Partnership for Quality Measurement
Powered by Battelle

National Consensus Development and Strategic Planning for Health Care Quality Measurement

2023 Pre-Rulemaking Measure Review (PRMR) Meeting Summary: Clinician Committee

Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, Maryland 21244-1850
January 2024

The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services.

Table of Contents

List of Tables

Page

List of Figures

Page

Pre-Rulemaking Measure Review (PRMR)

Battelle staff convened the PRMR Recommendation Group on January 16 and 17, 2024, for discussion and voting on the Measures Under Consideration (MUC) for 2023.

The goal of this meeting was to discuss the proposed additions to CMS programs through the perspective of interested parties impacted by the program. This meeting summary provides an overview of the meeting and its outcomes and will be followed by a comprehensive PRMR Meeting Recommendations Report and Recommendations Spreadsheet.

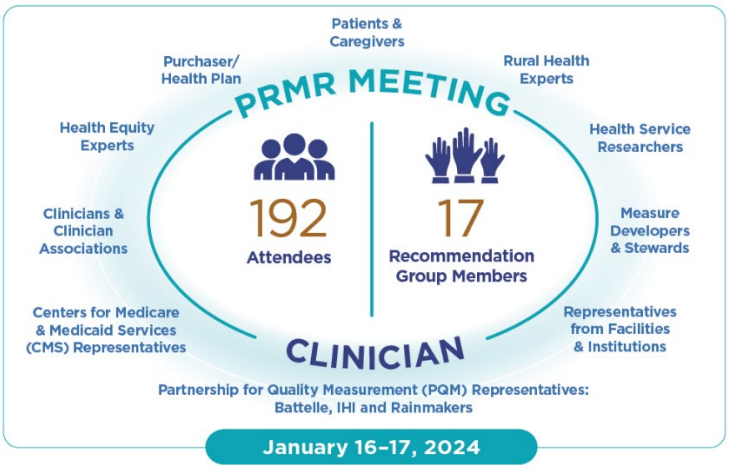

Figure 1. PRMR Meeting Attendance

Meeting participants joined virtually through the Zoom meeting platform. Figure 1 outlines overall meeting attendance, which was comprised of the PRMR Recommendation Group, the PRMR Advisory Group, the general public, and other interested parties. The PRMR Recommendation Group responsible for measure discussion and voting was comprised of 17 members. These members represented the interested parties shown in Figure 1 and were joined by CMS and Battelle's Partnership for Quality Measurement (PQM) representatives.

Overview and Purpose

Dr. Nicole Brennan, Executive Director of PQM and meeting co-facilitator, welcomed the attendees to the meeting and provided a brief overview of the agenda for the meeting. Ms. Kate Buchanan, Deputy PRMR Lead, reviewed housekeeping items and conducted roll call; during roll call, Recommendation Group members disclosed any conflicts of interest regarding the measures under review. One member reported they had a conflict and recused themselves from voting for MUC2023-137. Following the roll call, meeting co-facilitator and PQM Technical Director, Ms. Brenna Rabel, introduced the Recommendation Group co-chairs Mr. Reginald Barnes and Dr. Lisa Hines. The co-chairs each shared their relevant perspectives and motivation for serving in this role.

Dr. Michelle Schreiber, the Deputy Director of the Center for Clinical Standards and Quality (CCSQ) for CMS and Director of the Quality Measures and Value Based Incentives Group (QMVIG) within CCSQ, provided her opening remarks, starting by thanking Battelle, the

committee co-chairs, public commenters, her CMS colleagues, and other federal agencies that have provided input on the measures. Dr. Schreiber reminded the group that the programs are authorized by Congress and, as such, have statutory requirements that must be followed. Additionally, final decisions about what is proposed or finalized are the purview of the government and there are many layers of interested party input that go into each measure. Despite this, input from this committee is essential for ensuring consensus. Public opinion informs CMS decisions, and the committee's recommendations are extremely important to CMS.

Regarding feedback received in public comments, Dr. Schreiber noted that burden was a predominant theme. CMS has worked to reduce the measure inventory by 20%, and when a measure is proposed, a measure is often removed to keep the number of measures relatively stable. Digital measures will be a way of reducing burden as well. Other comments noted a lack of clarity around how measures would translate to improvement. Dr. Schreiber emphasized that all the measures are evidence based, and that measures are designed to show providers how they can improve.

Dr. Schreiber clarified that the pre-rulemaking meetings would focus largely on the Merit-based Incentive Payment System (MIPS) program and Medicare Shared Savings Program (MSSP); and that the meetings would not focus on measures used in the Center for Medicare and Medicaid Innovation (CMMI) Innovation Models, constructs in the MIPS Value Pathways, or measures used by Qualified Clinical Data Registries (QCDRs). Dr. Schreiber noted that measure choice in the MIPS program is voluntary, and clinicians have more than 200 measures from which to choose. Dr. Schreiber closed her remarks by emphasizing that CMS was looking forward to hearing the committee's comments over the coming days.

Before moving into the measure discussions, Ms. Rabel asked Recommendation Group members to introduce themselves.

MUC 2023 Clinician Committee Measure Discussion

After opening remarks, Battelle facilitators outlined the procedures for discussing and voting on measures. The discussion quorum required the attendance of at least 60% of the Recommendation Group members during roll call at the beginning of the meeting. The voting quorum required at least 80% of active Recommendation Group members who had not recused themselves from the vote. During the two meeting days, some members stepped away temporarily, so Battelle collected voting counts for each measure to ensure quorum was retained. For the determination of a vote outcome, a majority of greater than or equal to 75% of voting members was required. For votes of Recommend and Recommend with Conditions, a combination of 75% of voting members split between those two options would result in a determination of Recommend with Conditions.

At the beginning of each measure discussion, Battelle introduced the measure then a CMS program lead representative would give an overview of the measure and rationale for inclusion in CMS programs. Battelle provided a summary of the public comments and round 1 evaluations and opened the discussion up to Recommendation Group members.

PRMR Recommendation Group members were then asked to vote on each MUC 2023 measure for clinician committee relevant programs. The voting options were Recommend, Recommend

with Conditions[1], or Do Not Recommend. Members voted in real time via the Voteer platform. Table 1 shows the vote counts by measure.[2]

[1] Following confusion regarding what constituted conditions on Day 1 of the meeting, Battelle presented clarifying guidance on Day 2 and asked members to indicate their conditions for each measure, if applicable.
[2] The total number of votes by measure varied due to recusals. Additionally, committee members did not vote for a given measure if they did not participate in a significant portion of that measure's discussion.

Table 1. PRMR Recommendation Group Vote Counts per Measure (Clinician Committee, MUC 2023)

MUC ID	Measure Title	Program	Determination	Recommend N (%)	Recommend with Conditions N (%)	Do Not Recommend N (%)	Recusals
MUC2023-137	Initial Opioid Prescribing for Long Duration (IOP-LD)	Part C & D Star Rating	Consensus Not Reached	1 (7%)	8 (57%)	5 (36%)	1
MUC2023-179	Initiation and Engagement of Substance Use Disorder Treatment (IET)	Part C & D Star Rating	Consensus Not Reached	2 (14%)	8 (57%)	4 (29%)	0
MUC2023-212	Level I Denials Upheld Rate Measure	Part C & D Star Rating	Recommend	13 (87%)	1 (7%)	1 (7%)	0
MUC2023-164	Adult COVID-19 Vaccination	Merit-based Incentive Payment System	Consensus Not Reached	0 (0%)	5 (29%)	12 (71%)	0
MUC2023-211	Melanoma: Tracking and Evaluation of Recurrence	Merit-based Incentive Payment System	Consensus Not Reached	4 (24%)	6 (35%)	7 (41%)	0
MUC2023-141	Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy	Merit-based Incentive Payment System	Recommend with Conditions	6 (38%)	10 (63%)	0 (0%)	0

MUC ID	Measure Title	Program	Determination	Recommend N (%)	Recommend with Conditions N (%)	Do Not Recommend N (%)	Recusals
MUC2023-161	Appropriate Germline Testing for Ovarian Cancer Patients	Merit-based Incentive Payment System	Recommend with Conditions	6 (38%)	9 (56%)	1 (6%)	0
MUC2023-162	Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer	Merit-based Incentive Payment System	Recommend with Conditions	1 (6%)	11 (69%)	4 (25%)	0
MUC2023-190	Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer	Merit-based Incentive Payment System	Consensus Not Reached	0 (0%)	11 (69%)	5 (31%)	0
MUC2023-201	Cataract Removal with Intraocular Lens (IOL) Implantation	Merit-based Incentive Payment System	Recommend with Conditions	9 (56%)	5 (31%)	2 (13%)	0
MUC2023-205	Inpatient (IP) Percutaneous Coronary Intervention (PCI)	Merit-based Incentive Payment System	Recommend with Conditions	8 (50%)	4 (25%)	4 (25%)	0
MUC2023-203	Chronic Kidney Disease	Merit-based Incentive Payment System	Consensus Not Reached	1 (7%)	8 (53%)	6 (40%)	0

MUC ID	Measure Title	Program	Determination	Recommend N (%)	Recommend with Conditions N (%)	Do Not Recommend N (%)	Recusals
MUC2023-204	End-Stage Renal Disease	Merit-based Incentive Payment System	Consensus Not Reached	3 (20%)	3 (20%)	9 (60%)	0
MUC2023-206	Kidney Transplant Management	Merit-based Incentive Payment System	Consensus Not Reached	2 (14%)	4 (29%)	8 (57%)	0
MUC2023-207	Prostate Cancer	Merit-based Incentive Payment System	Consensus Not Reached	2 (14%)	2 (14%)	10 (71%)	0
MUC2023-208	Respiratory Infection Hospitalization	Merit-based Incentive Payment System	Consensus Not Reached	1 (7%)	5 (33%)	9 (60%)	0
MUC2023-209	Rheumatoid Arthritis	Merit-based Incentive Payment System	Do Not Recommend	1 (6%)	3 (19%)	12 (75%)	0

Note. Due to rounding, percentages may not sum exactly to 100.

MUC2023-137 Initial Opioid Prescribing for Long Duration (IOP-LD) [Pharmacy Quality Alliance (PQA)]

Description: The IOP-LD measure analyzes the percentage of Medicare Part D beneficiaries, 18 years or older, with at least one initial opioid prescription for more than 7 cumulative days' supply.

Program: Part C and D Star Ratings

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 1 (7%); Recommend with Conditions: 8 (57%); Do Not Recommend: 5 (36%)

Measure Discussion:

CMS Opening Remarks: CMS opened by noting that this measure is a plan-level retrospective performance measure across both Medicare Advantage and prescription drug plans. This measure does not penalize health plans or impede a patient's ability to get a prescription for opioids or get refills. This measure instead focuses on initial prescription duration and is evidence based. A greater supply of opioids is associated with adverse events such as long-term opioid use. Clinical practice guidelines also affirm that clinicians should prescribe no greater quantity than needed for the expected duration of pain severe enough to require opioids. The threshold for this measure is 7 days with an opioid initiation period of 3 days. CMS noted that some commenters requested a longer lookback period for some diagnosis codes related to exclusions; however, it is a common construct in quality measurement that a diagnosis code occurs within a given measurement year. This measure was reviewed by patients and caregivers in 2018 with 100% of the advisory panel supporting the measure. The measure is not intended to guide clinical decision-making for individual patients, and it does not represent a prescribing limit. Measures in Part D Star Ratings are usually limited to standardized claims and enrollment data. This measure was carefully developed to balance addressing opioid misuse without negatively impacting the patient-provider relationship. A coverage determination can be made for patients who need more than a 7-day supply of opioids. Medications for opioid use disorder are generally exempted. CMS has not identified any concerns with unintended consequences of the measure. The measure is not risk adjusted, as the developer did not identify a need to risk adjust. Additionally, this is a process measure, and these are generally not risk adjusted. The measure fills a gap in opioid safety in the Star Ratings program.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle reported that there were a number of public comments that did not support the measure. For example, a professional organization expressed their concerns that the measure was based on 2016 guidelines as opposed to the newly released 2022 guidelines. Another commenter called for an expansion of the exclusions for people who might be frail or have a disability. Commenters also expressed concerns about whether this measure would promote or limit equitable care. Comments in support of the measure included feedback from a professional organization that suggested the measure might help move people towards non-pharmacological interventions for pain. Additionally, several patients noted the importance of having a measure that promotes safe opioid use.

A committee member asked if any disparities were associated with this measure. The measure developer responded that they looked into stratifying the measure by Medicare status, age, and gender and did see that patients trended older. Another committee member asked how patients

being treated for opioid use disorder are impacted by this measure. The measure developer noted that methadone, Suboxone, and buprenorphine are not in the dataset used to calculate this measure. Another committee member asked what is being achieved by adding this measure to the Star Ratings. The measure developer challenged the assertion that different plans would not have different distributions and reported that they are seeing a lot of variation in this measure. They noted one in five prescriptions exceed the 7-day limit but there is variability in this when you look from plan-to-plan. CMS noted that Star Ratings is a pay-for-performance program so this incentivizes plans to improve in this area. A committee member expressed their organization's concern that the harms of this measure outweigh the benefits. They noted that CDC modified their recommendations around opioid prescribing because they believed their guidance was being misapplied. The committee member noted her organization believes this measure is creating harm for patients who need long-term opioids and that the exclusions do not go far enough.

Another committee member asked for clarification around the meaning of initial prescription, specifically, asking if the measure pertains to patients who were in the hospital but have been discharged and their provider is picking up where the hospital left off. The developer responded that the measure is based on Part D outpatient claims only and that the measure is a care coordination measure with safety overlaid. A patient representative on the committee wondered if patients who need opioids would be prevented from receiving them; this sentiment was echoed in the public comments. The developer responded that this measure is only for new prescriptions and is at the plan level. If a patient takes opioids in the preceding 90 days, the measure does not apply, and physicians are not penalized. Another committee member believed this measure would lead him, as a physician, to stop and think before refilling an opioid prescription. Another committee member indicated that the focus of the measure was important but noted that there was room for improvement, particularly given the concerns shared by an organizational representative on the committee. The organizational representative provided more context to their comments, stating that while the focus on opioid overuse is important, their concerns were related to the precision of and the evidence supporting the measure.

Conditions (if specified): None stated.

Future Directions:

While a majority of committee members voted to Recommend with Conditions, there were diverging perspectives on the measure; thus, the Recommendation Group did not reach consensus. The Recommendation Group acknowledged the importance of having a measure that assesses opioid prescriptions; however, they noted several concerns with this measure as specified.

MUC2023-179 Initiation and Engagement of Substance Use Disorder Treatment (IET) [National Committee for Quality Assurance (NCQA)]

Description: The percentage of new substance use disorder (SUD) episodes that result in treatment initiation and engagement.

Program: Part C and D Star Ratings

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 2 (14%); Recommend with Conditions: 8 (57%); Do Not

Recommend: 4 (29%)

Measure Discussion:

CMS Opening Remarks: CMS opened the discussion by noting this measure supports the Universal Foundation and seeks to close performance gaps. Only 36% of Medicare Advantage enrollees with new episodes have evidence of treatment engagement within 34 days of initiation. There are race and ethnicity differences in treatment and engagement. While risk-adjustment may be a strategy to mitigate discrepancies beyond the control of the measured entity, risk-adjustment may also mask differences in areas in which the health care system could potentially affect improvements. A prior version of the measure was endorsed in 2018 and re-evaluated in 2022. That version of the measure–which is the version being voted on by the Recommendation Group—will undergo full endorsement review in 2025.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle noted that committee members generally provided positive inputs for the measure in the round 1 evaluations with about 60% of those who returned evaluation rating the evidence for this measure as being complete and adequate. The public comments, on the other hand, raised more concern, the largest being around the accessibility of substance use disorder diagnosis information. Comments in support of the comment requested that telehealth visits for diagnosis and treatment be considered compliant for the measure. Battelle suggested that the discussion address whether there are enough eligible member to meaningfully report on the measure, the availability of treatment options across geographic regions, and the potential performance of the measure in low-volume or rural facilities.

The measure developer noted the intent of the measure remains the same, and the changes were made to better align the measure with its intent. A committee member shared that while their organization agreed with the intent of the measure, they had concerns about the availability of services, particularly in different communities and regions, that this is a claims-based measure, and that the measure does not capture patient choice.

Another committee member asked how 14 days and 34 days were selected, especially in considering clinical outcomes like relapse. The measure developer responded that those thresholds were mentioned in one state guideline but were mostly based on interested party expert consensus from their measurement advisory panels. The measure developer also noted that the measure utilizes a relatively narrow time window of assessment for treatment after initial diagnosis due to the high-risk of patients relapsing or falling out of contact with the health care system during this period. A committee member's organization was concerned that if plans require providers to submit electronic health record (EHR) data for this measure, it would increase burden on providers. The same committee member also noted that clinicians cannot force patients to get treated for substance use disorder. Another committee member countered, saying data issues are common and there is nothing specific about this measure that would lead to data issues. This committee member noted that telehealth services should be included, but that some states do not offer telehealth counseling for substance use disorder. They believed these challenges are similar across urban and rural areas.

Patient representatives on the committee were supportive of the measure. In response to a committee member question, CMS noted that it is getting the information for this measure through information reported by plans through their HEDIS submissions. Another committee member asked whether a patient who refuses is excluded. CMS stated that while patient refusal is a critical challenge not only for this measure but for other measures, providers and health plans are encouraged to identify ways to educate, communicate, and follow up with patients in

effort to minimize refusal and improve adherence to treatment recommendations. A committee member asked for a description of the changes that have been made to the current measure.

The measure developer reported that the changes made to the measure were to better align with its original intent, including extending the negative history period, and excluding denominator events in which an individual was solely undergoing medically managed withdrawal or receiving substance use treatment in the ED setting (i.e., if a patient is getting treatment for substance use disorder in an acute-care setting, they are excluded from the exclusions). Additionally, the receipt of pharmacotherapy as a standalone treatment would be considered sufficient for satisfying the numerator criteria of the measure. The measure is also now stratified by younger and older adults.

One committee member asked for clarification around same-day treatment events, and the measure developer confirmed that unless same-day treatment for substance use disorder was pharmacotherapy, same-day treatment satisfies numerator criteria if the services were provided by different providers.

Conditions (if specified): Undergo and receive consensus-based entity (CBE) endorsement.

Future Directions:

The Recommendation Group did not reach consensus on this measure and did not specifically identify future directions for the measure. Several committee members were interested in seeing this measure again pending the results of CBE endorsement.

MUC2023-212 Level I Denials Upheld Rate Measure [Federation of American Hospitals (FAH)]

Description: This rating shows how often a Medicare Advantage Organization review found their original determination decision to deny coverage to be reasonable.

Program: Part C and D Star Ratings

Committee Final Vote: Recommend

Vote Count: Recommend: 13 (87%); Recommend with Conditions: 1 (7%); Do Not Recommend: 1 (7%)

Measure Discussion:

CMS Opening Remarks: CMS opened the discussion by explaining that this measure is based on self-reported information submitted to CMS. The public comments focused on how this measure would potentially improve the quality of care with reduced unnecessary delays and increased transparency, while some commenters thought CMS should examine survey-specific information on prior authorization. Other concerns were around capturing proper initial denials that are overturned after receiving additional information, adding additional measures to the Part C and D Star Ratings programs that are not aligned across other CMS programs, and duplicating similar concepts in the current program. CMS clarified that this program is the only one that focuses on these types of appeals measures, therefore alignment is not something that needs to be addressed.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle noted that there was more consensus coming out of the round 1 evaluations. Committee members, clinical specialty societies, and interest groups were largely supportive of the measure, indicating that evidence was complete and adequate, and that this measure could improve quality of care by reducing unnecessary delays related to coverage denials and should increase the transparency around denials thus decreasing beneficiary frustration and provider burden. Concerns included added burden without perceived benefit, given the existence of the level 2 denials measure that is currently in the Star Ratings program, though some felt the measures would be complement one another. Given the concerns, Battelle wanted to hear comments from the patient perspective. A patient representative said that they saw a lot of face validity in the time savings and alleviating undue anxiety and delay for patients.

Committee Discussion: A committee member highlighted their organization's statement of support. Specifically, they thought the measure would reduce burden, improve transparency for patients and beneficiaries by providing denial information that could help in future health plan selections, could strongly complement the existing level 2 measure that is currently in the Star Ratings program, and could reduce frustration if it leads to the elimination of unnecessary prior authorizations with Medicare Advantage. The committee member noted that the measure had impressively high reliability. Several other committee members expressed their support for the measure, citing several reasons for their support including reduced provider burden and burnout. Additionally, a committee member noted that the measure is not duplicative since it is a different step in the process. The measure developer addressed a concern about data validation saying that, in the 2024 technical reporting requirements for Part C, all data submitted must be validated on the front end when plans submit the information to CMS.

Conditions (if specified): None stated.

Future Directions:

The committee reached consensus on this measure and voted to recommend the measure for inclusion in the Part C and D Star Ratings program.

MUC2023-164 Adult COVID-19 Vaccination Status [CMS]

Description: Percentage of patients aged 18 years and older seen for a visit during the performance period that are up to date on their COVID-19 vaccinations as defined by CDC guidelines on current vaccination.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 0 (0%); Recommend with Conditions: 5 (29%); Do Not Recommend: 12 (71%)

Measure Discussion:

CMS Opening Remarks: CMS opened the discussion by stating that COVID-19 remains a very important clinical condition, with numbers increasing during the winter season, and noting that ample evidence shows that the vaccine is safe and effective and decreases hospitalization and mortality. CMS has worked closely with the CDC on the measure and the definitions. As a

leading cause of death with over 100 deaths per day and climbing, vaccination rates remain low and that underscores the importance of this measure. CMS is considering this measure for use in MIPS to assess the percentage of adult patients who have been vaccinated and are up-to-date according to CDC guidelines (meaning aligned with CDC recommendations at the time of the most current vaccination). This measure excludes those who have a medical contraindication but not patient refusal for personal reasons. This was purposefully done to identify the true vaccine rate and best protect public health. CMS believes that clinicians have a vital role and a strong voice in encouraging patients to get vaccinated. This measure aligns with existing COVID vaccination measures.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle noted a few concerns emerged from the public comment and round 1 evaluation. Specifically, committee members and commenters expressed concerns about what it means for someone to be up-to-date and how quickly that changes. They also questioned where the measure would create harm in areas where there is more vaccine hesitancy or where clinicians might have more difficulty getting feedback from patients if they got vaccinated elsewhere. The feedback acknowledged the problem COVID-19 continues to pose and the importance of understanding how healthy communities are.

During discussion, concerns were brought up about the changing CDC recommendations and disparities based on geography, political affiliation, ethnicity, and income. Committee members agreed with CMS that physicians have a huge impact on vaccination rates. A committee member asked about the potential to implement COVID as part of the adult immunization status measure. CMS responded that if the COVID vaccine becomes an annual vaccine, it could possibly be implemented as part of the adult immunization measure in the future. A committee member expressed their organization's strong support of this measure within the MIPS program but would like to see another vote if it becomes part of the Medicare Shared Savings program. CMS confirmed another vote would happen if that were the case. A patient representative shared their support of the measure, acknowledging the noise and unpopular opinions around the topic of COVID vaccines. Battelle noted that the patient's perspective is important especially if they might face extra harms in the event of an outbreak.

A committee member shared a concern about the measure being at the physician level instead of at the health plan or system level to capture the bigger picture. They indicated that the measure does not provide useful information for physicians, and expressed their doubts that this measure will move the quality needle. CMS noted that physicians can garner valuable information by asking their patients whether they were vaccinated and by looking at their results in comparison to the results within the community they work. A committee member shared their organization's strong support for all vaccines; however, they do not support this measure for accountability at the physician level. They noted that a CBE endorsement would highlight some of the concerns expressed by others. They also highlighted the cost of vaccine delivery, which can prevent some primary care clinics from administering the vaccine in their practices. In addition, the same committee member touched upon the growing misinformation and hesitancy among some populations, which would make it unfair to hold clinicians accountable. They also discussed data challenges.

Conditions (if specified): Undergo and receive CBE endorsement.

Future Directions:

The committee did not reach consensus to recommend the measure for inclusion in MIPS, with most members voting "Do not recommend," largely due to concerns around holding clinicians

accountable for their patients' COVID vaccination status given the widely acknowledged impacts of vaccine hesitancy specific to COVID.

MUC2023-211 Melanoma: Tracking and Evaluation of Recurrence [American Academy of Dermatology (AAD)]

Description: Percentage of patients who had an excisional surgery for melanoma or melanoma in situ with initial American Joint Committee on Cancer (AJCC) staging of 0, I, or II, in the past 5 years in which the operating provider examines and/or diagnoses the patient for recurrence of melanoma.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 4 (24%); Recommend with Condition: 6 (35%); Do Not Recommend: 7 (41%)

Measure Discussion:

CMS Opening Remarks: CMS stated that this measure addresses a high-priority area as an outcome measure addressing care coordination. Evidence suggests a lack of communication between the excising provider and the provider continuing care. This measure will allow for the development of a system in which melanomas can be accurately tracked to increase the understanding of the effectiveness of care as well as to make sure that patients get timely follow-up. This measure fills a current MIPS quality measure inventory gap within the dermatological clinical topic and could potentially replace the current MIPS 137 Melanoma Continuity of Care measure.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle reported that the only public comment submitted for this measure was in support of it. In the round 1 evaluation, committee members indicated that there was a good measure concept; however, they expressed concerns about the burden of tracking and report. They also expressed concerns about how the measure would be tracked by socioeconomic status as that might lead to increased screening. A committee member noted that the measure seemed like a patient-centered care measure that prioritized patient outcomes.

A committee member shared their concern about the reliability of the measure. The measure developer reported that the reliability results were due to a low sample size. Based on the wide confidence interval, they believed that reliability would improve with a larger sample size. Committee members expressed concerns about the burden of tracking and reporting. One committee member questioned whether recurrence is the best way to track outcomes or if it is actually better to assess the number of people who are getting annual skin exams after melanoma. CMS indicated that both are important to track. One committee member stated that this measure felt patient-centered. CMS added that on face validity, a patient who had a melanoma excised would likely seek yearly follow-up to ensure it had not recurred. A committee member questioned whether it is feasible for a clinician to collect data and report on the measure. Another committee member echoed this concern and listed several scenarios that might make data collection challenging, for example, the patient moving or a different provider removing the melanoma. The measure developer clarified that "operating provider" allows for

another provider and pointed out that this measure has complementary measures. They also reminded the group that this measure already has been operationalized for two years, and it is possible to update the data, specifically the reliability and validity.

Conditions (if specified): Undergo and receive CBE endorsement; reliability testing with a larger sample size; additional validity testing.

Future Directions:

The Recommendation Group did not reach consensus on this measure. Some committee members noted that the measure should undergo the endorsement process before being considered for use in a CMS program. Whether undergoing CBE endorsement or not, committee members recommended the measure undergo additional reliability and validity testing with a larger sample size to better establish whether it would be appropriate for use in MIPS.

MUC2023-141 Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy [Society for Immunotherapy of Cancer (SITC)]

Description: Percentage of patients aged 18 years and older, with a diagnosis of metastatic non-small cell lung cancer or squamous cell carcinoma of head and neck on first-line immune checkpoint inhibitor (ICI) therapy, who had a positive PD-L1 biomarker expression test result prior to giving ICI therapy.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Recommend with conditions

Vote Count: Recommend: 6 (38%); Recommend with Conditions: 10 (63%); Do Not Recommend: 0 (0%)

Measure Discussion:

CMS Opening Remarks: CMS opened the discussion of this measure by explaining that the PD-L1 biomarker appears on some cancers but not all. First-line immune checkpoint inhibitor therapy targets this biomarker, but it is very expensive. This measure seeks to ensure that patients are tested for this biomarker before receiving the therapy.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle noted that of the two public comments received for this measure, one supported it while the other supported it with conditions. Feedback from the round 1 evaluations indicated that this measure is important in improving the equitable state of care for patients. Other feedback shared concerns about the burden of tracking and reporting associated with the measure.

The committee members had several questions about this measure. First, they asked whether the population base was large enough to have a sufficient denominator to be able to compare performance between providers or populations. CMS responded that there are sufficient case numbers to compare between populations. Another committee member noted that while the overall measure reliability was high (greater than 0.9), the interrater reliability was low (less than 0.5) and wondered why there were disparities in performing this testing. The measure developer responded that the testing is already being conducted by major academic health systems, but at rural hospitals, it may not be. The goals of the measure are to educate providers about testing

for this biomarker before therapy is offered and to increase access to immune checkpoint inhibitor therapy. The committee member was satisfied with this answer and noted that understanding what the measure hopes to achieve in the real world is very important. A

Another committee member stressed the importance of this measure for reducing disparities and wondered if anyone had data on the disparities for this therapy. The developer responded that controlling costs is especially important for those in rural areas, but they do not have data to show immunotherapy is prescribed less frequently in rural areas, although anecdotally they believe that this is true. The committee member encouraged exploring disparities between racial and ethnic minorities as well.

CMS noted that the knowledge around immunotherapy is expanding rapidly and that clinicians might be having a hard time keeping up with the latest science. Additionally, if a patient is dying, clinicians might prescribe a therapy that is not appropriate for that patient's cancer, just to try something. One committee member asked if conditions around feasibility and specificity could be placed before the measure is adopted by the program. The developer noted these comments were focused on side aspects of the measure rather than the intent of the measure. This measure was recently developed so it is not endorsed. However, the measure was developed based on oncology guidelines, which are consensus based. Feasibility and validity testing also achieve a variety of checkpoints for statistical relevance.

Conditions (if specified): Additional testing requested for examining measure performance and feasibility; additional testing to assess disparities in performance by race/ethnicity.

Future Directions:

The Recommendation Group reached consensus on this measure and voted to Recommend with Conditions. Notably, no committee members voted "Do not recommend." Committee members strongly supported the intent of the measure but sought more evidence to support its feasibility and usability across disparate populations.

MUC2023-161 Appropriate Germline Testing for Ovarian Cancer Patients [American Society of Clinical Oncology (ASCO)]

Description: Percentage of patients, aged 18 and older, diagnosed with epithelial ovarian, fallopian tube, or primary peritoneal cancer who undergo germline testing within 6 months of diagnosis.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Recommend with conditions

Vote Count: Recommend: 6 (38%); Recommend with Conditions: 9 (56%); Do Not Recommend: 1 (6%)

Measure Discussion:

CMS Opening Remarks: CMS opened the discussion on this measure by explaining that some genetic mutations in the germline (e.g., sperm and eggs) can predispose someone to cancer. The most well-known of these is the BRCA gene, which predisposes individuals to breast and ovarian cancer. The measure asks clinicians to consider germline testing in epithelial ovarian cancer patients. The measure is not endorsed but has met the statutory requirement for

inclusion in MIPS. It addresses the CMS priority of allowing for more personalized diagnostic strategies for patients. The measure fills a gap in MIPS in the oncology specialty set and may be considered for the Advanced or Cancer Care MIPS Value Pathway. The developer noted that exclusions for the measure include patients who have had germline testing before their diagnosis of ovarian cancer.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle reported that three public comments were received for this measure: two in support and one in support with conditions, specifically around CBE endorsement. Concerns discussed in the round 1 evaluation included reliability testing meeting thresholds, tracking socioeconomic status, and the burden associated with tracking and reporting. A member of the Advisory Group noted that the measure will close a gap in addressing this underutilized biomarker while another member indicated that the benefits of the measure were greater than the burden associated with it. A committee member indicated that genetic testing is not widely used in minority areas so this measure would help in making it more equitable.

One of the committee members noted that the validity testing is hovering around 50% and they prefer 70-75%. The same committee member was concerned that the measure was only tested in 12 practices. CMS noted that it is getting harder to test measures. They theorized that this might be related to survey fatigue or because clinicians do not have time for additional tasks. The developer agreed that recruiting practices is very labor intensive and costly, especially if a developer does not already have a dataset to work with.

Another committee member asked how specialty societies benefit when developing a measure. CMS and Battelle responded that the societies aim to improve care in their fields; another committee member noted the financial benefits to developers if their measures are adopted. A condition of this measure's adoption would be endorsement by a consensus-based entity. The measure developer noted that empiric validity was 0.78, adding that 12 practices, 32 physicians, and approximately 300 patients participated in testing for the measure. The overall performance score of the measure was 0.52. The reliability score for individual physicians was 0.86 and for group practices, it was 0.89 (well above the minimum reliability threshold required for MIPS). A committee member noted that it would be helpful for them to have access to more data. A patient representative on the committee encouraged other members to think about the impact to patients on a broader scale and to consider equity more in these discussions. A committee member asked how this measure promotes equity and Battelle summarized what had been previously described, stating measures like this might promote equitable care because clinicians must test all people who need it, not just those who they think might need it. CMS noted that the burden for germline testing is low, such as simple swab of the cheek sent to a lab. Lastly, a committee member noted it is difficult to evaluate the fit of a measure for a program if the measure has not been endorsed.

Conditions (if specified): Undergo and receive CBE endorsement.

Future Directions:

The committee reached consensus on this measure and voted to Recommend with Conditions. Committee members suggested that the measure undergo the endorsement to ensure that it has strong scientific acceptability results across different practice sizes.

MUC2023-162 Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer [Purchaser Business Group on Health (PBGH)]

Description: The patient-reported outcome performance measure (PRO-PM) will assess pain interference following chemotherapy administered with curative intent to adult patients with breast cancer.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Recommend with conditions

Vote Count: Recommend: 1 (6%); Recommend with Conditions: 11 (69%); Do Not Recommend: 4 (25%)

Measure Discussion:

Note: MUC2023-162 and MUC2023-190 were combined for the purposes of the discussion, given the similarities between them (see below for the measure description of MUC2023-190).

CMS Opening Remarks: CMS opened this discussion by noting these measures (MUC2023-162 and MUC2023-190) are endorsed. Many patients who go through chemotherapy have chronic pain or fatigue afterwards and this is not being quantified. A criticism of this measure is that it uses a single-page questionnaire, the PROMIS Tool. Commenters noted that other valid tools should be considered, and reporting can be fragmented if different tools are used. CMS will consider how to bring unity to this area. This measure fills a gap in understanding a patient's experience of care and fills a gap in the oncology clinical area of MIPS. These are the first outcome, specialty-specific oncology measures to address patient experience of care and could be considered in the MIPS Cancer Care Value Pathway.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle noted that there was nearly universal agreement from commenters that patient-reported measure are a good thing and that the focus of these two measures and cancer care broadly, is a good use of this kind of measure. The primary concerns identified in the feedback were related to feasibility; specifically, the burden to track and report associated with the measure and limitations of the measures due to their exclusive use of the PROMIS tool. Several specialty societies and representatives from cancer centers shared their support of the measure.

A committee member asked why the PROMIS Tool was used over other tools or if a similar tool could be used to reduce the burden of switching tools. The developer responded that they evaluated 13 different instruments and worked with a technical expert panel to narrow that list. They chose this tool specifically because it is already cross walked with the Functional Assessment of Cancer Therapy tool and can be cross walked with other instruments as well. The developer's patient and caregiver council also recommended PROMIS, as it is freely available and free to use. CMS agreed that an advantage of PROMIS is that it is non-proprietary. Another committee member asked about the literacy level of PROMIS. The developer responded that the tool has been tested for literacy amongst the general population and among cancer patients. The reliability of the tool has also been tested beyond the measure itself. PROMIS is also short, at four questions total. CMS added that the PROMIS tool came out of a decade-long National Institutes of Health (NIH) effort.

A committee member asked about the rate of refusal to take the survey and if certain groups were more likely to refuse. The developer responded that data collection was completed during the pandemic. They had hoped to administer the survey in person, but COVID made this challenging so the response rate was not as high. However, the developer still obtained

sufficient data to test and get statistically reliable results. Ultimately, the refusal rate was low. The developer did not see a pattern regarding certain groups refusing more than others.

Another committee member asked whether these measures were truly endorsed. Battelle clarified they are endorsed at the group level, and that they went through the National Quality Forum (NQF) endorsement process carried out by Battelle during the transition period. The member had concerns about low reliability at the individual clinician level and asked about the validity. The developer confirmed the validity is reported at the group level. The committee member reported that they have never seen a measure put forward in MIPS that does not perform well at the individual level. CMS responded that they will take this under consideration. Battelle noted that members could recommend with conditions, such that the measure would only be used at the group level. Another member noted that they did not believe this could be a condition and they would vote “do not recommend,” because they had concerns that CMS would have to implement the measure at the individual level. However, CMS confirmed later in the discussion that there was precedent for using measures at only the group level, and that as such it was a condition that the committee could recommend.

Another committee member asked about the minimum number of surveys needed to calculate the measure. The developer responded that they would need 23 surveys per group for a reliability score over 0.7. The survey response rate was 40%. The developer added that, in the future, the response rate is expected to be higher, because the survey was mostly done by mail and phone when it would normally be done in the office. Committee members were confused about what the reliability scores were, and Battelle responded that whatever scores were provided in the MUC submission were passed on to committee members and confirmed that reliability at the group level is 0.7 with an average sample size of 32 patients per group. The developer noted that their understanding is that CMS would implement this measure at the group level because they do this for cost measures. At the individual level, the reliability of the pain measure is 0.42 and fatigue is 0.39. The data collected was limited due to COVID, but the developer expects the reliability to increase over time. As stated above, CMS chimed in to indicate that it does have discretion to make determinations on which level a measure would be used in certain cases. For example, a Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey for MIPS that is group only. If members wanted to place this as a condition of use for this measure, that would be appropriate. A member responded that they had seen other measures be declined for having poor individual scores. A committee member asked how this measure could be applied only at the group level in practice. CMS noted that 85% of clinicians report as groups, although they have traditionally taken the stance that a measure needs to be tested with individuals because this is how the program is run. Additionally, CMS noted that, in some measures, only groups are allowed to report, such as the CAHPS for MIPS survey measure. CMS can put parameters in place for who could report a measure.

A member also asked about the desired effect of this measure and about data missingness. The measure developer responded that the rationale for the measure is for providers to reduce patients’ symptom burden. The numerator is absolute health status related to pain and fatigue so providers can see how they perform compared to others and improve if needed. Completion rates of the survey were not different based on race and ethnicity. The PROMIS tool is also integrated into several EHRs, and practices are able to use tools like REDCap to collect the data. Computer-adapted testing of the tool is also available. A patient representative on the committee noted that they receive IVIG medicine every month and their answers on the survey would change depending on when they took the survey. But they maintain the reporting of this information is a step in the right direction for patients, especially if more patients take the survey

and the measure is refined over time. The developer closed by saying that patient-reported measures serve to amplify the patient voice.

Conditions (if specified): Implementation at the clinician group-level only until further testing and improvements can be made to scientific acceptability at the individual clinician level.

Future Directions:

Consensus was not reached for this measure. Committee members suggested that CMS consider the fact that the measure performs well at the group-level but poorly at the individual level when making the decision about whether to include it in MIPS.

MUC2023-190 Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer [PBGH]

Description: The PRO-PM will assess fatigue following chemotherapy administered with curative intent to adult patients with breast cancer.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 0 (0%); Recommend with Conditions: 11 (69%); Do Not Recommend: 5 (31%)

Measure Discussion: MUC2023-162 and MUC2023-190 were combined for the purposes of the discussion. Refer to the measure discussion for MUC2023-162 above.

Conditions (if specified): See MUC2023-162

Future Directions: See MUC2023-162

MUC2023-201 Cataract Removal with Intraocular Lens (IOL) Implantation [CMS]

Description: Evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who undergo a procedure for cataract removal with IOL implantation. This procedural measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during each cataract removal episode from 60 days prior to the clinical event that opens, or "triggers," the episode through 90 days after the trigger.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Recommend with Conditions

Vote Count: Recommend: 9 (56%); Recommend with Conditions: 5 (31%); Do Not Recommend: 2 (13%)

Measure Discussion:

CMS Opening Remarks: CMS opened discussion of this measure by reminding the committee that cost measures are statutorily required for MIPS and must cover 50% of Part A and B spending. The cost performance category must also be 30% of the MIPS final score. Of the over 200 measures in MIPS, only 29 are cost measures. These are episode-based cost measures that are based on Medicare spending for a well-defined episode of care and are risk adjusted. The case minimums for episode-based cost measures are typically 20 episodes for acute or chronic conditions and 10 episodes for procedural measures. MIPS uses a 0.4 reliability threshold for adding measures to the cost performance category. This standard was developed in 2017 and was reaffirmed in 2022 with scientific evidence. All episode-based cost measures presented during the Recommendation Group meeting are tested at the individual clinician and clinician group level. Measure 201 is a revised version of the CBE-endorsed routine cataract removal with IOL implantation measure that has been in MIPS since 2019. Cataract removal is a very common procedure and is costly to Medicare. Many aspects of the original measure are being retained, but this version expands the patient cohort by no longer excluding patients with glaucoma or diabetic eye disease. The new measure also includes the cost of telehealth services, lenses, frames, and emergency department visits for ocular complaints. The new measure also adds new Part B medications. The new measure is risk adjusted, including for the new conditions that were previously excluded. Traumatic cataract is still an exclusion. Clinician-level reliability is 0.96 and group-level reliability is 0.97.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle reported of the six public comments received for this measure, two were in support of the measure and four were opposing. Advisory Group members expressed concern that the measure might disproportionately affect populations that have small patient sizes. They shared an additional concern that the measure might steer clinicians away from more complex patients if they felt they would reimbursed at the same level as they would for less complex patients. A specialty organization echoed this concern in their public comment, stating low-volume cataract practices that provide care to patients with complex conditions would have greater exposure to receiving a penalty. Cataract surgeons in particular would be inappropriately penalized because they see the majority of complex cases.

A committee member noted that hopefully the risk adjustment will limit cherry-picking and lemon-dropping of patients, but they were skeptical about the stability of this measure year over year, especially for small clinics. Another committee member wondered if low-volume cataract surgery providers have worse outcomes when compared to high-volume surgery providers. If so, this would disincentivize low-volume clinics with poor outcomes. The developer responded that performance looks similar across clinics regardless of how many cataract surgeries are performed. Another committee member asked if any testing was done on how social determinants of health affect the measure. The developer responded that they examined whether performance changes based on percentage of dual enrollees, and they did not note any differences. Performance is also similar among rural and urban practices. The committee member noted that other social determinants may affect performance on the measure.

A committee member noted that in other MIPS cost measures, variation exists between dual enrollees and non-dual enrollees. Much of this variation comes from Part B medications and prescribing around chronic conditions. Clinicians do not have the control to mitigate these costs if they are treating more complex patients. The developer responded that their clinician workgroup identified drugs that are clinically related to cataract surgery. The measure is risk adjusted for ocular conditions that make it more appropriate to use the list of specified drugs. However, use of these drugs is consistent across less risky and more risky episodes. The

episode covers costs prior to and after the procedure, and use of these medications can mitigate complications that might land a patient in the emergency room. The developer was asked to capture more variation in clinician performance, so they added emergency department visits for ocular pain after the procedure.

Another committee member asked whether the measure will disincentivize providers from performing the procedure. The developer re-emphasized that they do not see variation in performance based on the volume of procedures performed. Since the original measure was implemented in 2019, they have not seen the overall number of procedures performed change over time. A committee member noted that MIPS adjustments may not affect practice patterns because the adjustments are small.

Conditions (if specified): None stated.

Future Directions:

The Recommendation Group reached consensus and voted to Recommend with Conditions. The committee raised concerns related to cost measures in general. Specifically, committee members sought more information about how the implementation of cost measures may impact patient outcomes.

MUC2023-205 Inpatient (IP) Percutaneous Coronary Intervention (PCI) [CMS]

Description: Evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who present with a cardiac event and emergently receive PCI as treatment. This acute inpatient medical condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during each episode from the clinical event that opens, or "triggers," the episode through 30 days after the trigger.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Recommend with Conditions

Vote Count: Recommend: 8 (50%); Recommend with Conditions: 4 (25%); Do Not Recommend: 4 (25%)

Measure Discussion:

CMS Opening Remarks: CMS began the discussion by noting that this is a revised version of the measure that is currently in use in MIPS and has been since 2019. During the re-evaluation, CMS carefully considered public comments regarding risk adjustment and exclusions along with input from a clinical expert work group. The revised measure stratifies episodes based on the diagnosis associated with the PCI procedure and does risk adjust for factors such as cardiomyopathy, multivessel coronary artery disease, pulmonary hypertension, and recent cardiac arrest. The measure has specific exclusions such as cardiac arrest during the initial hospitalization and prior hospitalizations. CMS clarified that Part D costs are not included in the measure and that testing showed rural and urban providers performing similarly, showing no disadvantage for rural providers.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle indicated that of the two organization that provided public comment on the measure, one opposed the measure and the other supported it with conditions including improved reliability testing, adding social risk factors

to the risk adjustment model, testing to evaluate costs, and ensure that robust test was conducted at the individual level. In the round 1 feedback, a patient on the Advisory Group supported the measure because it focuses on an important aspect of care and reducing cost is crucial. Another reviewer noted that the expanded population for the measure seems appropriate given that more patients undergo PCI than those who present with an acute STEMI.

A committee member expressed their organization's concern with the reliability level. Specifically, they shared their belief that reliability needs to meet a high threshold; thus, they consider CMS' minimum threshold of 0.4 to be too low. CMS pointed out that the reliability for clinicians was 0.52 and 0.63 for groups at a 20-episode case minimum so they exceeded the threshold established through rulemaking. CMS also confirmed that this and all the cost measures being reviewed on Day 2 of the meeting are at the individual- and group-levels. The measure developer added that they have undergone extensive engagement with specialty societies in both the development of the original measure and in the development of the current reevaluated measure.

A committee member speaking from the patient perspective recognized that the information the committee was discussing was complex and challenging. However, they indicated they had no issues following the documentation provided as well as the meeting discussion.

A committee member asked CMS to provide additional information about their reliability threshold and the implications for the intended function of the measure in plain language. The measure developer responded to this question and provided clarity on how they define reliability. Battelle clarified that while committee recommendations regarding the reliability thresholds had been noted, the committee was charged with evaluating measures against the established criteria used in the program (meaning a minimum reliability threshold of .4); any changes to those requirements would have to happen over time or via a separate rulemaking process and could not be addressed in this forum.

A committee member inquired whether there were significant differences in cardiac level of care between PCI providers and non-PCI providers and the implications in situations where patients need to be transferred to a PCI-performing facility. Another committee member added to the question and asked about attributions to both institutions. The measure developer stated that the measure is going to be attributed to groups and clinicians that are involved with the hospital stay for the inpatient PCI. They explained that when a transfer is involved, those cases would be excluded, and providers would be stratified by the complexity of patients.

Conditions (if specified): Undergo and receive CBE endorsement; skeptical of year-to-year stability so would like to see results over time.

Future Directions:

The Recommendation Group voted to Recommend with Conditions. Committee members suggested that the measure undergo the endorsement process where the scientific properties of the measure could be assessed with rigor. A committee member noted that they would like to see longitudinal data to assess the stability of the measure.

MUC2023-203 Chronic Kidney Disease [CMS]

Description: Evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for patients who receive medical care to manage and treat stage 4 or 5 chronic

kidney disease (CKD). This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Chronic Kidney Disease episode.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 1 (7%); Recommend with Conditions: 8 (53%); Do Not Recommend: 6 (40%)

Measure Discussion:

CMS Opening Remarks: CMS stated that this measure, together with the End-Stage Renal Disease (ESRD) and Kidney Transplant Management measures, fill cost measure gaps in MIPS and have the potential to create a MIPS Value Pathway for kidney care. They reported that the measure was developed with extensive input from clinical experts and persons with lived experience. CMS summarized comments they received about the measure, including concerns about rural providers being disadvantaged. They indicated that they tested this and found that rural and urban providers perform similarly.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle indicated that there were six public comments for this measure, all of which came from the professional organization and specialty societies one would expect to provide feedback on the measure. Support for the measure was primarily related to the value of measurement in reducing healthcare costs as well as some specific support for certain measure exclusions. Concerns for the measure were related to the impact of the measure on rural practices and scientific acceptability. One specialty society shared their perspective that the low reliability of the measure might be linked to the fact that comorbidities tend to be on the problem lists and therefore, not captured correctly. Battelle noted there were some comments about the measure exclusion of patients with extremely low treatment costs with concerns surrounding what the threshold for extremely low treatment cost was and why those patients were excluded. A professional organization had questions related to attribution which Battelle noted was an artifact of them thinking the measure had only been tested at the group level; however, the measure was actually tested at both the group and individual levels.

A committee member shared their organization's comments on the measure, indicating that they recognized the importance of working to improve affordability and efficiency and appreciated the efforts to field test the measure. However, they stated that they shared the concerns expressed by other stakeholders and could not support the measure for several reasons. First, they indicated, the measure has only been tested by the measure developer. They shared their organization's position that the measures should go through a rigorous endorsement process and be further tested and refined before considered for implementation. Secondly, the measure appeared to have been tested at the clinician group level, and they do not support use of this measure at the individual clinician level. The committee member indicated that for the denominator exclusions, their organization recommends defining what constitutes extremely low treatment costs. They also stated that the measure developer did not supply evidence showing mechanisms by which clinicians can improve chronic kidney disease care and slow progression of ESRD through implementation of cost-containment measures. Additionally, the committee member stated that an entity does not have a clear path forward as to how they may improve performance on the measure's focus. They indicated that it is important to recognize that the physicians cannot control drug costs or coverage of drugs needed to treat patients with chronic kidney disease or to slow the progression in the ESRD. They noted that not all communities

have access to nephrologists so patients with chronic kidney disease would likely be attributed to a primary care physician in those communities. Another reason the committee member's organization was not in support of the measure is because its reliability was below 0.6, which they believe should be the minimum reliability threshold (Note: this is not aligned with the requirements for inclusion in MIPS). They suggested that CMS would need to increase the case minimum to improve reliability. Lastly, they stated that the measure's risk-adjustment model uses CMS-HCC Version 24; however, the committee member suggested CMS update the specifications because, as of the start of the year, they have been asked to use CMS-HCC Version 25, no. 28. Battelle clarified that the measure is evaluated at the individual and group level.

CMS highlighted the fact that extremely low-cost episodes are only those episodes where the cost is less than one evaluation and management equivalent services episode. The measure developer provided information about the reliability of the measure, stating reliability increased as the case minimum increased. They indicated that they have other considerations related to the measure specifications to more aggressively handle outliers. The measure developer reported that there were no significant differences in average performance measure scores between urban and rural providers. They also looked at performance scores by different specialties and found that at the group level, nephrology was the most attributed entity. The measure developer responded to the issue of providers potentially avoiding complex patients and indicated they do not see systematic evidence of this.

A committee member noted that risk adjusting a measure based on diagnosis codes is contingent on the on the provider appropriately billing those codes They indicated that doctors do not always enter a complete list of relevant diagnosis codes in the EHR and they may be judged badly for not doing so. Another committee member pushed back, from a primary care perspective, on the idea of physicians cherry picking and dropping patients. They noted that primary care is increasingly being bought up into large health systems so it is often the system, not the physicians themselves, dropping patients for not following guidelines or their physician's recommendations. The systems are looking at quality scores and dropping patients who pull down those scores. A committee member expressed their concern that that approach places the blame on the patient.

Another committee member noted that physician practices are being consolidated in locations which might be inconvenient to get to, making it challenging for socially disadvantaged or vulnerable patients to access care. They also stated that there have historically been disparity issues with the algorithms for placing patients on the kidney care transplant lists and it is unclear how patients impacted by those outdated algorithms might impact the data. The committee member indicated that CMS seems to be changing their policy on the case minimum to add the measure into the program. CMS responded, stating that they have been trying to change all the online calculators and make sure the appropriate calculators are available to calculate GFR. The measure developer noted that the CKD and ESRD measures have built in incentives to help promote transplantation. It tries to ensure that providers are incentivized to conduct transplant-related activities.

Conditions (if specified): Undergo and receive CBE endorsement; use at group-level (and not the individual clinician level).

Future Directions:

The Recommendation Group did not reach consensus on this measure, though a slight majority voted to recommend with conditions, specifically that the measure undergo the CBE endorsement process. They also indicated that the measure should be used at the group-level as opposed to the individual-level.

MUC2023-204 End-Stage Renal Disease [CMS]

Description: Evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for patients who receive medical care to manage End-Stage Renal Disease (ESRD). This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during an ESRD episode.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 3 (20%); Recommend with Conditions: 3 (20%); Do Not Recommend: 9 (60%)

Measure Discussion:

CMS Opening Remarks: CMS began by pointing out that this measure was developed through the same process as the Chronic Kidney Disease measure (MUC2023-203) to ensure that the measures would work together cohesively to assess kidney care in MIPS or as a MIPS Value Pathway. Like CKD, this measure has a robust risk-adjustment model that includes CMS hierarchical condition categories and ESRD-specific variables such as crash starts to dialysis. It also risk adjusts for patients who are dually enrolled in Medicare and Medicaid. CMS reiterated that the MIPS program uses 0.4 as the threshold for reliability. As part of rulemaking, CMS considered the tradeoffs between reliability and case volume as well as between reliability and validity and how each relates to the overall goals of the program. For example, if a measure's reliability did not meet that 0.4 threshold, CMS can set a higher case minimum, using it at only the group level, or make other minor adjustments that improve the statistical reliability.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle explained that the public comment and round 1 evaluation feedback was similar to what was received for the CKD measure, with about a 50/50 split in terms of support. The support was based on the intent of the measure. Battelle noted that reliability testing was a challenge and there were questions around appropriate risk adjustment. The question of equity among Black patients was raised because an old algorithm was used to determine risk of chronic kidney disease severity and eligibility for transplant that systematically underestimated severity within that population.

The measure developer acknowledged the concerns around the equitability of the measure and emphasized the importance of care coordination between all specialties to address the equity issues. A committee member shared their organization's feedback that the measure was not CBE endorsed and had not undergone that rigorous review process. Committee members expressed concern that testing had only been done at the clinician group level and not the individual level.

Conditions (if specified): None stated.

Future Directions:

The committee did not reach consensus to recommend the measure for inclusion in MIPS, with a small majority voting “No not recommend” outright.

MUC2023-206 Kidney Transplant Management [CMS]

Description: Evaluates a clinician’s or clinician group’s risk-adjusted and specialty-adjusted cost to Medicare for patients who receive medical care related to kidney transplant, beginning 90-days post-transplant. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician’s role in managing care during a Kidney Transplant Management episode.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 2 (14%); Recommend with Conditions: 4 (29%); Do Not Recommend: 8 (57%)

Measure Discussion:

CMS indicated that this measure, along with the CKD and ESRD measures, address a measurement gap in MIPS with regards to costs for ongoing kidney transplant-related care and management. The measure includes risk-adjustment variables related to the transplanted organ (e.g., whether the donor is living or deceased), the kidney donor profile index, and whether the kidney was from a blood-type-incompatible donor. The measure also adjusts for dual enrollments. These adjustments help ensure that the measure is accounting for higher-complexity and higher-cost patients. CMS indicated that this measure meets or surpasses the mean reliability threshold of 0.4.

A committee member noted that generally, cost measures are important to have when they are appropriate. They asked whether there was a reason these types of cost measures, the present measure as well as the ESRD and CKD measures, were not endorsed by a CBE. CMS responded, indicating that MIPS does not require endorsement of measures to be included in the program. They stated that in the past, cost measures have gone through the endorsement process after their inclusion in the program. The committee member shared that many of the committee’s concerns would have been vetted through the endorsement process and this would have made them feel more confident about whether or not to include measures in the program. Another committee member echoed this concern, indicating that they felt increasingly uncomfortable about making recommendations without adequate information. They agreed that a process like endorsement would help assure them that the measure had been adequately evaluated.

The measure developer shared feedback from the workgroup who reviewed this measure. They clarified that the measure focuses on the maintenance stage, 90 days following transplantation, to ensure that they are appropriately managed given the high cost and patient consequences that are associated with complications. The measure developer indicated that the workgroup debated the use of immunosuppressives given the lack of scientific consensus on which are appropriate. They also noted the importance of considering diabetes and its interactive effects with immunosuppressives. The measure developer noted that the transplant center conducted a

provider-based analysis and they were not penalized more than other providers, suggesting that the risk adjustment addressed complexity of care.

A committee member asked whether there was a stipulation related to a minimum number of patients as the reliability testing indicated that the reliability of the measure increased based on the number of participants. CMS indicated that they did take the case minimum into account and noted there were tradeoffs with increasing that minimum. They are statutorily required to cover 50% of the spending so they want to cover enough spending while also recognizing the increase in reliability with an increased case minimum.

Conditions (if specified): Undergo and receive CBE endorsement; larger minimum case size.

Future Directions:

The Recommendation Group did not reach consensus for this measure. Committee members expressed their discomfort with making recommendations on measures about which they felt they did not have enough information. They suggested that measures undergo endorsement prior to their consideration to ensure that any concerns regarding their scientific merits are adequately evaluated prior to the review of the Recommendation Group.

MUC2023-207 Prostate Cancer [CMS]

Description: Evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for patients who receive medical care to manage and treat prostate cancer. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Prostate Cancer episode.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 2 (14%); Recommend with Conditions: 2 (14%); Do Not Recommend: 10 (71%)

Measure Discussion:

CMS Opening Remarks: CMS indicated that the overarching theme with prostate cancer is there are many ways to treat it. They noted that the field has moved toward less-aggressive treatments based on disease progression and the Gleason Score, which minimize patients' risk. This measure will fill a measurement gap for cancer care in MIPS and enhance the Advancing Cancer Care MIPS Value Pathway. The developer sought input from a clinician expert panel and from persons with lived experience. CMS noted that the measure has a robust risk-adjustment model.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle reported that this measure received four written public comments as well as comments from the listening session and from the round 1 evaluation. They described a robust comment from a specialty association which noted the potential for providers to perform poorly on the measure even though they are providing high-quality care. They indicated that though their association was engaged in the development of the measure, they did not feel their concerns were addressed. An Advisory Group member noted that the measure would help promote better adherence to clinical decision guidelines for prostate cancer. Another Advisory Group member shared their concerns around

undue burden related to how to attribute costs where the triggering event happens versus where the care happens as well as socioeconomic disparities.

A committee member noted that Black men are disproportionately more affected by prostate cancer; this can be exacerbated by factors such as socioeconomic status and education. The measure developer indicated that the measure risk adjusted for dual status and once it is adjusted, there were no significant inequity in outcomes, especially when comparing Black patients to white patients. Thus, no evidence was found that providers will be systematically penalized if they treat more vulnerable populations.

Another committee member inquired whether data was available to differentiate performance between high-risk localized or locally advanced patients and low-to-intermediate risk localized patients. CMS indicated that a major problem in oncology care is ICD-10 codes do not capture stage. The measure developer added that a lot of the clinical data elements are not available in claims, which is a limitation of a claims-based measure. However, registries are available, although they also have limitations, such as interoperability (e.g., the QCDR registry is only available in certain states). To address this issue, rather than relying solely on diagnoses, the measure developer reported that they include the care trajectory of the patient prior to the episode in the risk-adjustment model. The measure will be evaluated every 3 years, and the measure will continue to look for new data sources that could make the measure more robust. They indicated that the measure is performing well at this stage with reliability over 0.6, even with the 20-case minimum.

A committee member asked whether the validity findings themselves had undergone any clinical validation, to show that the model was effective. The measure developer reported that the measure has gone through a consensus-based development process that included 15 specialty societies as well as a round of national field testing. The measure then went through an iterative process where they considered what was included and not included in the measure. The measure developer indicated that their findings were consistent with the literature, which indicates that there are opportunities for improvement in avoiding adverse events. For example, for every $1,000 increase in adverse events, the measure score declines by 6%, showing that these are the same cost drivers as identified in the literature.

Another committee member asked if evidence-based treatment guidelines are available for the treatment of prostate cancer, and, if not, would there not be variation in treatment. CMS confirmed that guidelines exist for different kinds and stages of cancer, including prostate cancer. These guidelines are not definitive so significant clinician judgment and variability as well as patient choice and judgment are involved. The measure developer added that life expectancy is a key differentiative guideline, which they account for in the model by including age.

A committee member shared their concern about the signal this measure sends and the unintended consequences on costs, treatment options, the doctor-patient relationship, and decision-making given the need for more granularity in the measure in terms of staging. They reiterated concerns about racial disparities with prostate cancer, with Black men being more likely to be diagnosed and in a later stage. Thus, physicians who treat certain populations will disproportionately look worse, which might incentivize physicians to prioritize patients expected to have better outcomes. The committee member also responded to an assertion from CMS in the chat regarding a reported 2-year lag in getting data from registries, stating that in their experience, they get the data in real time. CMS acknowledged the challenge with reporting stage and molecular markers, which require the creation of a separate system. They indicated that risk adjustment is the best they can do to account for race and other factors. The committee

member reiterated a previously expressed concern that it is challenging to make decisions on a measure that has not been endorsed and for which the committee does not have access to the risk-adjustment model or validity testing data. The measure developer noted that the clinical workgroup did have the opportunity to review plan data and analyses and these were used in developing the measure specifications. They indicated that the clinical experts agreed that treatment procedure was a reasonable proxy for stage in the risk-adjustment model. The committee member expressed their appreciation for the helpful description but noted that they lacked the source of truth to be able to review the validity testing data themselves.

Conditions (if specified): None stated.

Future Directions:

The Recommendation Group did not reach consensus on this measure. Several committee members expressed their concerns about the fact that the risk-adjustment model did not directly account for stage. While the measure developer noted that they did use a proxy for stage, committee members said that they did not have adequate information available to assess the validity of the measure.

MUC2023-208 Respiratory Infection Hospitalization [CMS] †

Description: Evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who receive inpatient treatment for a respiratory infection. This acute inpatient medical condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during each episode from the clinical event that opens, or "triggers," the episode through 30 days after the trigger.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Consensus not reached

Vote Count: Recommend: 1 (7%); Recommend with Conditions: 5 (33%); Do Not Recommend: 9 (60%)

Measure Discussion:

CMS Opening Remarks: CMS indicated that this measure fills a gap left by a previous version of the measure, which was removed due to coding changes that unevenly affected clinicians. The measure evaluates groups of individuals on the risk-adjusted cost to Medicare for the inpatient treatment of respiratory infections, and it is attributable to clinicians and groups that would treat patients during the hospitalization. This measure addresses the concerns of the measure's previous version by expanding the patient cohort to include beneficiaries hospitalized for pneumonia, and related respiratory infections. The measure also reflects the coding changes that have occurred and addresses concerns regarding appropriate risk adjustment and exclusions.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle reported that there were two public comments in support of the measure and two in opposition of the measures. In their public comment, a specialty society indicated that while they opposed the use of the measure in the MIPS Program, they would be supportive of the measure concept in other programs attributable at broader system level. They also expressed concern around the attribution

methodologies and fact that hospitalists would have cases attributed to them. In the round 1 evaluation, Advisory Group members noted concerns around the use of the measure at the provider level and the scientific acceptability of the measure.

A committee member expressed concerns that the measure is gameable for a couple of reasons: 1) how it is handled in a clinical versus an urgent care and 2) the diagnosis used. They indicated that the measure sounds simple but may not be doable in real life. The measure developer indicated that the big focus of their reevaluation is accounting for differences in respiratory conditions. They previously only included one set of Medicare Severity Diagnosis Related Groups (MS-DRGs) or type of hospitalization in the simple pneumonia measure. They saw in their testing that the set of MS-DRGs was no longer detecting all the pneumonia cases because they were appearing in a different set of diagnosis-related groups (DRGs). One of the ways the measure developer account for the differences between those groups is by stratifying by the DRG and then further risk adjusting based on the MS-DRG. They also include a risk adjuster for COVID-19 and have exclusions for other respiratory conditions. The committee member noted that the model is only as good as the data so there may be some variation because what happens in the real world often skews the data.

Another committee member asked how the measure is currently being used in the program. The measure developer indicated that the measure was used at both the group and clinician level. They reported that CMS removed the measure from MIPS for the 2024 performance period. The committee member asked the type of providers this measure was pulling in and why the measure was removed for 2024. The measure developer reported that internal medicine, hospitalists, and nurse practitioners were the top three specialists when looking at clinicians who were in a group that met the case minimum. CMS added infectious disease and pulmonary to the list of the types of providers that were pulled in. At a high level, the measure was removed because of coding changes. The measure developer provided additional clarity on why the measure was removed, stating that they saw a shift such that the measure as specified was not picking up the types of pneumonia that it was supposed to pick up. The current version of the measure is an adjustment to reflect this change and ensure the continued assessment of pneumonia hospitalization episodes.

A committee member inquired whether there were any experiences related to the measure that could be shared, given that it had previously been in use in MIPS. The measure developer responded that this is something they are still evaluating as the previous version of the measure was implemented in the 2019 performance period and there have been reporting delays due to the COVID-19 pandemic and other changes to MIPS scoring. CMS echoed this sentiment, confirming a lag in receiving and then processing MIPS data, and this was exacerbated by the public health emergency. The committee member shared their feedback that it would be helpful to see results from the implementation of a measure that has been included in a program to inform the recommendation decision. The measure developer shared that they did reconvene a clinical expert workgroup 3 years after the previous measure was implemented and they continued to support the use of the measure in MIPS. They acknowledged the importance of monitoring analysis going forward.

Another committee member asked if they had examined the impact of regional variations or environmental factors such as wildfires on the number of hospitalizations. They also expressed their concern that 2021 data was used in developing this updated version of the measure given the impact of the public health emergency. The measure developer indicated that although they submitted initial results on 2021 data, they used 2022 data for beta testing. CMS said that in the case of weather-related issues (such as major hurricanes) that triggered an emergent situation, doctors in the impacted region might be exempted from the program depending on the

magnitude. The committee member requested clarification, noting that while doctors might not be scored, on whether submitted claims would be included in calculations and setting the benchmark. They indicated that this might skew the data and impact its integrity. CMS responded that this was a challenging question to answer as they have never encountered a situation like that. The measure developer addressed the question of regional variability, reporting that they conducted testing by geographic region. Their clinician expert workgroup knew of certain respiratory infections that disproportionately affected a certain area (for example, bird flu). This has been excluded from the measure; however, it is a very rare case.

Another committee member asked a clarifying question about whether, for measures that are at the physician level and, potentially, group level, there is a low-end size, such that if physician does not meet that low end, it is up to the group. The measure developer reported a case minimum of 20 for this episode-based cost measure as it was used in MIPS, meaning the measure does have a low end where a clinician is only scored if they meet the case minimum. For scoring in MIPS, clinicians can either be scored at the clinician- or group-level based on how they submit the data.

A committee member asked if the measure was adjusted for specialists versus internists or by intensive care unit (ICU) versus non-ICU. The measure developer said that they do not adjust for specialty; however, MS-DRGs are based on condition and complexity. Because they risk adjust by MS-DRG, that would account for differences in those types of cases.

Conditions (if specified): Measured at group level.

Future Directions:

The Recommendation Group did not reach consensus on this measure. Committee members noted that for measures that have previously been included in a program, it would be helpful for them to see data related to the implementation of the measure to inform their decision-making.

MUC2023-209 Rheumatoid Arthritis [CMS]

Description: Evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for patients who receive medical care to manage and treat rheumatoid arthritis. This chronic condition measure includes the cost of services that are clinically related to the attributed clinician's role in managing care during a Rheumatoid Arthritis episode.

Program: Merit-based Incentive Payment System (MIPS)

Committee Final Vote: Do not recommend

Vote Count: Recommend: 1 (6%); Recommend with Conditions: 3 (19%); Do Not Recommend: 12 (75%)

Measure Discussion:

CMS Opening Remarks: CMS stated that this measure will fill a measurement gap as no cost measures exist for this condition and rheumatologists do not have any episode-based cost measures for their specialty. They indicated that it would also enhance the Advancing Rheumatology Patient Care MIPS Value Pathway.

Battelle Summary of Public Comment and Round 1 Evaluation: Battelle indicated that one

specialty society shared their concerns that the measure is flawed and may have unintended consequences for beneficiary care. They noted that rheumatologists believe the measure does not generate actionable results and thus, they will not be able to take meaningful action. They also mentioned that the costliness of rheumatoid arthritis treatment is associated with the cost of medication, a comment which echoed by other commenters. Another specialty society noted that many of their practices have performed well on the cost category in prior years are now facing penalty for a number of reasons, including the costs of the medications. A specialty society also asked that extremely low treatment costs be defined. In the round 1 evaluation, an Advisory Group member expressed concern that they might be unattended consequences to patients by reducing the total costs rather than defining the appropriate processes needed for care. They noted that patients who need extensive medications could be at a disadvantage.

A committee member shared that they had concerns about a lot of cost measures; however, they described the need for the measure at the community level. They stated that in the community, when looking at cost data between the two major rheumatology groups, one is twice as expensive as the other. They indicated that it is simple utilization where they order a lot of labs and x-rays. Thus, the member indicated a big need for this measure despite their concerns.

The measure developer provided clarity on issues that came up in comments. They indicated that this measure is not yet in use in MIPS. They reported that any feedback about a rheumatologist being scored on cost measures might be on a TPCC measure, which is a different measure in use in MIPS. They indicated that the present measure is more clinically targeted, not just for rheumatologists, but specifically for rheumatoid arthritis. The measure developer shared that they thought carefully about drug costs during the development of this measure. The measure stratifies episodes based on patients with and without Part D enrollment to account for differences between the drugs that patients are on. They found that performance was similar for episodes with Part D enrollment, where the beneficiary has just Part D medication cost or just Part B medication costs. In terms of the extremely low-cost exclusion, they reported that it is like a data completeness exclusion that they do when information is insufficient to compare performance. Lastly, the measure developer expressed their agreement about the importance of having measures looking at things like quality and process; however, the current measure is a cost measure, which is why they focused on cost as the outcome.

A committee member shared their perspective that this measure is needed but they could see the frustration from the rheumatologists' comments, which begged the question if this is the right measure. They noted that this measure shares the same problems identified in the other cost measures that were discussed during the Recommendation Group meeting and this is reflected in the voting results. The measure developer noted that the measure reflects input specifically from rheumatologists and others involved with the care of rheumatoid arthritis. The committee member shared their observation that it seemed from the comments received that some of the large rheumatology associations were in opposition of the measure. CMS acknowledged that doctors often do not like to report on cost as they work with the aim to do the best for their patients. However, the program is legislated by Congress and cost is part of it; thus, they try to come up with the best measure they can to assess this. While the measures are not perfect, they are as good as they can be; they level the playing field for all rheumatologists, including those that are using more expensive medications; and then they try to risk adjust. The goal is to make it as palatable as it can be, given the requirements of the program in the legislation. The committee member shared their perspective that while this is true, physicians aim to improve the care of their patients so it is important to address the fundamental issues in assessing quality rather than simply focusing on what Congress has mandated. CMS agreed with this

sentiment but noted that there are cases where, for example, a doctor might order more tests than needed and that is associated with added costs that might have been avoided.

A committee member noted some confusion from the comments and requested clarification on whether another measure is now in use, in addition to the current one, or if this measure is intended to replace it. The measure developer acknowledged an additional global cost measure that is not being replaced by this one. This measure is intended to provide a specific cost measure for rheumatoid arthritis as required by MIPS.

CMS noted the legislative mandate is that 50% of measures in the program are cost measures with the aim of minimizing taxpayer costs and dealing with the reality of a few bad actors who make the system more challenging or even doctors with limited experience who might incur more costs than needed. A committee member shared their perspective that they have no constitutional opposition to payment measures that contributed to the decisions made by the Recommendation Group. The committee simply had concerns based on the information they had available to them, and this is just one step in the process.

Conditions (if specified): Undergo and receive CBE endorsement.

Future Directions:

The Recommendation Group voted to not recommend this measure for inclusion in the program. Battelle noted that this measure received more votes than any other, which likely led to consensus being reached (as compared to “consensus not reached” for the other cost measures reviewed during the meeting). This highlights how narrow the margins are for voting and the importance of each committee member contributing to the voting process.

Next Steps

Dr. Brennan offered an opportunity for meeting participants to stay on the call to provide feedback on the meeting, including recommendations for future meetings. She described what is statutorily required of the process; specifically, following the meeting, Battelle will summarize the votes and publish a spreadsheet listing the measures and votes that came out of the meeting as well as a report detailing everything that came out of the meeting. The recommendations will be posted by February 1, and this will launch a 15-day public comment period. No recommendations that came from the group will change based on the public comment; it serves as a last opportunity to reflect on the recommendations made and to give CMS feedback.

Dr. Brennan thanked the measure developers for being available throughout the meeting to answer questions and CMS for helping the committee in understanding the context of why the measures are being considered. She thanked committee members for their engagement and dialogue as the first ones to go through the new process. One of the committee co-chairs, Lisa Hines, expressed her appreciation to Battelle for making changes to the process and for their humility and willingness to accept feedback. Dr. Hines commended committee members for their engagement and contributions as well as the measure developers for their work. She noted that the voting did not reflect the rigor of the work they did but rather the lack of comfort from the committee with the information with which they were provided. Dr. Hines expressed her gratitude for the effort to include the patient perspective and thanked co-chair, Reginald Barnes. Mr. Barnes thanked Battelle and the Recommendation Group members for the discussion as well as the measures developer and CMS.

Dr. Brennan stated the committee will convene in an in-person meeting in the spring to collect additional feedback and to discuss feedback, changes, and potential strategy.

Closing Acknowledgements

Dr. Schrieber shared how impressed CMS was with the engagement of the committee across the 2-day meeting. She acknowledged that consensus was not reached on several measures but noted that the discussion on how to improve the measures has been invaluable. Dr. Schreiber stated that "no consensus" reflected the reality of the measurement world and that it is an honest evaluation. Dr. Schreiber noted that the spring in-person meeting will likely follow the CMS Quality Conference, which is in April 2024. The meeting will focus on debriefing on the Recommendation Group process, how to improve it, and strategies around measurement. Dr. Schreiber thanked Battelle for coordinating the meeting as well as CMS and the measure developers for the work they put into the measures. Dr. Schreiber closed by thanking the committee whose expertise and dedication has greatly informed CMS's work. She specifically thanked the co-chairs for their contribution.

The remainder of the meeting was opened up to committee members to share their feedback. Committee members shared areas of improvement and offered suggestions for future consideration.

- Improved communications, particularly in relation to roles and meeting details
- Post meeting materials to a central location (e.g., SharePoint, a dedicated webpage)
- Separate the meeting days as it can be challenging to hold focus for two consecutive days
- Adjust review times for measures as a lot of the work fell around the holidays, making it challenging to coordinate
- Give committee members access to round 1 evaluations to help inform the measure discussion
- Provide a subject matter overview of certain topics like cost to ensure that everyone is using and interpreting terminology in the same way
- Develop "measure at a glance" document that provides key information about a measure (e.g., whether or not it has been endorsed)
- Label the most up-to-date testing data so it is easily identified
- Consider committee member stated that it may have been easier to think about the actionability of cost measures if the committee was thinking about the related quality measures in the program at the same
- Continue to center conversations around health equity and unintended consequences
- Find ways to help people on the ground such as office managers and doctors understand the impact of measurement
- Given the lack of consensus and the small numbers of the Recommendation Group, consider the utility in having two separate groups and potentially include more voices
- Require CBE endorsement prior to submitting to the MUC List to ensure that measures are thoroughly evaluated for their scientific merits

Committee members also shared successes from the meeting.

- Strong facilitation by Battelle which allowed opposing ideas to come forth
- Good time management
- Presence of CMS and measure developers helpful in understanding the rationale for measures and clarifying discussion points
- Public comments were presented in a helpful format
- Presence of patient voice was greatly appreciated and invaluable.

Dr. Brennan and Dr. Schreiber closed the meeting by thanking everyone for their participation.