hierarchical condition categories (HCC) risk-adjustment model that the developer used, and why variables specific to the procedure were not included. Weighing all of the validity subcriteria, the Committee ultimately passed the measure on validity.

The Committee did not have any concerns on the feasibility. Several Committee members stated that they were unsure how the usability of the measure could enable physicians to drive down costs. The developer responded that physicians conduct a high volume of these procedures and in aggregate the

savings will increase. Overall, the Committee did not have any major concerns on the use or usability. The Standing Committee recommended this measure for NQF endorsement.

3510 Screening/Surveillance Colonoscopy

Description: The Screening/Surveillance Colonoscopy cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care from the day of the clinical event that opens or 'triggers' the episode, through 14 days after the trigger. Beneficiary populations eligible for the Screening/Surveillance Colonoscopy measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period. **Measure Type:** Cost/Resource Use; **Level of Analysis**: Clinician: Group/Practice, Clinician: Individual; **Data Source**: Claims, Enrollment Data, Other

This cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive a screening/surveillance colonoscopy. The developer provided data demonstrating that routine screening/surveillance colonoscopy has a range of cost performance at the clinician group and the clinician levels. The developer also provided citations demonstrating that poor bowel preparation increases the potential for missed lesions, canceled procedures, adverse events, and ultimately higher episode costs. The Committee agreed that there is an opportunity for improvement to decrease costs associated with screening/surveillance colonoscopy. Several Committee members did express concern about the narrow interquartile range.

The SMP evaluated the scientific acceptability criteria and voted to pass on reliability but did not come to consensus on the validity testing. The Committee agreed that the reliability testing scores were high and voted to uphold the SMP rating of high. With regards to validity, the Committee encouraged the developer to continue to test social risk factors. The Committee also encouraged the developer to test within and between clinician differences in performance to further understand the role of social risk, particularly for those that have a disproportionate share of high-risk patients. There were no major concerns on the feasibility of the measure. The Committee expressed similar concerns as with 3509 regarding the usability of the measure. The Standing Committee recommended this measure for NQF endorsement.

3512 Knee Arthroplasty

Description: The Knee Arthroplasty cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 30 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Knee Arthroplasty measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period. **Measure Type:** Cost/Resource Use; **Level of Analysis**: Clinician: Group/Practice, Clinician: Individual; **Data Source**: Claims, Enrollment Data, Other

The Standing Committee agreed that this measure addresses a significant, high-volume, and high-resource use area of measurement. The SMP evaluated the scientific acceptability criteria and voted to pass on reliability but did not come to consensus on the validity testing. The reliability assessment was conducted using a test-retest approach with Pearson correlations of 0.8 for the group level and 0.75 for the clinician level. The Committee noted that this was lower than the previous two measures discussed but demonstrated moderate reliability. For validity, the Committee discussed significance of the dual status variable in the bivariate testing but recognized the high correlation between the dual-adjusted and non-dual adjusted model. In addition, several Committee members expressed concern that the risk-adjustment model may not account for appropriate differences in post-acute care services. The developer's rationale was that post-acute care should be seen as a free-floating source of variation in the measure. The Committee encouraged the developer to consider concerns about unintended consequences of sending patients home too soon, without appropriate post-acute care. The developer noted that this measure can be used with related readmissions measures within the post-acute care setting to monitor for unintended consequences. The Standing Committee ultimately recommended the measure for NQF endorsement.

References

1 Centers for Medicare & Medicaid Services. National Health Expenditure Data Fact Sheet. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/index.html. Published April 17, 2018. Last accessed July 2019.

2 Medicaid and Medicare Services. Context for Medicare Payment Policy. http://www.medpac.gov/docs/default-source/reports/mar19_medpac_ch1_sec.pdf?sfvrsn=0. Published March 2019. Last accessed July 2019.

Appendix A: Details of Measure Evaluation

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

Endorsed Measures

3509 Routine Cataract Removal with Intraocular Lens (IOL) Implantation

Submission | Specifications

Description: The Routine Cataract Removal with Intraocular Lens (IOL) Implantation cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 60 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Routine Cataract Removal with IOL Implantation measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

Numerator Statement: The sum of the ratio of observed to expected payment-standardized cost to Medicare for all Routine Cataract Removal with IOL Implantation episodes attributed to a clinician or clinician group. Expected costs refer to costs predicted by the risk adjustment model. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

Denominator Statement: The total number of episodes from the Routine Cataract Removal with IOL Implantation episode group attributed to a clinician or clinician group within a performance period (i.e., MIPS performance year).

Exclusions:

- The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day.
- No attributed clinician is found for the episode.
- The beneficiary's date of birth is missing.
- The beneficiary's death date occurred before the trigger date.
- The beneficiary's death date occurred before the episode ended.
- The beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window.
- The episode trigger claim was not performed in an outpatient hospital or ASC setting.
- Episodes where the beneficiary has ocular comorbidities (impacting visual outcome of surgery or surgical complication rate).
- Episodes classified as outlier cases.

Adjustment/Stratification: Stratification by risk adjustment

Level of Analysis: Clinician : Group/Practice, Clinician : Individual

Setting of Care:

Type of Measure: Cost/Resource Use

Data Source: Claims, Enrollment Data, Other

Measure Steward: Centers for Medicare & Medicaid Services

STANDING COMMITTEE MEETING 06/27/2019

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

(1a. High Impact or High Resource Use, 1b. Opportunity for Improvement)

Importance to Measure and Report: **H-2; M-7; L-1; I-1**

Rationale:

- The developer provided data demonstrating that routine cataract surgery is the most common surgical procedure in the United States with a range of cost performance at the taxpayer idenfication number (TIN) and the TIN-national provider identifier (NPI) level. Specifically, the interquartile range of performance for TIN level scores is $238, and mean performance of $3,041. The interquartile range of performance for TIN-NPI is $232, and mean performance of $3,038.
- The developer provided citations demonstrating that complications from cataract surgery, specifically endophthalmitis after surgery, resulted in an 83% greater episode costs.
- The Committee agreed that it was important to measure but noted that the there was little variation because the interquartile range was very small. Several Committee members also questioned whether an individual physician or physician group would be able to influence the outcome of this measure given the low incidence of complications that drive the variation in the measure. These Committee members questioned whether a direct quality measure of complications would be a more appropriate method to evaluate provider performance.
- The developer responded that although the interquartile difference appears small, there is a high volume of these procedures and savings across all episodes can be significant.

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: SMP: **H-3; M-3; L-0; I-0;** Standing Committee: **Yes-10; No-0**; 2b. Validity: **H-2; M-9; L-1; I-0**

Rationale:

Reliability

- Reviewed by the Scientific Methods Panel who passed the measure on the reliability criteria.
- Data element testing was conducted via CMS auditing programs for Parts A &B Claims data. The demonstration of data element validity did not meet NQF standards (i.e., description of CMS audits, fraud detection efforts).
- There were 490,714 Medicare beneficiaries included in the TIN level testing analysis and 485,216 beneficiaries included in the TIN-NPI level measure testing.
- Measure score reliability testing included test-retest with correlations, and signal to noise.
- The test-retest correlation coefficients were 0.93 (clinician groups) and 0.92 (clinicians)
- Relatively stable movement across quintiles.
- Signal to Noise used the Adams' method with the mean reliability scores 0.95 for groups (TIN) and 0.94 for clinicians (TIN-NPI).
- Overall, the Committee agreed that the reliability testing was appropriate and voted to uphold the SMP rating of moderate.

Validity

- Reviewed by the NQF Scientific Methods Panel who did not reach consensus on validity.

- The developer used a clinical subcommittee, a TEP, a person and family committee, and a national stakeholder feedback survey to provide input on measure and cost components attributable to this procedure episode of care measure.
- The face validity testing information provided by the developer does not meet NQF validity testing requirements since NQF requires that an expert group has been convened and a systematic assessment of the measure score has been conducted.
- Due to the inadequacy of face validity, the NQF SMP focused its evaluation on the empirical validity testing.
- Empirical validity was assessed by examining correlation with other known indicators of resource utilization in administrative claims data, specifically complications related to cataract removal. Correlation analysis showed expected correlation of higher cost and complications.
- The mean observed to expected cost for episodes with services related to complications during the post-trigger period is 1.04, compared to 0.95 for episodes without services related to complications during the post-trigger period.
    - Some members of the SMP expressed concern about the approach to empirical validity testing. Specifically, there was concern that the measure construct which relies on administrative claims was compared to another measure with the same underlying data elements – which were also generated using administrative claims and used in the performance measure score. As such, the SMP members were concerned the method used by the developer did not represent correlation to an independent variable or measure.
    - Other SMP members, noted that this approach was sufficient.
- Risk adjustment: The developer assessed potential disparities by analyzing social risk factors of gender, dual status, income, education and unemployment.
    - The developer tested the impact of including social risk factors using T-tests and F-tests of variable coefficients and p-values, testing with step-wise regression models, and testing the final models with and without social risk factors. The developer noted that testing demonstrated significance of the social factors, but inconsistent direction of the social risk factors and high correction between the measure scores with and without the social risk factors.
    - No social factors were ultimately included in risk adjustment based on results of empirical analysis.
    - The Committee discussed the HCC risk-adjustment model that the developer used. The Committee questioned why the developer did not include variables specific to the procedure.
- TEP Feedback: Generally agreed that the clinical population was appropriate but expressed some concerns and sought clarity on the rationale for some of the specifications and decision logic. Exclusions eliminated a large number of patients (~40%) but left a significant number of patients to implement measure. Developer clarified that after exclusions applied, 87.1 TIN, 83.7 TIN-PINs met case minimum of 10 after exclusions applied.
- Meaningful Differences: The developer stratified the clinician measure scores by meaningful characteristics and investigated the clinician score distribution by percentile. Characteristics included: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician. Large performance difference among clinicians but clinicians with more episodes perform similarly to those who perform fewer procedures.
- Committee members and SMP questioned the appropriateness of the exclusions. The developer replied that the exclusions are consistent with two other NQF-endorsed measures related to

cataract surgery outcomes and that they wanted consistency among measures to better align cost to quality.
- Weighing all of the validity sub criteria, the Committee ultimately passed the measure on validity.

---

3. Feasibility: H-10; M-2; L-0; 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:

- Claims-based measure.
- The Committee did not have any concerns on the feasibility.

---

4. Use and Usability

4a. Use; 4a1. Accountability and transparency; 4a2. Feedback on the measure by those being measured and others; 4b. Usability; 4b1. Improvement; 4b2. The benefits to patients outweigh evidence of unintended negative consequences to patients; 4b3.Measure con be deconstructed to facilitate transparency and understanding)

4a. Use: **Pass-12; No Pass-0** *4b. Usability:* **H-2; M-7; L-2; I-1**

Rationale:

- This is a new measure and is not currently publicly reported. The measure is used in a quality payment program, Merit-based Incentive Payment System (MIPS). While the measure has been implemented into the MIPS program, the measure results are first scheduled to be calculated for performance year 2019 (payment year 2021).
- Several Committee members stated that they were unsure how the usability of the measure could allow physicians to drive down costs. The developer responded that physicians conduct a high volume of these procedures and in aggregate, the savings will increase.

---

5. Related and Competing Measures

- No related or competing measures noted.

---

6. Standing Committee Recommendation for Endorsement: **Yes-11; No-1**

---

7. Public and Member Comment

A commenter raised concerns around the measure's inclusion of drugs with pass-through status because it may serve as a disincentive to surgeons using innovative treatments in cataract surgery. Further, the commenter disagrees with the inclusion of drugs used to treat post-operative conditions after cataract surgery since this treatment is not part of the cataract surgery itself.

COMMITTEE RESPONSE:

Thank you for your comment. While the Committee understands the commenters' concerns with the inclusion of pass through status and post-operative drugs in the measure, it also

believes that episode-based cost measurement should include all relevant costs in the episode, including drugs used during the episode that are clinically related to the procedure and treatment of the condition during the episode. Therefore, it is appropriate to include pass through or other drugs used during the episode of care based on the definition of the episode and the intent of the measure. The addition or inclusion of any drug to the measure specifications will be considered on a case-by-case basis by the measure developers and the Committee through the maintenance review process. For example, drugs will be considered for inclusion as new drugs become available and alternatives to current treatments are introduced to the market and clinical practice.

MEASURE STEWARD/DEVELOPER RESPONSE:

We appreciate the input from Eyepoint Pharmaceuticals. To provide some context and to recap the discussions from the Standing Committee meeting, we consider services for potential assignment to the measure by gathering expert input on their clinical relatedness to the procedure or condition that the measure focuses on. For drugs on pass-through status, we assess these on a case-by-case basis to ensure that this measure is capturing clinically relevant services. In this case, the expert clinician committee that we convened had considered that an intraoperative drug on pass-through status was appropriate to include as they considered it to be an important source of cost variation. We continue to monitor the inclusion of this intraoperative drug – as we do with other assigned services – as part of ongoing measure maintenance and will make updates to the measure specifications through the maintenance process where appropriate.

Part D drugs, including eye drops, were not included in the cost measure due to the lack of standardized cost data for Part D claims, as the expert committee believed it would not be appropriate to include non-standardized costs in the measure. An additional challenge is that not every Medicare beneficiary is enrolled in Part D. The expert committee discussed the implications of not including Part D data at this point in time and specifically noted that certain intraoperative Part B drugs, which may be substitutes for eye drops, should nonetheless be included in the measure due to the high cost of those drugs. Once Part D standardized costs are available, we will review the inclusion of Part D costs including eye drops.

---

8. Consensus Standards Approval Committee (CSAC) Vote (November 12, 2019): Y-10; N-0

CSAC Decision: Approved for endorsement

---

9. Appeals

No appeals were received

3510 Screening/Surveillance Colonoscopy

Submission | Specifications

Description: The Screening/Surveillance Colonoscopy cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care from the day of the clinical event that opens or 'triggers' the episode, through 14 days after the trigger. Beneficiary populations eligible for the Screening/Surveillance Colonoscopy measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

Numerator Statement: The sum of the ratio of observed to expected payment-standardized cost to Medicare for all Screening/Surveillance Colonoscopy episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

Denominator Statement: The total number of episodes from the Screening/Surveillance Colonoscopy episode group attributed to a clinician within a performance period (i.e., MIPS performance year).

Exclusions:

- The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day.
- No attributed clinician is found for the episode.
- The beneficiary's date of birth is missing.
- The beneficiary's death date occurred before the trigger date.
- The beneficiary's death date occurred before the episode ended.
- The beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window.
- The episode trigger claim was not performed in an outpatient hospital, office or Ambulatory Surgical Center (ASC) setting.
- The trigger event includes endoscopic mucosal resection.
- The trigger event includes upper GI endoscopy.
- The patient has a history of inflammatory bowel disease.
- Episodes classified as outlier cases.

Adjustment/Stratification: Stratification by risk adjustment

Level of Analysis: Clinician : Group/Practice, Clinician : Individual

Setting of Care:

Type of Measure: Cost/Resource Use

Data Source: Claims, Enrollment Data, Other

Measure Steward: Centers for Medicare & Medicaid Services

STANDING COMMITTEE MEETING 06/27/2019

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

(1a. High Impact or High Resource Use, 1b. Opportunity for Improvement)

Importance to Measure and Report: **H-3; M-6; L-3; I-0**

Rationale:

- The developer provided data demonstrating that routine screening/surveillance colonoscopy has a range of cost performance at the TIN and the TIN NPI level. Specifically, the interquartile

range of performance for TIN level scores is $176, and mean performance of $936. The interquartile range of performance for TIN-NPI is $173, and mean performance of $979.

- The developer also provided citations demonstrating that poor bowel preparation increases the potential for missed lesions, canceled procedures, adverse events, and higher episode costs.
- The Committee agreed that there is an opportunity for improvement to decrease costs associated with screening/surveillance colonoscopy. Several Committee members did express concern about the small interquartile range and the relationship between the variation in performance and its relationship to poor quality. The developer replied that overutilization of services such as anesthesia or inadequate bowel prep that requires a repeated colonoscopy drive up costs.

---

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: H-4; M-2; L-0; I-0**; Standing Committee: **Yes-12; No-0** 2b. Validity: **H-0; M-11 L-1; I-0**

<u>Rationale</u>:

Reliability

- Reviewed by the Scientific Methods Panel who passed the measure on the reliability criteria.
- Data Element: Data element testing conducted via CMS auditing programs for Parts A &B Claims data. The developer did not provide information on confirmation of the procedure and diagnosis code. The demonstration of data element validity did not meet NQF standards (i.e., description of CMS audits, fraud detection efforts)
- Score-Level: Measure score reliability testing included test-retest with correlations, and signal to noise. The test-retest results found a Pearson correlation of 0.93 at the group level and 0.88 correlation at the clinician level. Relatively stable movement across quintiles. The signal to noise analyses relied on the Adams' method (ratio of between variance to total variance). Mean reliability scores were 0.96 for clinician groups (TIN) and 0.93 for clinicians (TIN) again indicating high reliability based on signal to noise test.
- The Committee agreed that the reliability testing scores were high and voted to uphold the SMP rating of high.

Validity

- Reviewed by the Scientific Methods Panel who did not reach consensus on validity.
- The developer used a clinical subcommittee, a TEP, a person and family committee, and a national stakeholder feedback survey to provide input on measure and cost components attributable to this procedure episode of care measure. The face validity testing information provided by the developer does not meet NQF validity testing requirement requirements since NQF requires that an expert group has been convened and a systematic assessment of the measure score has been conducted.
- Due to the inadequacy of face validity, the SMP focused its evaluation on the empirical validity testing.
- Empirical validity was assessed by examining correlation with other known indicators of resource utilization in administrative claims data, specifically emergency room (ER) visits or complications related to the colonoscopy.
- Correlation analysis showed expected correlation of higher cost and complications. The mean observed to expected cost is 1.49 for episodes with an ER visit and 1.0 for episodes with no ER

visit. The mean observed to expected cost for episodes with services indicating services related to a complication is 1.33 compared to 1.0 for episodes that do not indicate such services in the post trigger period.

- The NQF SMP encouraged the Cost and Efficiency Standing Committee to consider the empirical testing conducted by the developer to determine if it is adequate to meet the NQF endorsement criteria.
- Risk adjustment: The risk adjustment model was based on the CMS-HCC risk adjustment model. The developer's clinical subcommittee considered other factors for inclusion in addition to those in the HCC model to ensure it was clinically appropriate. The R-squared is modest at 0.12. The developer noted that while individual bivariate testing demonstrated significance of the social factors, the inconsistent direction of the social risk factors and high correlation between the measure scores with and without the social risk factors indicated that the final model sufficiently accounts for the effects of social risk factors on clinician measure scores. No social factors included in risk adjustment based on results of empirical analysis.
- TEP Feedback: Generally agreed that the clinical population was appropriate.
- Meaningful Differences: The developer assessed meaningful differences by stratifying the clinician measure scores by meaningful characteristics and investigating the clinician score distribution by percentile. Characteristics included: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician. Large performance difference among clinicians for the measure: The measure score at the 99th percentile is approximately 96 percent greater than the score at the 1st percentile at the TIN level, and more than 92 percent greater at the TIN-NPI level. The mean Colonoscopy score for hospital outpatient department (HOPD) sub-group is 30-40 percent higher than for Ambulatory Surgical Center (ASC) and Office sub-groups at both the TIN and TIN-NPI levels.
- The Committee encouraged the developer to continue to test social risk factors. The Committee also encouraged the developer to test within and between clinician differences in performance to further understand the role of social risk.

---

3. Feasibility: H-8; M-4; L-0; 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:

- Claims-based measure.
- The Committee did not have any concerns on the feasibility.

---

4. Use and Usability

4a. Use; 4a1. Accountability and transparency; 4a2. Feedback on the measure by those being measured and others; 4b. Usability; 4b1. Improvement; 4b2. The benefits to patients outweigh evidence of unintended negative consequences to patients; 4b3.Measure con be deconstructed to facilitate transparency and understanding)

4a. Use: **Pass-11; No Pass**-1 *4b. Usability:* **H-2; M-8; L-2; I-0**

Rationale:

- This is a new measure and is not currently publicly reported. The measure is used in a quality payment program, Merit-based Incentive Payment System (MIPS). While the measure has

technically been implemented into the MIPS program, the measure results are first scheduled to be calculated for performance year 2019 (payment year 2021).
- The Committee expressed similar concerns as with 3509 with regards to the usability of the measure.

---

5. Related and Competing Measures

- No related or competing measures noted.

---

6. Standing Committee Recommendation for Endorsement: **Y-10**; **N-2**

---

7. Public and Member Comment

Validity Testing. Commenters expressed concerns with various aspects of the validity of the measure. One commenter sought clarity on Committee's response to the Method Panel's concerns with validity testing approach. Concerns with the low r-squared result for the risk adjustment model testing and the developer's decision not to include social risk factors in the model were also raised; other commenters sought clarity from the developers on their specified approach to risk adjustment. Another commenter questioned whether the measure developer demonstrated meaningful differences in performance and highlighted the implications of program design and benchmarking on distinguishing differences in performance on the measure.

Measure Intent. One commenter expressed concerns with the intent of the measure and its alignment with the measure specifications future improvements as it relates to capturing bowel prep issues. The commenter emphasized that the intent of the measure should be to capture costs related to a colonoscopy episode, and not the efficacy of bowel prep.

COMMITTEE RESPONSE:

Thank you for your comments. The Committee appreciates the concerns raised by the commenters and considered them carefully. For the comments regarding risk adjustment, the Committee reconsidered both the calibration testing (r-squared) and the (sociodemographic) factors included in the model. The Committee acknowledges that the r-squared value for this model is considered low, however, this is expected as the measure specifications (e.g., exclusion criteria) have inherently removed risk factors that would otherwise be related to cost variation and have accounted for factors that would be associated with cost variation between providers. This is balanced by a high mean reliability score (~0.92; n=10 clinicians) indicating that the variation observed in the measure score is mostly attributed to differences in performances rather than noise. At this time, the Committee has accepted the developer's rationale for not risk adjusting the for social factors based on the analysis completed to date. The Committee encouraged the developer to continue to explore approaches for accounting for social risk, particularly through performing analyses focused on clinicians with a high proportion of patients with social risk factors.

The Committee also carefully considered the concerns regarding the validity testing approach used to demonstrate construct validity. The Scientific Methods Panel (SMP) did not reach

consensus in their review of this measure due to concerns that the comparator measure used to demonstrate construct validity was not an independent construct and shared data elements with the measure under review. The Committee acknowledges this concern, but was unable to offer an alternative approach for administrative claims-based cost measures to demonstrate construct validity. Given the developer's extensive field testing and in the absence of guidance on validity testing, the Committee agreed to accept this testing for this evaluation and to request guidance from the Scientific Methods Panel on how to consider this testing in the future.

The Committee also recognizes the commenter's concerns with benchmarking and program design how meaningful differences will be distinguished on this measure in use. However, these aspects of measure use are not considered a part of the measure specifications and cannot be evaluated as part of the measure endorsement process.

The intent of this measure as understood by the Committee and described by the measure developer is to capture screening colonoscopy episode costs which is triggered by the colonoscopy procedure itself. While the Committee acknowledges that inadequate bowel prep can have an impact on the quality of a colonoscopy, procedures that are not performed due to poor bowel prep are not included in the measure. The Committee agrees that this cost measure is not and should not be designed to capture the quality of bowel prep.

MEASURE STEWARD/DEVELOPER RESPONSE:

We appreciate ACG's ongoing engagement with this measure, as the input from specialty societies and the clinician community is vital to the creation of valid, clinically sound episode-based cost measures. Our responses below are structured in parallel with the sections in the submitted comment.

Section 2b1 "Validity Testing": Our validity testing is intended to assess whether the measure is accurately measuring what it is designed to measure. This cost measure is intended to measure the costs to Medicare and beneficiaries for a clinician's performance of an episode of care for colonoscopy, including costs under the reasonable influence of the attributed clinician. For this reason, our analyses of empirical validity examine whether clinicians with higher costs in key categories of services also have higher measure scores.

We understand that, beyond the question of the validity of the measure, there are program-level questions about how this measure is related to existing quality measures. A valid cost measure may have a high correlation with a valid quality measure if clinicians who provide high quality care according to the available measures do so at higher cost. A valid cost measure may have a low correlation with a valid quality measure if a clinician who performs procedures with few complications and targeted utilization of services also has high quality outcomes (in terms of complications or other metrics). In the future, we will investigate such program-level questions to provide additional information to stakeholders.

Section 2b3 "Risk Adjustment/Stratification": As the MIPS Cost Category is currently scored based on percentile rankings of providers, we performed an analysis to look at the degree to which provider percentile rankings are affected by the inclusion of social risk factors. This

analysis shows that 90% of TINs would move no more than +-2 percentiles if social risk factors were included, and that the overall Pearson correlation of the measure with and without social risk factors is 0.999.

Risk adjustment for episode-based cost measures should be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted. This means that much of the variation out of the influence of the clinician is captured by service assignment, rather than risk adjustment. This is a point of distinction between this episode-based cost measure and measures that include all costs. A low R-squared does not necessarily indicate that a measure reflects variation unrelated to clinical care, while a high R-squared does not necessarily indicate the opposite; so, the risk adjustment models must be evaluated in concert with the service assignment rules.

The interpretation of R-squared results for the Colonoscopy measure was also discussed during the Standing Committee meeting, where members noted that the R-squared of the model largely reflects how standardized treatment is, and the variability across patients. As such, it is specific to the procedure or condition for which the costs are being evaluated under each measure, and the clinically specific factors that can affect costs within each measure.

Section 2b4 "Identification of Statistically Significant and Meaningful Differences in Performance": An analysis addressing this question can be found in Section 2a2, Reliability Testing. Reliability is a metric of the precision of the measure, or the ability of a measure to distinguish between "low" and "high" performers, where a result close to 1.00 indicates higher reliability, and a result closer to 0.00 indicates lower reliability. Our testing shows that mean reliability for this measure at a 10-case minimum is 0.956 for TINs, and 0.926 for TIN-NPIs. A mean reliability of 0.4 is generally seen as 'moderate', so these reliability results show that the measure has very high precision.

Our validity testing is intended to assess whether the measure is accurately measuring what it is designed to measure. This cost measure is intended to measure the costs to Medicare and beneficiaries for a clinician's performance of an episode of care for colonoscopy, including costs under the reasonable influence of the attributed clinician. For this reason, our analyses of empirical validity examine whether clinicians with higher costs in key categories of services also have higher measure scores.

We understand that, beyond the question of the validity of the measure, there are program-level questions about how this measure is related to existing quality measures. A valid cost measure may have a high correlation with a valid quality measure if clinicians who provide high quality care according to the available measures do so at higher cost. A valid cost measure may have a low correlation with a valid quality measure if a clinician who performs procedures with few complications and targeted utilization of services also has high quality outcomes (in terms of complications or other metrics). In the future, we will investigate such program-level questions to provide additional information to stakeholders.

We appreciate ASGE's ongoing engagement with this measure, as the input from specialty societies and the clinician community is vital to the creation of valid, clinically sound episode-based cost measures. Our responses below are structured in parallel with the sections in the submitted comment.

IM 1.2: The low minimum scores reflected idiosyncratic values on a few paid claims and the payment standardization algorithm's handling of these values. The payment standardization methodology prices services using Medicare payment rules. In rare situations, a colonoscopy is paid a positive amount but one which is substantially lower than the Medicare Physician Fee Schedule (PFS) amount. In these cases, the payment standardization methodology would use the actual paid amount instead of the Medicare PFS amount. The corrected testing results for the measure show that the actual minimum scores are $461.21 at both TIN and TIN-NPI levels, reflecting the measure as used in MIPS which does not include these extremely low paid colonoscopy cases.

IM 1.4: We will continue to study the effect of accounting for social risk factors, including race/ethnicity, that would allow us to view disparities and provide information to CMS to assist in any potential improvements. We routinely analyze the effect of accounting for social risk factors on risk adjustment models and clinician measure performance for this measure, as well as the other episode-based cost measures we have developed so far. Our analyses to date have shown that accounting for social risk factors in risk adjustment would not meaningfully affect measure performance, and that the correlation of measures with and without inclusion of social risk factors is very high (above 0.99 for both TIN and TIN-NPI level testing).

IM 2.1: The language in the submission form relating to bowel preparation was intended to indicate one potential factor in reducing downstream costs associated with a colonoscopy, and we recognize that other examples of services – including over-utilization of services or complications – may be more compelling examples for this measure. It was not intended to indicate that this measure is designed to capture data specifically around bowel preparation. Thank you for providing us the opportunity to clarify this point.

U 2.3: Thank you for your note, these bullets were created from development materials for the purpose of this submission form. In future, we will edit these for greater clarity."

We appreciate AGA's ongoing engagement with this measure, as the input from specialty societies and the clinician community is vital to the creation of valid, clinically sound episode-based cost measures. We thank the AGA for their support for NQF endorsement of this measure. Please find below responses to the questions in the submitted comment.

Social Risk Factors: As part of the testing process during development and on the final measure, we performed analyses on the impact of social risk factors on the measure. These analyses showed very minimal impacts on measure scores. The Pearson correlation between the measure with and without social risk factors was 0.999 at TIN and TIN-NPI level. We will continue to monitor the potential effect of social risk factors on the measure.

Within- and Between-Clinician Differences: An analysis of the within and between clinician differences in performance is available in Section 2a2, Reliability Testing. Our signal-to-noise analysis in this section looks at the within-clinician and between-clinician performance variance, with a reliability score close to 1.00 indicating higher reliability, and a result closer to 0.00 indicating lower reliability. Our testing shows that mean reliability for this measure at a 10-case minimum is 0.956 for TINs, and 0.926 for TIN-NPIs.

Risk Adjustment Variables: A list of factors used in the risk adjustment model can be found in the publicly available measure specifications at this location: https://qpp-cm-prod-content.s3.amazonaws.com/uploads/344/2019%20Cost%20Measure%20Code %20Lists.zip. In addition, risk factor variables and their respective coefficients and p-values can be found in the testing appendix in the NQF submission packet.

---

8. Consensus Standards Approval Committee (CSAC) Vote (November 12, 2019): Y-10; N-0

CSAC Decision: Approved for endorsement

---

9. Appeals

No appeals were received

3512 Knee Arthroplasty

Submission | Specifications

Description: The Knee Arthroplasty cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 30 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Knee Arthroplasty measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

Numerator Statement: The sum of the ratio of observed to expected payment-standardized cost to Medicare for all Knee Arthroplasty episodes attributed to a clinician or clinician group. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

Denominator Statement: The total number of episodes from Knee Arthroplasty episode group attributed to a clinician or clinician group within a performance period (i.e., MIPS performance year).

Exclusions:

- The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day.
- No attributed clinician is found for the episode.
- The beneficiary's date of birth is missing.
- The beneficiary's death date occurred before the trigger date.
- The beneficiary's death date occurred before the episode ended.
- The beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window.
- Episodes with inpatient procedures without relevant DRG codes
- Episodes for bilateral partial knee arthroplasties.
- Episodes where the beneficiary has reinsertion/reimplantation of prosthetic knee after infection or spacer during the trigger event or in a 120-day lookback period.
- Episodes with where the beneficiary has history of infections
- Episodes classified as outlier cases.

Adjustment/Stratification: Stratification by risk adjustment

Level of Analysis: Clinician : Group/Practice, Clinician : Individual

Setting of Care:

Type of Measure: Cost/Resource Use

Data Source: Claims, Enrollment Data, Other

Measure Steward: Centers for Medicare & Medicaid Services

STANDING COMMITTEE MEETING 06/27/2019

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

(1a. High Impact or High Resource Use, 1b. Opportunity for Improvement)

Importance to Measure and Report: **H-6; M-6; L-0; I-0**

Rationale:

- The developer provided citations demonstrating that medical and surgical readmissions following knee arthroplasty are costly. Specifically, 90-day readmissions for surgical complications cost an average of $28,000, and medical complications cost an average of $12,000. In a 2015 study of readmissions, approximately 4-5% resulted in a 30-day readmission representing significant opportunity for cost and performance improvement.
- The Standing Committee stated that this measure addresses significant, high-volume spending and high resource use, and it provides an opportunity for improvement.

---

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: SMP: **H-1; M-4; L-1; I-0**; Standing Committee: **H-3; M-9; L-0; I-0**; 2b. Validity: **H-1; M-8; L-3; I-0**

Rationale:

Reliability

- Reviewed by the Methods Panel who passed the measure on the reliability criteria.
- Data Element: Data element testing was conducted via CMS auditing programs for Parts A &B Claims data. The developer did not provide information on confirmation of the procedure and diagnosis code. The demonstration of data element validity did not meet NQF standards (i.e., description of CMS audits, fraud detection efforts)
- Score-Level: There were 237,376 Medicare beneficiaries included in the TIN level testing analysis and 227,075 beneficiaries included in the TIN-NPI level measure testing. Measure score reliability testing included test-retest with correlations, and signal to noise. Test-retest results found a Pearson correlation of 0.8 at group level and 0.75 at clinician level. Relatively stable movement across quintiles. The signal to noise analyses relied on the Adams' method (ratio of between variance to total variance). The mean reliability score was 0.87 for groups and 0.81 for clinicians. However, with mean reliability of the TIN at 0.72 for the lowest 10th percentile and 0.98 at 90th percentile indicates measure may be less reliable at lowest ranked levels.
- The Committee noted that the measure correlation of 0.8 for the group level and 0.75 for the clinician level was lower compared to the previous two measures discussed. The developer

provided additional analysis. The Committee commented that this measure can be applied at the group level.

- While the Committee believed that the measure, for reliability, should be rated as high in contrast to the SMP vote based on the information that the developer provided, the Committee ultimately voted this criterion as moderate.

Validity

- Reviewed by the Scientific Methods Panel who did not reach consensus on validity.
- The developer used a clinical subcommittee, a TEP, a person and family committee, and a national stakeholder feedback survey to provide input on measure and cost components attributable to this procedure episode of care measure. The face validity testing information provided by the developer does not meet NQF validity testing requirements since NQF requires that an expert group has been convened and a systematic assessment of the measure score has been conducted.
- Due to the inadequacy of face validity, the SMP focused its evaluation on the empirical validity testing.
- Empirical validity was assessed by examining correlation with other known indicators of resource utilization in administrative claims data, specifically hospital admissions (including readmissions) and post-acute care (PAC) services. They examined observed to expected spending for episodes with and without acute hospital readmission and with and without PAC. The mean observed to expected cost ratio for episodes without a hospital (re)admission was 0.99. The mean observed to expected cost ratio was 1.45 for episodes with a hospital (re)admission during the post-trigger period. The mean observed to expected cost ratio for episodes without PAC is 0.84. The mean observed to expected cost ratio compared was 1.09 for episodes that do contain some PAC.
- The NQF SMP encouraged the Cost and Efficiency Standing Committee to consider the empirical testing conducted by the developer to determine if it is adequate to meet the NQF endorsement criteria.
- Risk adjustment: The overall R-squared for the cost measure was 0.279 with an adjusted value of 0.278. Calibration demonstrated that the average observed to predicted cost is between 0.99 and 1.01 across risk score deciles. The developer noted that while individual bivariate testing demonstrated significance of the social factors, the inconsistent direction of the social risk factors and high correction between the measure scores with and without the social risk factors indicated that the final model sufficiently accounts for the effects of social risk factors on clinician measure scores.
- TEP Feedback: The NQF Clinical TEP generally agreed that the clinical population was appropriate but expressed some concerns and sought clarity on the rationale for some of the specifications and decision logic. In particular, the TEP was concerned that orders for unnecessary imaging by a primary care provider, (e.g., MRI) would impact a surgeon's costs in the episode, particularly given the evidence of overuse of this type of imaging.
- Meaningful Differences: The developer assessed meaningful differences by stratifying the clinician measure scores by meaningful characteristics and investigating the clinician score distribution by percentile. Characteristics included: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician. Large performance difference among clinicians: The measure score at the 99th percentile is over 1.6 times the measure score at the 1st percentile at both the TIN and TIN-NPI level; The measure score at the 90th percentile is approximately 30 percent greater than the score at the 10th percentile at the TIN level and TIN-NPI level; and The mean Knee Arthroplasty score for providers with Total Knee/Bilateral

sub-groups is 2.3 times the mean score for providers with Partial Knee/Unilateral sub-groups at the TIN and TIN-NPI levels.
- The Committee discussed the significance of dual status in the bivariate testing. The Committee noted that they would have preferred that the measure be stratified for the duals.
- In addition, the Committee noted that there was no analysis that the risk adjustment accounted for appropriate differences in post-acute care services. The developer responded to this point by commenting that post-acute care should be seen as a free-floating source of variation as it is rapidly changing to become narrower. The Committee replied to the developer, noting concern about unintended consequences of sending patients home too soon, without appropriate post-acute care.

---

3. Feasibility: H-8; M-4; L-0; 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:

- Claims-based measure.
- The Committee did not have any concerns on the feasibility.

---

4. Use and Usability

4a. Use; 4a1. Accountability and transparency; 4a2. Feedback on the measure by those being measured and others; 4b. Usability; 4b1. Improvement; 4b2. The benefits to patients outweigh evidence of unintended negative consequences to patients; 4b3.Measure con be deconstructed to facilitate transparency and understanding)

4a. Use: **Pass-12; No Pass-0** *4b. Usability:* **H-3; M-7; L-2; I-0**

Rationale:

- This is a new measure and is not currently publicly reported. The measure is used in a quality payment program, Merit-based Incentive Payment System (MIPS). While the measure has technically been implemented into the MIPS program, the measure results are first scheduled to be calculated for performance year 2019 (payment year 2021).
- The Committee provided similar comments as it had for 3509 and 3510 for feasibility and use.
- In addition, during its discussion of the usability criterion, the Committee discussed avoidable cost and if this measure will discourage appropriate post-acute services as a potential unintended consequence. The developer responded that they attempt to align their quality measures with previously endorsed measures and have aligned the cost measures with related readmissions measures within the post-acute care setting.

---

5. Related and Competing Measures

- No related or competing measures noted.

---

6. Standing Committee Recommendation for Endorsement: **Y-11**; **N-1**

---

7. Public and Member Comment

- NQF did not receive any comments on this measure.

---

8. Consensus Standards Approval Committee (CSAC) Vote (November 12, 2019): Y-10; N-0

CSAC Decision: Approved for endorsement

---

9. Appeals

No appeals were received

Appendix B: Cost and Efficiency Portfolio—Use in Federal Programs[a]

NQF #	Title	Federal Programs: Finalized or Implemented as of May 31, 2019
1598	Total Resource Use Population-Based PMPM Index	None
1604	Total Cost of Care Population-Based PMPM Index	None
2431	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Acute Myocardial Infarction (AMI)	Hospital Inpatient Quality Reporting (Implemented)
2436	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Heart Failure	Hospital Compare (Implemented) Hospital Inpatient Quality Reporting (Implemented)
2579	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Pneumonia	Hospital Compare (Implemented) Hospital Inpatient Quality Reporting (Implemented)
2158	Medicare Spending Per Beneficiary	None
3474	Hospital-Level, Risk Standardized Payment Elective for THA/TKA	None

[a] Per CMS Measures Inventory Tool as of 07/10/2019

Appendix C: Cost and Efficiency Standing Committee, Technical Expert Panels, and NQF Staff

STANDING COMMITTEE

Cheryl Damberg, PhD (Chair)
RAND Distinguished Chair in Healthcare Payment Policy
Santa Monica, California

Kristine Martin Anderson, MBA
Booz Allen Hamilton
Rockville, Maryland

Troy Fiesinger, MD, FAAFP
Village Family Practice
Houston, Texas

Nancy Garrett, PhD
Hennepin County Medical Center
Minneapolis, Minnesota

Andrea Gelzer, MD, MS, FACP
AmeriHealth Caritas
Philadelphia, Pennsylvania

Rachael Howe, MS, BSN, RN
3M HIS
Murray, Utah

Sunny Jhamnani, MD
Yale University
New Haven, Connecticut

Lisa Latts, MD, MSPH, MBA, FACP
Watson Health, IBM
Denver, Colorado

Jason Lott, MD, MHS, MSHP, FAAD
Bayer US LLC
Whippany, New Jersey

Jack Needleman, PhD
UCLA Fielding School of Public Health
Los Angeles, California

Janis Orlowski, MD, MACP
Association of American Medical Colleges
Washington, DC

John Ratliff, MD, FACS, FAANS
Stanford University Medical Center
Stanford, California

Srinivas Sridhara, PhD, MHS
Optum
New York, New York

Lina Walker, PhD
AARP — Public Policy Institute
Washington, District of Columbia

Bill Weintraub, MD, FACC
MedStar Washington Hospital Center
Washington, District of Columbia

Herbert Wong, PhD
Agency for Healthcare Research and Quality
Rockville, Maryland

CARDIOLOGY TECHNICAL EXPERT PANEL

Tim Dewhurst, MD FACC
Kaiser Permanente
Seattle, Washington

William Downey, MD
Interventional Cardiology Sanger Heart and Vascular Institute, Atrium HealthCare System
Charlotte, North Carolina

Sunny Jhamnani, MD
Dignity Health & Banner Health
Phoenix, Arizona

Mladen Vidovich, MD
University of Illinois at Chicago; Jesse Brown VA Medical Center
Chicago, Illinois

FAMILY/INTERNAL MEDICINE TECHNICAL EXPERT PANEL

Lindsay Botsford, MD, MBA/FAAFP
Physicians at Sugar Creek – Memorial Hermann Healthcare System
Sugar Land, Texas

Scott Fields, MD
OCHIN
Portland, Oregon

GASTROENTEROLOGY TECHNICAL EXPERT PANEL

Audrey Calderwood, MD
Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth-Hitchcock Medical Center
Lebanon, New Hampshire

Doug Corley, MD, PhD
Kaiser Permanente Northern California
Oakland, California

Johannes Koch, MD
Virginia Mason Medical Center
Seattle, WA

Larissa Temple, MD
Memorial Sloan-Kettering Cancer Center
New York, New York

Brian Jacobson, MD
Boston University School of Medicine
Boston, Massachusetts

NEUROLOGY TECHNICAL EXPERT PANEL

Ketan Bulsara, MD
University of Connecticut
Mansfield, Connecticut

James Burke, MD
University of Michigan
Ann Arbor, Michigan

Millie Hepburn, PhD, RN, SCRN, ACNS-BC
Pace University
Pleasantville, New York

David Seidenwurm, MD
Sutter Health, Sutter Medical Group
Sacramento, California

OPHTHALMOLOGY TECHNICAL EXPERT PANEL

Frank Burns, MD
Suburban Excimer Laser Center
Middletown, Kentucky

Kristen Carter, MD
Clarity Eye Care & Surgery
Tucson, Arizona

David Vollman, MD, MBA
Washington University School of Medicine; VA St. Louis Health Care System;
Center for Clinical Excellence BJC Healthcare
St. Louis, Missouri

ORTHOPEDIC SURGERY TECHNICAL EXPERT PANEL

Timothy Henne, MD
Orthopedic Associates of Michigan
Grand Rapids, Michigan

Bryan Little, MD
Detroit Medical Center, Detroit Medical Center
Detroit, Michigan

Anthony Mascioli, MD
University of Tennessee/Campbell Clinic
Memphis, Tennessee

Kimberly Templeton, MD
University of Kansas Medical Center
Kansas City, Kansas

VASCULAR SURGERY TECHNICAL EXPERT PANEL

Salvatore T. Scali, MD, FACS, RPVI
University of Florida-Gainesville
Gainesville, Florida

Clifford Sales, MD
Overlook Medical Center
Westfield, NJ

Sherene Shalhub, MD
University of Washington
Seattle, Washington

Christopher Smolock, MD
Cleveland Clinic
Cleveland, Ohio

Theodore Hwan Yuo, MD
University of Pittsburgh
Pittsburgh, Pennsylvania

NQF STAFF

Elisa Munthali, MPH
Senior Vice President, Quality Measurement

Ashlie Wilbon, MS, MPH, FNP-C
Senior Director

Kathryn Buchanan, MPH
Senior Project Manager

Janaki Panchal, MSPH
Project Manager

Hannah Ingber, MPH
Project Analyst

Navya Kumar, MPH
Project Analyst

Taroon Amin, PhD, MPH
NQF Consultant

Appendix D: Measure Specifications

3509 Routine Cataract Removal with Intraocular Lens (IOL) Implantation: Specifications

STEWARD

Centers for Medicare & Medicaid Services (CMS)/Acumen

DESCRIPTION

The Routine Cataract Removal with Intraocular Lens (IOL) Implantation cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 60 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Routine Cataract Removal with IOL Implantation measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

TYPE

Cost/Resource Use: Per episode

DATA SOURCE

Claims, Enrollment Data, Other

LEVEL

Clinician: Group/Practice, Clinician: Individual

NUMERATOR STATEMENT

The sum of the ratio of observed to expected payment-standardized cost to Medicare for all Routine Cataract Removal with IOL Implantation episodes attributed to a clinician or clinician group. Expected costs refer to costs predicted by the risk adjustment model. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

DENOMINATOR STATEMENT

The total number of episodes from the Routine Cataract Removal with IOL Implantation episode group attributed to a clinician or clinician group within a performance period (i.e., MIPS performance year).

EXCLUSIONS

The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day, no attributed clinician is found for the episode, the beneficiary's date of birth is missing, the beneficiary's death date occurred before the trigger date, the beneficiary's death date occurred before the episode ended, the beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window, the episode trigger claim was not performed in an outpatient hospital or

ASC setting, Episodes where the beneficiary has ocular comorbidities (impacting visual outcome of surgery or surgical complication rate), episodes classified as outlier cases

EXCLUSION DETAILS

Episodes are excluded for the following conditions, with the rationale for each provided below.

The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day:

This population is excluded to ensure that we have complete claims data for beneficiaries as there may be other claims (e.g., for services provided under Medicare Part C) that we do not observe in Medicare Parts A and B claims data. Including episodes that do not meet this criterion could potentially misrepresent a clinician's resource use. This exclusion also allows us to accurately construct HCCs for each episode by examining the episode's lookback period without missing claims.

No attributed clinician is found for the episode:

These episodes are excluded as the measure assesses clinician performance. The measure is intended to assess a homogeneous patient cohort to provide meaningful comparisons between attributed clinicians, so to include these episodes could potentially misrepresent these comparisons.

The beneficiary's date of birth is missing:

These episodes are excluded as a data cleaning step.

The beneficiary's death date occurred before the trigger date:

These episodes are excluded as a data cleaning step.

The beneficiary's death date occurred before the episode ended:

Episodes ending in death are excluded as they are - by definition - truncated episodes and do not have a complete episode window. Including episodes without all observable claims or a complete episode window could potentially make clinicians appear to have lower cost episodes not due to efficiencies of their own performance, but because the data are missing services that would be included in the measure calculation.

The beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window:

Similarly to above, these episodes are excluded as these beneficiaries may receive services not observed in the data. Including these episode could make the attributed clinician appear to have lower cost episodes due to incomplete data.

The episode trigger claim was not performed in an outpatient hospital or ASC setting

Episodes where the Part B Physician/Supplier claim with the CPT/HCPCS trigger code is not performed in an outpatient hospital or ASC are excluded to ensure that this measure captures a homogenous patient cohort, focusing on uncomplicated cataract removal procedures. Performing this procedure in other settings could indicate more complex procedures.

Episodes where the beneficiary has ocular comorbidities (impacting visual outcome of surgery or surgical complication rate)

Beneficiaries with significant ocular conditions are excluded from this measure, as defined by the presence of ICD-10 diagnosis and CPT/HCPCS codes on Part B Physician/Supplier, Outpatient, and Inpatient claims in a 120-day lookback period. These diagnosis and procedure

codes indicate significant ocular conditions (e.g., diabetic retinopathy) that impact the outcomes of surgery and expected resource use. Patients with these conditions are more likely to require more complex care that differs from the routine care in this measure. Exclusion of these patients is consistent with MIPS quality measures assessing the outcome of routine cataract surgery (NQF #0564 and #0565).

Episodes classified as outlier cases.

To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1st percentile and above the 99th percentile are considered outliers and removed from measure calculation.

RISK ADJUSTMENT

Stratification by risk category/subgroup.

The Routine Cataract Removal with IOL Implantation measure uses Medicare Part A and Part B claims data, which is maintained by CMS. Part A and B claims data are used to build episodes of care, calculate episode costs, and construct risk adjustors. Data from the Medicare Enrollment Database (EDB) are used to determine beneficiary-level exclusions and supplemental risk adjustors, specifically Medicare Parts A, B, and C enrollment; primary payer; disability status; end-stage renal disease (ESRD); beneficiary birth dates; and beneficiary death dates. The risk adjustment model also accounts for expected differences in payment for services provided to beneficiaries in long-term care, and that information comes from the Minimum Data Set (MDS). The MDS is used to create the Long Term Care Indicator variable in risk adjustment.

For measure testing, data from the American Census, American Community Survey (ACS), and Common Medicare Enrolment (CME) are used in the analyses evaluating social risk factors in risk adjustment.

STRATIFICATION

The Routine Cataract Removal with IOL Implantation measure is stratified into four sub-groups: ASC/Bilateral, ASC/Unilateral, HOPD/Bilateral, and HOPD/Unilateral. The stratification for site of service accounts for access factors, as some clinicians may not have access to an ASC, a lower cost setting than HOPD, due to regional availability or as a result of health plan contracting arrangements. Sub-groups for unilateral and bilateral surgery are used to account for scenarios where some services may be applied to a second surgery performed in close succession, meaning that bilateral procedures will likely be more expensive than unilateral ones. These sub-groups represent more homogenous patient cohorts to enable meaningful clinical comparisons based on information available on the trigger claim. These sub-groups are useful in ensuring clinical comparability so that the corresponding cost measure fairly compares clinicians with a similar patient case-mix. A separate risk adjustment model is created for each stratified group, so that clinically meaningful distinctions in the beneficiary population are preserved.

TYPE SCORE

Ratio

ALGORITHM

Grouping methodology and assignment algorithm: The Routine Cataract Removal with IOL Implantation cost measure evaluates resource use through the unit of episodes of care. The cost measure episodes are constructed by including select Medicare Part A and Part B claims

(assigned services) which occur during the episode window, defined as 60 days prior to the episode trigger to 90 days after the trigger. The episode trigger and assigned services are contained in the Measure Codes List file (see Section S.1. for details), along with risk adjustors, sub-groups, and exclusions.

3510 Screening/Surveillance Colonoscopy: Specifications

STEWARD

Centers for Medicare & Medicaid Services (CMS)/Acumen

DESCRIPTION

The Screening/Surveillance Colonoscopy cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care from the day of the clinical event that opens or 'triggers' the episode, through 14 days after the trigger. Beneficiary populations eligible for the Screening/Surveillance Colonoscopy measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

TYPE

Cost/Resource Use: Per episode

DATA SOURCE

Claims, Enrollment Data, Other

LEVEL

Clinician: Group/Practice, Clinician: Individual

NUMERATOR STATEMENT

The sum of the ratio of observed to expected payment-standardized cost to Medicare for all Screening/Surveillance Colonoscopy episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

DENOMINATOR STATEMENT

The total number of episodes from the Screening/Surveillance Colonoscopy episode group attributed to a clinician within a performance period (i.e., MIPS performance year).

EXCLUSIONS

The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day, no attributed clinician is found for the episode, the beneficiary's date of birth is missing, the beneficiary's death date occurred before the trigger date, the beneficiary's death date occurred before the episode ended, the beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window, the episode trigger claim was not performed in an outpatient hospital, office or ASC setting, the trigger event includes endoscopic mucosal resection, the trigger event

includes upper GI endoscopy, the patient has a history of inflammatory bowel disease, and episodes classified as outlier cases.

EXCLUSION DETAILS

Episodes are excluded for the following conditions, with the rationale for each provided below:

The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day.

This population is excluded to ensure that we have complete claims data for beneficiaries as there may be other claims (e.g., for services provided under Medicare Part C) that we do not observe in Medicare Parts A and B claims data. Including episodes that do not meet this criterion could potentially misrepresent a clinician's resource use. This exclusion also allows us to accurately construct HCCs for each episode by examining the episode's lookback period without missing claims.

No attributed clinician is found for the episode.

These episodes are excluded as the measure assesses clinician performance. The measure is intended to assess a homogeneous patient cohort to provide meaningful comparisons between attributed clinicians, so to include these episodes could potentially misrepresent these comparisons.

The beneficiary's date of birth is missing.

These episodes are excluded as a data cleaning step.

The beneficiary's death date occurred before the trigger date.

These episodes are excluded as a data cleaning step.

The beneficiary's death date occurred before the episode ended.

Episodes ending in death are excluded as they are - by definition - truncated episodes and do not have a complete episode window. Including episodes without all observable claims or a complete episode window could potentially make clinicians appear to have lower cost episodes not due to efficiencies of their own performance, but because the data are missing services that would be included in the measure calculation.

The beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window.

Similarly to above, these episodes are excluded as these beneficiaries may receive services not observed in the data. Including these episode could make the attributed clinician appear to have lower cost episodes due to incomplete data.

The episode trigger claim was not performed in an outpatient hospital, office or ASC setting.

Episodes where the Part B Physician/Supplier claim with the CPT/HCPCS trigger code is performed in an inpatient facility or emergency room are excluded. Screening and surveillance colonoscopy occurs in healthy individuals, and patients typically do not get their screening colonoscopies while hospitalized as they are sicker, and it is not appropriate to perform this elective procedure while being treated for an illness that requires hospitalization. Only episodes triggered in a clinician's office, an outpatient hospital or ambulatory surgery center are included.

The trigger event includes endoscopic mucosal resection.

Episodes will be excluded if this CPT/HCPCS code is found on Part B or outpatient claims during trigger event. Endoscopic mucosal resection is indicated for polyps that are larger and may be

more technically challenging to remove. They can be associated with higher risks of complications, so episodes are excluded to ensure clinical coherence in this patient cohort.

The trigger event includes upper GI endoscopy.

Episodes will be excluded if CPT/HCPCS codes for upper GI endoscopy are present on Part B Physician/Supplier claims or outpatient claims during the trigger event. Colonoscopies done in conjunction with an upper GI endoscopy suggests that the patient is having symptoms that necessitate the endoscopy, so the presence of these codes may indicate that the colonoscopy is diagnostic rather than being for screening.

The patient has a history of inflammatory bowel disease.

Episodes will be excluded if ICD-10 diagnosis codes for inflammatory bowel disease are present on Part B Physician/Supplier, outpatient, or inpatient claims during the 120-day lookback period. Beneficiaries with history of such conditions (e.g., Crohn's disease, ulcerative colitis, diverticulitis) have a higher risk of colon cancer and higher risk of complications from colonoscopy given their underlying comorbid condition of an inflamed large bowel.

Episodes classified as outlier cases.

To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1st percentile and above the 99th percentile are considered outliers and removed from measure calculation.

RISK ADJUSTMENT

Stratification by risk category/subgroup.

The risk adjustment model includes variables from the CMS-HCC V22 2016 Risk Adjustment Model, as well as other standard risk adjustors (e.g., beneficiary age brackets using information in the Medicare beneficiary enrollment database) and disease interaction terms. The model also includes variables specific to this cost measure, identified through the incorporation of detailed clinical input. These variables account for factors that are likely to affect cost, such as types of hypertension or history of anesthesia difficulties, among others.

The CMS-HCC V22 model uses 79 Hierarchical Condition Category (HCC) indicators derived from the beneficiary's claims in the period 120 days prior to the episode trigger day. Other risk adjustors are originally "Disabled without end-stage renal disease (ESRD)" or "Disabled with ESRD" using the original reason for joining Medicare in the Medicare beneficiary enrollment database. The risk adjustment model also identifies beneficiaries who have spent at least 90 days in a long-term care institution without having been discharged to the community for 14 days, based on MDS assessment data. Additional information about the risk adjustment model is included in Section S.8.6.

The Screening/Surveillance Colonoscopy episode group includes all services identified as being clinically relevant to this procedure. There are logic rules to determine when and what conditions each particular service will be assigned, as detailed in the Measure Codes List file (see Section S.1 for URL).

STRATIFICATION

The Screening/Surveillance Colonoscopy measure is stratified by place of service into three sub-groups: ambulatory surgery centers, hospital outpatient department and office. These sub-groups represent more homogenous patient cohorts to enable meaningful clinical comparisons based on information available on the trigger claim. These sub-groups are useful in ensuring

clinical comparability so that the corresponding cost measure fairly compares clinicians with a similar patient case-mix. A separate risk adjustment model is created for each stratified group, so that clinically meaningful distinctions in the beneficiary population are preserved. Since Medicare pays different amounts for the same service in the three settings, the decision was made to create sub-groups so providers would not be affected by the site of service payment differential. Site of service sub-groups ensure that costs were compared across homogeneous patient populations and account for cost differences specifically related to facility type.

TYPE SCORE

Ratio

ALGORITHM

Grouping methodology and assignment algorithm: Screening/Surveillance Colonoscopy cost measure evaluates resource use through the unit of episodes of care. The cost measure episodes are constructed by including select Medicare Part A and Part B claims (assigned services) which occur during the episode window, defined as from the day of the episode trigger to 14 days after the trigger. The episode trigger and assigned services are contained in the Measure Codes List file (see Section S.1. for details), along with risk adjustors, sub-groups, and exclusions.

3512 Knee Arthroplasty: Specifications

STEWARD

Centers for Medicare & Medicaid Services (CMS)/Acumen

DESCRIPTION

The Knee Arthroplasty cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive this procedure. The cost measure score is a clinician's average risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 30 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Knee Arthroplasty measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

TYPE

Cost/Resource Use: Per episode

DATA SOURCE

Claims, Enrollment Data, Other

LEVEL

Clinician: Group/Practice, Clinician: Individual

NUMERATOR STATEMENT

The sum of the ratio of observed to expected payment-standardized cost to Medicare for all Knee Arthroplasty episodes attributed to a clinician or clinician group. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

DENOMINATOR STATEMENT

The total number of episodes from Knee Arthroplasty episode group attributed to a clinician or clinician group within a performance period (i.e., MIPS performance year).

EXCLUSIONS

The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day, no attributed clinician is found for the episode, the beneficiary's date of birth is missing, the beneficiary's death date occurred before the trigger date, the beneficiary's death date occurred before the episode ended, the beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window, episodes with inpatient procedures without relevant DRG codes, episodes for bilateral partial knee arthroplasties, episodes where the beneficiary has reinsertion/reimplantation of prosthetic knee after infection or spacer during the trigger event or in a 120-day lookback period, episodes with where the beneficiary has history of infections, and episodes classified as outlier cases.

EXCLUSION DETAILS

Episodes are excluded for the following conditions, with the rationale for each provided below:

The beneficiary has a primary payer other than Medicare for any amount of time overlapping the episode window or in the 120 days prior to the episode trigger day.

This population is excluded to ensure that we have complete claims data for beneficiaries as there may be other claims (e.g., for services provided under Medicare Part C) that we do not observe in Medicare Parts A and B claims data. Including episodes that do not meet this criterion could potentially misrepresent a clinician's resource use. This exclusion also allows us to accurately construct HCCs for each episode by examining the episode's lookback period without missing claims.

No attributed clinician is found for the episode.

These episodes are excluded as the measure assesses clinician performance. The measure is intended to assess a homogeneous patient cohort to provide meaningful comparisons between attributed clinicians, so to include these episodes could potentially misrepresent these comparisons.

The beneficiary's date of birth is missing.

These episodes are excluded as a data cleaning step.

The beneficiary's death date occurred before the trigger date.

These episodes are excluded as a data cleaning step.

The beneficiary's death date occurred before the episode ended.

Episodes ending in death are excluded as they are - by definition - truncated episodes and do not have a complete episode window. Including episodes without all observable claims or a complete episode window could potentially make clinicians appear to have lower cost episodes

not due to efficiencies of their own performance, but because the data are missing services that would be included in the measure calculation.

The beneficiary was not enrolled in Medicare Part A and B for the entirety of the 120-day lookback period plus episode window, or is enrolled in Part C for any part of the lookback period plus episode window. Similarly to above, these episodes are excluded as these beneficiaries may receive services not observed in the data. Including these episode could make the attributed clinician appear to have lower cost episodes due to incomplete data.

Episodes with inpatient procedures without relevant DRG codes

Episodes will be excluded if the procedure occurred in the inpatient setting and if its concurrent inpatient stay does not have MS-DRG codes that indicate that the reason for admission was for this procedure. These cases are excluded to limit the measure to only capture admissions where the reason for admission is for the knee arthroplasty because cases admitted for other reasons are likely to be more expensive because of the cost of care for the reason for admission as well as for the knee joint replacement.

Episodes for bilateral partial knee arthroplasties.

Beneficiaries who undergo same-day or staged bilateral partial knee arthroplasties (within 90 days of the first procedure) as identified by modifier codes are excluded from this measure, as rare occurrences. With such a small sample size, these cases are excluded as the effect on expected episode cost is unclear.

Episodes where the beneficiary has reinsertion/reimplantation of prosthetic knee after infection or spacer during the trigger event or in a 120-day lookback period.

Episodes for procedures for reinsertion or reimplantation either during the trigger event or identified during a 120-day lookback period are excluded because these patients require considerably different clinical care, and are likely to be more expensive.

Episodes with where the beneficiary has history of infections

Episodes where the beneficiary has recent infections to the knee indicated through CPT/HCPCS and ICD-10 diagnosis codes during a 120-day lookback are excluded because it indicates that these patients require considerably different clinical care, and are likely to be more expensive.

Episodes classified as outlier cases.

To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1st percentile and above the 99th percentile are considered outliers and removed from measure calculation.

RISK ADJUSTMENT

Stratification by risk category/subgroup.

Clinical hierarchies are embedded in the risk adjustment model. The risk adjustment model includes variables from the CMS-HCC V22 2016 Risk Adjustment Model, as well as other standard risk adjustors (e.g., beneficiary age brackets using information in the Medicare beneficiary enrollment database) and disease interaction terms. The model also includes variables specific to this cost measure, identified through the incorporation of detailed clinical input. These variables include conditions which may influence the episode cost and risk of complication, for example; osteoporosis, post-infectious osteoarthritis, or psoriatic arthritis, amongst others.

The CMS-HCC V22 model uses 79 Hierarchical Condition Category (HCC) indicators derived from the beneficiary's claims in the period 120 days prior to the episode trigger day. Other risk adjustors are originally "Disabled without end-stage renal disease (ESRD)" or "Disabled with ESRD" using the original reason for joining Medicare in the Medicare beneficiary enrollment database. The risk adjustment model also uses an indicator for beneficiaries identified as having had recent need of long-term care (90 days in a long-term care institution without having been discharged to community for 14 days) using MDS assessment data. Additional information about the risk adjustment model is included in Section S.8.6.

The Knee Arthroplasty episode group includes all services identified as being clinically relevant to this procedure. There are logic rules to determine when and what conditions each particular service will be assigned, as detailed in the Measure Codes List file (see Section S.1 for URL).

STRATIFICATION

The Knee Arthroplasty measure is stratified into three sub-groups: Partial Knee/Unilateral, Total Knee/Unilateral and Total Knee/Bilateral. These sub-groups represent more homogenous patient cohorts to enable meaningful clinical comparisons based on information available on the trigger claim. These sub-groups are useful in ensuring clinical comparability so that the corresponding cost measure fairly compares clinicians with a similar patient case-mix. A separate risk adjustment model is created for each stratified group, so that clinically meaningful distinctions in the beneficiary population are preserved.

The Knee Arthroplasty measure stratifies cases by partial and total knee replacements because of differing surgical durations and differing rates of complication, readmission and revision. These differences lead to substantial differences in total costs.

Unilateral partial knee procedures have been outpatient procedures for past years and are likely to be less expensive than unilateral total knee replacements. Bilateral total knee replacements are generally require different resources than unilateral procedures. There will be some services that will not have to be repeated for the second procedure but the patient is likely to require additional rehabilitation, for example at an inpatient rehabilitation hospital stay after the bilateral procedure.

The Knee Arthroplasty measure accounts for the removal of knee arthroplasty procedures from the inpatient-only list through the use of the CPT/HCPCS code to trigger the episode. The current trigger code is based on CPT/HCPCS codes and does not require an inpatient stay. Additionally, risk adjustment for the DRG of the inpatient stay was included, if one is associated with the knee arthroplasty. Specifically, the episode should be included only when the trigger code appears concurrently with MS-DRGs 461, 462, 469 and 470, indicating that the hospital stay was for the knee arthroplasty. With total knee arthroplasties now being allowed in an outpatient setting, patients who receive a knee arthroplasty in an inpatient setting are likely sicker, and clinicians taking care of these patients should not be penalized for the necessary precaution of a longer inpatient stay. The Knee Arthroplasty cost measure also includes Partial Knee Arthroplasty which is sub-grouped in order to create a more homogenous comparison.

TYPE SCORE

Ratio

ALGORITHM

Grouping methodology and assignment algorithm: The Knee Arthroplasty cost measure evaluates resource use through the unit of episodes of care. The cost measure episodes are

constructed by including select Medicare Part A and Part B claims (assigned services) which occur during the episode window, defined as 30 days prior to the episode trigger and 90 days after the trigger. The episode triggers and assigned services are contained in the Measure Codes List file (see Section S.1. for details), along with risk adjustors, sub-groups, and exclusions.

Appendix E: Related and Competing Measures

There were no related and competing measures for the spring 2019 cycle submitted measures.

Appendix F: Pre-Evaluation Comments

Comments received as of June 19, 2019.

NQF 3509 Routine Cataract Removal with Intraocular Lens (IOL) Implantation

Submitted by American Society of Cataract and Refractive Surgery

The American Society of Cataract and Refractive Surgery (ASCRS) is a medical specialty society representing nearly 9,000 ophthalmologists in the United States and abroad who share a particular interest in cataract and refractive surgical care.

Thank you for the opportunity to provide comments on the pending endorsement of the cost measure, Routine Cataract Removal with Intraocular Lens (IOL) Implantation (cataract episode measure). **ASCRS opposes NQF endorsement of this measure unless it excludes any FDA-approved Medicare Part B drug on pass-through status administered during cataract surgery. Specifically, we are concerned that if pass-through drugs are included in the cataract episode measure, it will disincentivize surgeons from using the drugs and negatively impact the utilization data CMS is collecting for this purpose. Ultimately, including Medicare Part B drugs on pass-through in the episode measure defeats the purpose of pass-through.**

In addition to our concerns related to the inclusion of pass-through drugs in this measure, we are providing comments on the following elements:

Support for the trigger code, exclusions for risk adjustment, and the sub-groups of the episode measure. These measure specifications were developed with the input of ASCRS and other ophthalmologists through Acumen's technical expert panel (TEP), of which I was a member. These factors reflect the TEP's goal of only measuring physicians on the costs they can influence.

Ongoing concern that when physicians are scored on this measure they are given an adjusted cost, expressed as a dollar figure that may differ considerably from the actual cost, and then compared to total national average cost. We recommend the cost score be expressed as the difference between the physician's expected cost versus the physician's observed cost.

EXCLUDE DRUGS ON PASS-THROUGH

ASCRS urges NQF to withhold its endorsement of the cataract surgery episode-based cost measure unless CMS and Acumen remove the current FDA-approved drug administered during cataract surgery included in the measure on pass-through, and signal that any drug that has since come onto the market and is paid on pass-through, or will come onto the market, will not be included in the measure.

Pass-through status is a vital tool in ensuring that new and innovative drugs are introduced to the market and is used by CMS in the formula to calculate the increase in the ambulatory payment classification (APC) group, which is the facility fee for the procedure and other related procedures, to account for the drug. Pass-through status helps introduce a new drug into the marketplace that is used during or immediately after surgical procedures with an average estimated cost that exceeds a certain percentage of the procedure's ambulatory payment classification (APC) payment amount. It is initially put on pass-through status and paid separately for up to three years under Medicare Part B. This encourages the use of new drugs in the facility by allowing physicians time to become familiar with their use without their adding to facility cost. Separate payment for pass-through drugs is also essential to ASCs, in particular, because their lower facility reimbursements would make it difficult to afford new, high-cost drugs. Over this time period, CMS measures the utilization of the drug and, when the drug goes off pass-through status, adjusts the reimbursement level for the bundled facility fee based on the

utilization data gathered and the formula. **Pass-through status is intended as a means for CMS to gather data not influenced by other factors. If drugs on pass-through status are included in the measure, physicians mindful of their score on the cataract surgery measure may modify their use of the drug for reasons other than clinical appropriateness, and thus impact the gathering of utilization data, thereby defeating the purpose of pass-through.**

Currently, there are several ophthalmic drugs that have either recently been approved or will be approved in the near future for use during cataract surgery. One such drug—injection, phenylephrine and ketorolac, 4 ml vial—is included in the measure for 2019. Specifically, these new FDA-approved drugs administered during cataract surgery that are on now on pass-through, or soon will be, have a post-operative indication, such as post-operative pain and inflammation and/or other sequela of the surgery, and eliminate the need for some or all post-operative eye drops. Reducing or eliminating the need for post-operative eye drops, which are currently furnished under Medicare Part D, represents a substantial cost-saving both to the Medicare program and the patient. In addition, eliminating the need for post-operative eye drops improves patient compliance and leads to better clinical outcomes. However, since Part D costs are not a factor in the cataract episode measure, using these Medicare Part B pass-through medications during cataract surgery and including them in the episode calculation would increase the total episode cost and would inaccurately designate the surgeon as high-cost. **Beyond the primary goal of preserving pass-through status to ensure accurate utilization calculations, we believe including these drugs with a post-operative indication on pass-through would go against the goal of the episode-based cost measures of encouraging physicians to make more efficient use of resources.**

ASCRS believes that episode-based cost measures are a more effective method of measuring clinician resource use than population-based measures because they only include the costs of care that are within the physician's control. However, physicians have no control over the cost of drugs as they enter the market, and therefore, including the cost of these drugs in the measure is contrary to the goals of episodic-based measurement. To ensure that clinicians are not penalized for using drugs on pass-through and that pass-through status is preserved to collect accurate, market-based utilization data, we recommend that any FDA-approved Medicare Part B drug administered during, or at the end of, cataract surgery that is on pass-through status be excluded from the cataract surgery episode-based cost measure. NQF should withhold its endorsement of the measure until there are assurances from CMS that the current drug on pass-through included in the measure, and any other pass-through drug that has entered the market or will in the future, be excluded.

TRIGGER CODE

ASCRS supports the use of CPT code 66984 as a trigger for the cataract episode measure. Routine cataract removal with 66984 requires homogeneous and comparable resources for nearly all patients. As a high-volume code, it will provide enough data to identify outlier physicians who are practicing outside of established patterns. **66984 is the only code included in the specifications as a trigger code. No other codes should be considered.**

ASCRS does not support including other codes, such as complex cataract surgery, 66982, in this measure, as it will not yield comparable enough data to measure a physician's resource use accurately. Patients undergoing cataract surgery that requires the use of the complex cataract code may suffer from a wide variety of ocular co-morbidities, or other non-ocular co-morbidities, which could require varying levels of resource use depending on the condition. For example, patients taking Tamsulosin or similar medications very frequently require the use of iris retractors, leading to the use of code 66982 instead of the usual 66984. Furthermore, these patients often have complications requiring further surgery, such as a vitrectomy. **Complex cataract surgery may require additional supplies and increases the**

likelihood of potential complications, resulting in a range in value too significant to provide a homogenous patient group for cost measures and, therefore, should not be used as a trigger code.

EXCLUSIONS

ASCRS supports the use of criteria to exclude patients with significant ocular co-morbidities from the cost measure. We support the use of the exclusionary criteria from quality measure 191, Cataracts: 20/40 or Better Visual Acuity within 90 Days Following Cataract Surgery. The criteria exclude patients with documentation of significant ocular conditions. The exclusions include such chronic conditions as diabetic retinopathy, macular degeneration, and glaucoma.

The rationale for excluding these patients is that since patients with significant ocular co-morbidities are excluded from quality measurement, they should also be excluded from cost measurement. The developers of quality measure 191 excluded these patients because ocular co-morbidities play a significant role in whether the patient will have a predictably good outcome, and whether complications may arise. Surgeons do not have control over a patient's co-morbidities and should not be held accountable for additional costs in an episode if a patient suffers from one of these conditions. **If physicians are not measured on the quality outcomes of these surgeries, therefore, they should not be held responsible for the cost of these surgeries.**

These exclusions ensure a greater level of risk adjustment than has previously been incorporated in cost measures, such as the current measures total cost per capita and Medicare spending per beneficiary. While we are aware that CMS includes a basic level of risk adjustment calculation for existing cost measures, and for the episode-based measures, CMS has not been able to demonstrate that its risk adjustment properly reflects the cost of cataract surgery. This risk adjustment includes factors such as beneficiary age, dual-eligibility, and some co-morbidities, but does not include ocular co-morbidities. Cataract surgery is performed on a relatively older patient base, and while some systemic co-morbidities may require additional resource use as discussed above, ocular co-morbidities play a much larger role in determining the resource use and likelihood of a good outcome than do the factors used in CMS' current risk adjustment. **Using the exclusionary criteria from the cataract quality measure is a much more accurate means of risk adjustment to ensure that physicians are not held accountable for the cost of care related to factors outside their control.**

EPISODE SUB-GROUPS

ASCRS supports the inclusion of the measure sub-groups, which relate to site of service and laterality. We believe the four sub-groups in the cataract episode measure represent significant differences in the cost of cataract surgery—largely related to Medicare's own billing and reimbursement policies—and should be separated for basis of comparison in this measure.

Cataract surgery can be performed in either hospital outpatient departments (HOPDs) or ambulatory surgery centers (ASCs), with reimbursements for ASCs well below HOPDs. Cataract surgery is reimbursed 45% less in the ASC than in the HOPD. While some ophthalmologists have the option of building and owning their own ASC, some states with certificate of public need laws limit the number of existing ASCs or prevent physicians from opening new ASCs, so they may be forced to operate in HOPDs. In addition, some physicians, especially solo practitioners, may not have the resources to construct and manage their own ASC, and must operate in whatever facility, either ASC or HOPD, is available. Despite these limitations and given the choice, ophthalmic surgeons would likely prefer to operate in the lower-cost ASC. ASCs are not subject to the same requirements as HOPDs, such as extensive pre-operative testing, that are not relevant to treating ophthalmic disease. In addition, patients may prefer to undergo surgery in ASCs, which are generally easier to navigate since they are smaller, less intimidating, and have shorter wait times. Ophthalmic surgeons want to make the cost-effective choice but cannot always do

so. **Given that, the episodes must include sub-groups for ASCs and HOPDs, since the site of service is not always within the physician's control.**

ASCRS supports the sub-groups for laterality because they reflect whether the surgeon removed cataracts in either one or both eyes during the episode window. Patients frequently develop cataracts in both eyes, and while both eyes are rarely operated on at the same time, many patients find it convenient to have the second surgery shortly after the first, usually still within the 90-day global post-operative period. Medicare has specific billing rules for physicians performing multiple procedures on the same patient related to pre- and post-operative care; therefore, the expected cost of cataract surgery performed during the global period of previous cataract surgery would be substantially different from two surgeries performed more than 90 days apart.

MEASURE SCORE BASED ON NATIONAL AVERAGE PRICE

As noted above, we support the measure sub-groups, which were determined by the clinical TEP. The physicians on the TEP selected those sub-groups to separate and avoid comparing surgeries where fundamentally different factors are contributing to the amount Medicare is reimbursing in total for the episode that are not always in the control of the physician. Given that fact, it is confusing to the physician receiving feedback on cost performance to see a dollar figure representing his or her "average" cost compared to a national average. **We believe physicians would better understand their performance on the measure and be able to take action in response if they were shown how their average observed cost compared to their average expected cost.**

The cataract episode should compare each physician or TIN's average observed cost to that same physician or TIN's average expected cost and not compare overall to a national average. The clinical TEP determined that the sub-groups broadly represent the main drivers of cost in relation to cataract surgery. CMS and Acumen will determine risk-adjusted expected costs for each of the sub-groups. Each of a physician's attributed episodes' observed costs will then be compared to the expected cost for the respective sub-group and assigned a ratio to represent the divergence between the expected and observed costs. The ratios for each of the episodes are then averaged to determine the frequency of the physician's divergence from the expected cost. **We support this approach, as it ensures that the varying costs of the sub-groups outside of the physician's control, such as the facility fee, are not impacting the physician's score.**

However, the steps following the calculation of the average ratio should be re-thought to make the final average cost more meaningful to the physician or group practice. Following the above calculation, the ratios of observed and expected costs are converted back to a dollar figure and compared to a national average. When draft field test reports of the measure were distributed in late 2017, we heard from several of our members who reviewed their feedback reports and questioned how their final average cost and the national average costs were determined, and how it relates to the actual reimbursement they received for the surgery. In addition, physicians are aware that geographic differences contribute to the reimbursement level and may question why they should be compared to a national dollar average. While we understand that supplemental methodology documents are available and discuss this process, CMS does not generally distribute these explanatory resources in conjunction with performance feedback, and if a clinician does find them posted on a separate web page, they may be difficult to understand.

To overcome this issue, we recommend the physician's score be based on his or her own expected and observed costs, and not based on a misleading national average. For example, if a surgeon's case mix means that 70% of his or her surgeries are unilateral and performed in the ASC, and 30% are unilateral and performed in the HOPD, then his or her expected cost should be a weighted average of those two sub-groups. Then, the surgeon's actual observed costs are compared to that expected average of a

70/30 mix of those two sub-groups. The average cost for this surgeon would be substantially different from a physician who performed 90% of surgeries in an HOPD and 10% in an ASC. **The surgeon would then be evaluated on the extent, above or below, that he or she deviates from the expected cost for his or her specific case mix. We believe this is a much more useful value to a physician than a national average.**

CONCLUSION

While ASCRS believes that episode-based measures are a more accurate means of measuring resource use because they only include the costs that are within the clinician's control, we oppose NQF endorsement of this measure until CMS and Acumen remove the current drug on pass-through from the measure and indicate that no other pass-through drug will be included in the future. We are deeply concerned that including these drugs will improperly influence utilization data collection and create an adverse incentive for physicians to avoid using innovative drugs that improve patient outcomes because they will negatively impact cost scores. If CMS and Acumen make these changes, the measure should move forward for endorsement with the current specifications for trigger code, sub-groups, and exclusions. We continue to recommend that the final score of the measure not be compared to one national average figure.

Thank you again for the opportunity to provide input.

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

ISBN