NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Applications Partnership (MAP) Clinician Workgroup

Preliminary Analyses

December 1, 2022

This report is funded by the Centers for Medicare & Medicaid Services under HHSM-500-T0003, Option Period 4

Contents

Merit-based Incentive Payment System-Cost

Preliminary Analysis – MUC2022-097 Low Back Pain

Measure Description:

The Low Back Pain episode-based cost measure evaluates risk adjusted cost to Medicare of a clinician or clinician group for patients receiving ongoing medical care to manage and treat low back pain. This chronic condition measure includes the costs of services that are clinically related to the role of the attributed clinician in managing care during a Low Back Pain episode.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). When compared to the two population-based cost measures currently used in MIPS, the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure (), this episode-based cost measure focuses on items and services related to the episode for lower back pain, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers. For example, the developer notes that extensive literature suggests that some clinicians routinely and unnecessarily conduct diagnostic tests for low back pain (e.g., magnetic resonance imaging [MRI]) in the absence of symptoms suggesting serious low back pain problems.

Low back pain is estimated to impact 84 percent of adults at some point in their lives (Deyo et al., 1987; Cassidy et al., 1998). Roughly 20 percent of Americans experience low back pain each year, and Medicare patients seeking low back treatment grew at nearly triple the rate of Medicare beneficiary growth (131 percent versus 42 percent) (Will et al. 2018; Blanpied et al. 2017). The frequency of low back pain translates to high overall healthcare spending. A 2020 study found that low back and neck pain contributed the most to healthcare spending among 154 mutually exclusive diagnoses, at $134.5 billion in 2016 (Dieleman, 2020).

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is an economic outcome measure of health care costs, or utilization. The developer provides two primary areas for provider intervention to drive improvement in performance variation of healthcare costs, specifically reducing low back pain complications through early physical therapy/conservative care, and promoting cost efficiency by avoiding expensive treatment options (e.g., imaging, injections) during the initial stages of care.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: The developer provides a performance range of the measure with a mean performance of $1,712 and median performance of $1,640. The minimum performance was $395 and maximum performance was $10,179 (SD $518). This range of performance on the measure suggests an opportunity for improvement in low back pain episode costs.

Does the measure contribute to the efficient use of measurement resources and/or support the alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The MIPS program cost domain does not currently include a low back pain episode cost measure. Given the prevalence and variation in episode costs associated with this clinical condition, this measure captures an important clinical area. The developer noted that there are several quality measures included in the program focused on a similar patient cohort, that are clinically related to the care provided for the measure, or that focus on complementary care that may not be directly captured by the cost measure (e.g., lumbar fusion). These quality measures are important to ensuring cost incentives are matched with quality incentives for this clinical area.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the developer, data elements used in this measure are in defined fields in electronic sources. This is a claims-based measure that uses codes for services billed in Medicare claims that are covered by Medicare Parts A, B, and D. It does not require any additional submission of data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This is a fully developed measure that is specified for the clinician and clinician group level of analysis. The measure has not been submitted for endorsement by a consensus-based entity (CBE) and has not been submitted or reviewed by MAP in the past. The developer demonstrates reliability performance of greater than 0.65 using a Weighted Heteroscedastic Within-Group Variance Estimation across all volume thresholds and across TIN and TIN-NPIs. This reliability performance suggests that measure has high reliability for clinician and clinician groups across a range of volume thresholds. To demonstrate construct validity, the developer notes that higher frequencies of high-costs events are associated with higher scores. The developer also assessed the correlation between the Lower Back pain episode-based cost measure with the Total Per Capita Cost (TPCC) all-cost measure. The results indicate that there is a statistically significant positive correlation between the two measures, around 0.5 at both the TIN and TIN-NPI level.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweigh the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure is not currently in use.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High-Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value of the program measure set?

This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode for a lower back pain, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Summary: What is the potential impact of this measure on the quality of care for patients?

Low back pain is estimated to impact 84 percent of adults at some point in their lives (Deyo et al., 1987; Cassidy et al., 1998). Roughly 20 percent of Americans experience low back pain each year, and Medicare patients seeking low back treatment grew at nearly triple the rate of Medicare beneficiary growth (131 percent versus 42 percent) (Will et al. 2018; Blanpied et al. 2017). The frequency of low back pain translates to high overall health care spending. A 2020 study found that low back and neck pain contributed the most to health care spending among 154 mutually exclusive diagnoses, at $134.5 billion in 2016 (Dieleman, 2020).

Preliminary Analysis – MUC2022-100 Emergency Medicine

Measure Description:

The Emergency Medicine episode-based cost measure evaluates a clinician's risk-adjusted cost to Medicare for patients who have an emergency department (ED) visit during the performance period. The measure score is the clinician's risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This measure includes costs of Part A and B services during each episode from the start of the ED visit that opens, or triggers the episode through 14 days after the trigger, excluding a defined list of services for each ED visit type that are unrelated to the ED care.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in MIPS, the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode for an emergency department (ED) visit, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Patients who are 65 and older have 29.2 million emergency department visits annually, representing a cost of $20.2 billion dollars to Medicare and other payors (Healthcare Cost and Utilization Project (HCUP) Statistical Brief #268). Hospital visits for those 65 years of age and older are disproportionately costly relative to other age groups.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is an economic outcome measure of health care costs, or utilization. The developer submitted literature identifying a variety of ways health care providers, particularly ED physicians, can affect the total cost of a visit. For example, the decision of whether to admit a patient from the ED varies considerably by physician, even with similar patient characteristics (Smulowitz et al., 2021) – however, choosing to admit a patient to the ED instead of discharging them to the community can increase costs by a factor of 14 (Burke et al., 2020). Likewise, to prevent costly readmissions or repeat ED visits during the 14-day period following the ED visit trigger, ED physicians can improve care transitions (Mansukhani et al., 2015; Nelson and Pulley, 2015).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: The developer provides a performance range of the measure with median performance of $5,172 and a standard deviation of $1,018. The 25th percentile of costs is $4,832, while the 75th percentile is $5,643, nearly 17 percent higher. This range of performance on the measure suggests an opportunity for improvement in ED visit costs.

Does the measure contribute to the efficient use of measurement resources and/or support the alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The MIPS program cost domain does not currently include an emergency ED cost measure. Given the prevalence and variation in episode costs, this measure captures an important clinical area. There are measures in the MIPS Quality Measures set that reflect clinical quality of care in emergency medicine (e.g., utilization measures of CT, preventive care and screening, and stroke care). These measures complement this episode of care for the ED visit measure.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: All data elements used in this measure are in defined fields in electronic sources. This is a claims-based measure that uses codes for services billed in Medicare claims that are covered by Medicare Parts A and B. It does not require any additional submission of data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This is a fully developed measure that is specified for the clinician and clinician group level of analysis. The measure has not been submitted for endorsement by the consensus-based entity (CBE) and has not been submitted or reviewed by the Measure Applications Partnership (MAP) in the past. The developer demonstrates reliability performance of greater than 0.836 using a Weighted Heteroscedastic Within-Group Variance Estimation across all volume thresholds and across TIN and TIN-NPIs. This reliability performance suggests that the measure has high reliability for clinician and clinician groups across a range of volume thresholds. To demonstrate construct validity, the developer notes that higher frequencies of high-costs events are associated with higher scores. The developer also assessed the correlation between this MUC with the Medicare Spending per Beneficiary (MSPB) Clinician measure, finding a statistically significant ($p<0.001$) correlation of 0.229.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweigh the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure is not currently in use, and no possible unintended negative consequences were raised by the developer.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value of the program measure set?

This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode for an emergency department (ED) visit, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Summary: What is the potential impact of this measure on the quality of care for patients?

Patients who are 65 and older have 29.2 million emergency department visits annually, representing a cost of $20.2 billion dollars to Medicare and other payors (Healthcare Cost and Utilization Project (HCUP) Statistical Brief #268). Hospital visits for those 65 years of age and older are disproportionately costly relative to other age groups. Implementing this measure will incentivize ED physicians to take steps to reduce the total cost of care for an ED visit. For example, the decision of whether to admit a patient from the ED varies considerably by physician, even with similar patient characteristics (Smulowitz et al., 2021) – however, choosing to admit a patient to the ED instead of discharging them to the community can increase costs by a factor of 14 (Burke et al., 2020). Likewise, to prevent costly readmissions or repeat ED visits during the 14-day period following the ED visit trigger, ED physicians can improve care transitions (Mansukhani et al., 2015; Nelson and Pulley, 2015).

Preliminary Analysis – MUC2022-101 Depression

Measure Description:

The Depression episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients receiving medical care to manage and treat depression. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Depression episode.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in MIPS, the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode for a depression episode, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

One study of the Medicare population suggests that prevalence of major depressive disorder (MDD) reached 4.96 percent in 2013, and had increased by 51 percent in the 5-year period between 2008-2013 (Bashyal et al., 2016). A review of the public health impact of burden in older adults found that depression is the leading cause of psychiatric hospitalization in older adults, and can exacerbate comorbidities, leading to higher health care utilization and costs (Zivi et al., 2013).

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is an economic outcome measure of health care costs, or utilization. The developer submitted literature identifying a variety of ways health care providers can affect the total cost of an episode of depression. For example, clinicians can improve adherence to antidepressant medication by implementing a shared decision-making approach, involving family caregivers in treatment, and closely monitoring short-term outcomes (Dell'Osso et al., 2020).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: The developer provides a performance range of the measure with a median performance of $1,380 and a standard deviation of $543. The 20th percentile of costs is $1,090, while the 80th percentile is $1,765. This range of performance on the measure suggests an opportunity for improvement in depression episode of care costs.

Does the measure contribute to the efficient use of measurement resources and/or support the alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The MIPS program cost domain does not currently include a depression episode

cost measure. Given the prevalence and variation in episode costs, this measure captures an important clinical area. There are measures in the MIPS Quality Measures set that reflect clinical quality of care in depression (e.g., suicide risk assessment, depression remission, anti-depressant medication management). These measures complement this episode of care for depression measure.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: All data elements used in this measure are in defined fields in electronic sources. This is a claims-based measure that uses codes for services billed in Medicare claims that are covered by Medicare Parts A, B, and D. It does not require any additional submission of data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This is a fully developed measure that is specified for the clinician and clinician group level of analysis. The measure has not been submitted for endorsement by the consensus-based entity (CBE) and has not been submitted or reviewed by the Measure Applications Partnership (MAP) in the past. The developer demonstrates reliability performance of greater than 0.835 using a Weighted Heteroscedastic Within-Group Variance Estimation across all volume thresholds and across TIN and TIN-NPIs. This reliability performance suggests that the measure has high reliability for clinician and clinician groups across a range of volume thresholds. To demonstrate construct validity, the developer notes that higher frequencies of high-cost events are associated with higher scores. The developer also assessed the correlation between this measure under consideration (MUC) with the Total Per Capita Cost (TPCC) all-cost measure, finding a statistically significant ($p<0.0001$) correlation of 0.412.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweigh the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure is not currently in use, and no possible unintended consequences were raised by the developer.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High-Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value of the program measure set?

This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode for a depression episode, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Summary: What is the potential impact of this measure on the quality of care for patients?

One study of the Medicare population suggests that prevalence of major depressive disorder (MDD) reached 4.96 percent in 2013, and had increased by 51 percent in the 5-year period between 2008-2013 (Bashyal et al., 2016). A review of the public health impact of burden in older adults found that depression is the leading cause of psychiatric hospitalization in older adults, and can exacerbate comorbidities, leading to higher health care utilization and costs (Zivi et al., 2013).

Implementing this measure under consideration (MUC) in the MIPS program could incentivize clinicians to more cost-effectively manage episodes of care of depression. The developer submitted literature identifying a variety of ways health care providers can affect the total cost of an episode of depression. For example, clinicians can improve adherence to antidepressant medication by implementing a shared decision-making approach, involving family caregivers in treatment, and closely monitoring short-term outcomes (Dell'Osso et al., 2020).

Preliminary Analysis – MUC2022-106 Heart Failure

Measure Description:

The Heart Failure episode-based cost measure evaluates a clinicians or clinician groups risk-adjusted cost to Medicare for patients receiving medical care to manage and treat heart failure. This chronic condition measure includes the costs of services that are clinically related to the role of the attributed clinician in managing care during a Heart Failure episode.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode of care for heart failure, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Heart failure costs of care represent at least $31 billion annually (Benjamin et al., 2019), affecting 6.2 million adults (Virani et al, 2020). Heart failure represents 41.5 percent of all inpatient Medicare admissions, and patients with heart failure account for over three times the cost of the typical Medicare patient (Fitch et al., 2016).

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is an economic outcome measure of health care costs, or utilization. The developer submitted literature identifying a variety of ways health care providers can affect the total cost of an episode of heart failure. For example, one study found that a nurse-directed intervention consisting of patient education, discharge planning, medication review, follow-up, and a social service consultation reduced readmissions by 56.2 percent (Rich et al., 1995). Likewise, a systematic review of self-management interventions found higher self-reported quality of life and hospitalization rates (Zhao et al., 2021).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: The developer provides a performance range of the measure with a median performance of $12,118 and a standard deviation of $3,510. The 20th percentile of costs is $9,241, while the 80th percentile is $14,750. This range of performance on the measure suggests an opportunity for improvement in heart failure episode of care costs.

Does the measure contribute to the efficient use of measurement resources and/or support the alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The MIPS program cost domain does not currently include a heart failure episode cost measure. Given the prevalence and variation in episode costs, this measure captures an important clinical area. There are measures in the MIPS Quality Measures set that reflect clinical quality of care in depression (e.g., Functional Status Assessments for Heart Failure Quality ID 377; Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) or Angiotensin Receptor-Neprilysin Inhibitor (ARNI) Therapy for Left Ventricular Systolic Dysfunction (LVSD) Quality ID 005; and, Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) Quality ID 008). These measures complement this episode of care for the heart failure visit measure.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: All data elements used in this measure are in defined fields in electronic sources. This is a claims-based measure that uses codes for services billed in Medicare claims that are covered by Medicare Parts A, B, and D. It does not require any additional submission of data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This is a fully developed measure that is specified for the clinician and clinician group level of analysis. The measure has not been submitted for endorsement by the consensus-based entity (CBE) and has not been submitted or reviewed by the Measure Applications Partnership (MAP) in the past. The developer demonstrates reliability performance of greater than 0.609 using a Weighted Heteroscedastic Within-Group Variance Estimation across all volume thresholds and across TIN and TIN-NPIs. This reliability performance suggests that the measure has high reliability for clinician and clinician groups across a range of volume thresholds. To demonstrate construct validity, the developer notes that higher frequencies of high-costs events are associated with higher scores.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweigh the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure is not currently in use, and no possible unintended consequences were raised by the developer.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High-Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value of the program measure set?

This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the episode of care for heart failure, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Summary: What is the potential impact of this measure on the quality of care for patients?

Heart failure costs of care represent at least $31 billion annually (Benjamin et al., 2019), affecting 6.2 million adults (Virani et al, 2020). Heart failure represents 41.5 percent of all inpatient Medicare admissions, and patients with heart failure account for over three times the cost of the typical Medicare patient (Fitch et al., 2016).

Implementing this measure under consideration (MUC) in the MIPS program could incentivize clinicians to more cost-effectively manage episodes of care of heart failure. The developer submitted literature identifying a variety of ways health care providers can affect the total cost of an episode of heart failure. For example, one study found that a nurse-directed intervention consisting of patient education, discharge planning, medication review, follow-up, and a social service consultation reduced readmissions by 56.2 percent (Rich et al., 1995). Likewise, a systematic review of self-management interventions found higher self-reported quality of life and hospitalization rates (Zhao et al., 2021).

Preliminary Analysis – MUC2022-129 Psychoses and Related Conditions

Measure Description:

The Psychoses/Related Conditions episode-based cost measure represents the cost to Medicare for the items and services provided to a patient during an episode of care (episode). This measure evaluates a clinician's risk-adjusted cost to Medicare for patients who receive inpatient treatment for psychoses or related conditions during the performance period. The measure score is the clinician's risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician during the episode and up to 45 days after the trigger.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the psychoses and related conditions, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Psychotic conditions are mental disorders associated with disturbances in thought processing and behaviors that result in a loss of contact with reality, and can occur throughout a patient's lifetime. This measure consists of Medicare beneficiaries enrolled in Medicare Parts A and B with an ICD-10 principal diagnosis for schizophrenia, delusional disorders, brief psychotic disorder, schizoaffective disorder, manic episode with psychotic symptoms, bipolar disorder with psychotic symptoms, major depressive disorder with psychotic symptoms, or unspecific psychosis on an inpatient claim that triggers a Psychoses and Related Conditions episode. The developer provides supplementary material demonstrating the percentage of the United States population, and Medicare specifically, with schizophrenia spectrum disorders, intellectual development disorder, dementia, major depressive disorder, and the direct and indirect costs associated with these conditions.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is an economic outcome measure of health care costs, or utilization. The developer provides a conceptual model linking treatment choices, and unobservable treatment choices (e.g., appropriate care plan, following clinical guidelines, proactive monitoring, medication reconciliation, care coordination, and patient education), and the direct effect of cost of services on the measure score.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: The developer provides a performance range of the measure for tax

identification number groups (TINs) with a mean performance of $17,092. The minimum performance was $6,328 and a maximum performance of $37,637 (standard deviation, $3,539). This range of performance on the measure suggests an opportunity for improvement in psychoses episode costs.

Does the measure contribute to the efficient use of measurement resources and/or support the alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The MIPS program cost domain does not currently include a psychoses episode cost measure. Given the prevalence and variation in episode costs, this measure captures an important clinical area. The developer noted that there are several quality measures included in the program focused on a similar patient cohort, that are clinically related to the care provided for the measure, or that focus on complementary care that may not be directly captured by the cost measure (e.g., mortality). These quality measures are important to ensuring cost incentives are matched with quality incentives for this clinical area.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the developer, all data elements used in this measure are in defined fields in electronic sources. This is a claims-based measure that uses codes for services billed in Medicare claims that are covered by Medicare Parts A, B, and D. It does not require any additional submission of data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This is a fully developed measure that is specified for the clinician and clinician group level of analysis. The measure has not been submitted for endorsement by a consensus-based entity (CBE) and has not been submitted or reviewed by the Measure Applications Partnership (MAP) in the past. The developer demonstrates reliability at the accountability entity level, the measure is highly reliable for both the TIN and TIN-NPI reporting levels, at 0.833 and 0.857 respectively.

To demonstrate construct validity, the developer provides results demonstrating that the cost measure is reflective of both the cost directly related to treatment choices, as well as cost of adverse outcomes as a result of care. These analyses help to provide evidence that the measure is capturing what it purports to measure. Further, correlation was assessed between this measure and the Medicare Spending Per Beneficiary Clinician measure. The results indicate that there is a statistically significant positive correlation between the two measures, greater than 0.35 at both the TIN and TIN-NPI level. Finally, face validity was assessed by a Psychoses and Related Conditions Clinician Expert Workgroup; a Technical Expert Panel (TEP); and the Person and Family Partners workgroup. Overall, there is very strong consensus among the members that all of the actions outlined in logic model for the measure are often or always within a reasonable influence of the attributed clinician, with every action receiving above 50 percent of responses that rated often or always.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweigh the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The Psychoses and Related Conditions measure is not currently in use, but is intended for use in a payment program and could eventually be publicly reported. The measure was specifically developed for potential use in the Cost performance category of MIPS to assess clinicians reporting as individuals or groups.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High-Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value of the program measure set?

This cost measure addresses a critical program objective and aligns with the Patient-Focused Episode of Care goal of CMS's Meaningful Measures initiative, and the domain of Efficiency and Cost Reduction. The development of episode groups for resource use analysis is also required by section 101(f) of MACRA. When compared to the two population-based cost measures currently used in the Merit-based Incentive Payment System (MIPS), the Medicare Spending Per Beneficiary Clinician measure and the Total Per Capita Cost measure, this episode-based cost measure focuses on items and services related to the psychoses and related conditions, which allows for focused improvement efforts for a narrowed clinical area and a focused group of clinical providers.

Summary: What is the potential impact of this measure on the quality of care for patients?

Psychotic conditions are mental disorders associated with disturbances in thought processing and behaviors that result in a loss of contact with reality, can occur throughout a patient's lifetime. This measure consists of Medicare beneficiaries enrolled in Medicare Parts A and B with an ICD-10 principal diagnosis for schizophrenia, delusional disorders, brief psychotic disorder, schizoaffective disorder, manic episode with psychotic symptoms, bipolar disorder with psychotic symptoms, major depressive disorder with psychotic symptoms, or unspecific psychosis on an inpatient claim that triggers a

Psychoses and Related Conditions episode. The developer provides supplementary material demonstrating the percentage of the U.S. population, and Medicare specifically with schizophrenia spectrum disorders, intellectual development disorder, dementia, major depressive disorder, and the direct and indirect costs associated with these conditions.

Merit-based Incentive Payment System-Quality

Preliminary Analysis – MUC2022-007 Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician and Clinician Group Level)

Measure Description:

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: The measure addresses outcome and digital measures as priorities for the Merit-based Incentive Payment System (MIPS) program, and the Safety Meaningful Measures 2.0 Healthcare Priority. The focus of this measure is to reduce radiation doses from computerized tomography (CT) scans, which increases the risk of cancer.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is an intermediate-outcome electronic clinical quality measure (eCQM) using electronic health data at the clinician and clinician group level that provides a standardized method for monitoring the performance of diagnostic computed tomography (CT) scan radiation doses, a risk factor for cancer, while preserving image quality. According to evidence provided in the developer's submission for consensus-based entity (CBE) endorsement in 2021, CT scans are used in most acute care facilities, and statistical inference suggests these scans cause approximately 2 percent of all new U.S. cancers diagnoses every year (Berrington de Gonzalez et al, 2009). The developer cites a retrospective cohort study finding a threefold increase in leukemia and brain cancer for pediatric patients who were CT scanned (Pierce et al., 2012).

The developer cites a randomized clinical trial of two interventions designed to reduce CT doses, finding "detailed feedback on CT radiation dose combined with actionable suggestions and quality improvement education significantly reduced doses, particularly organ doses" (Smith-Bindman et al., 2020).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: This measure addresses a patient safety concern of increased radiation dose from CT exams, as well as limited image quality of CT exams. The developer notes that doses used for CT vary substantially across imaging facilities for patients imaged for the same clinical indication. Specifically, the developer notes a study of 151 imaging facilities and hospitals where, after adjusting for patient characteristics, abdominal CT exams had a four-fold range in mean effective radiation dose and a 17-fold range in the proportion of high dose exams (Smith-Bindman et al., 2019). The developer does

not argue that there is a persistent quality challenge in image quality for CT scans; rather, this component of the measure is included as a "balancing" element, to prevent an unintended consequence where an excessive reduction in CT doses might compromise image quality and the diagnostic process.

In testing with 16 clinician groups, the developer found mean performance at 30 percent, with a standard deviation of seven percent, and minimum and maximum rates of 20 percent and 43 percent, respectively, indicating variation in performance and a performance gap. For this measure, a lower score indicates better quality.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: There are three process measures currently implemented in the MIPS program that address a similar concept, limiting radiation doses for CT scans:

- Optimizing Patient Exposure to Ionizing Radiation: Count of Potential High Dose Radiation Imaging Studies: Computed Tomography (CT) and Cardiac Nuclear Medicine Studies (Quality ID 360);
- Radiation Consideration for Adult CT: Utilization of Dose Lowering Techniques (Quality ID 436);
- Radiology: Exposure Dose Indices or Exposure Time and Number of Images Reported for Procedures Using Fluoroscopy (Quality ID 145)

However, as a clinical outcome measure instead of a process, this measure is not duplicative and presents an advantage over the measures currently in the set. The measure is being concurrently submitted for rulemaking in the Hospital Outpatient Quality Reporting (OQR) program, as well as the Hospital Inpatient Quality Reporting (IQR) program, encouraging alignment across these settings and across two otherwise separate levels of analysis. Reporting on this measure is applicable to a broad population: over a third of acute care hospitalizations involved at least one CT scan (Vance .et al., 2013).

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: This measure is an eCQM and according to the developer, all data elements are in defined fields in electronic sources. The developer conducted a feasibility assessment across eight different EHR systems with 16 clinician groups, finding all data elements were available in structured fields, and no impact on clinician workflow. During the fall 2021 CBE review of this measure by the National Quality Forum's Patient Safety Standing Committee, the Committee rated the measure "High" for feasibility.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The measure is endorsed by a consensus-based entity (CBE) (NQF #3633e and #3662e), is fully developed, and measure testing has demonstrated reliability and validity for the level of analysis. The measure was tested for reliability through a split-half correlation intraclass correlation coefficient (ICC), yielding a score of 0.99, indicating high reliability. NQF's Scientific Methods Panel and

Patient Safety Standing Committee both rated the measure "High" for reliability in the Fall 2021 evaluation. Likewise, both rated the measure "Moderate" for validity in the same evaluation, based on a face validity assessment of a Technical Expert Panel (TEP) composed of both clinicians and patient advocates.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The developer identified a possible unintended consequence, suggesting that image quality of CT scans might deteriorate if the radiation dose was lowered. The developer notes that by specifying the measure to also capture CT exams reported as having low image quality, the incentives are aligned to produce CT scans that are within an appropriate range that balances safety considerations with image quality.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Support for Rulemaking

Summary: What is the potential value to the program measure set?

The measure addresses outcome and digital measures as priorities for the Merit-based Incentive Payment System (MIPS) program, and the Safety Meaningful Measures 2.0 Healthcare Priority. The focus of this measure is to reduce radiation doses from computerized tomography (CT) scans, which increases the risk of cancer. There are three process measures currently implemented in the MIPS program that address a similar concept, limiting radiation doses for CT scans. However, as a clinical outcome measure instead of a process, this measure is not duplicative and presents an advantage over the measures currently in the set. The measure is being concurrently submitted for rulemaking in the Hospital Outpatient Quality Reporting (OQR) program, as well as the Hospital Inpatient Quality Reporting (IQR) program, encouraging alignment across these settings and across two otherwise separate levels of analysis. The measure is endorsed by a consensus-based entity (CBE) (National Quality Forum #3633e and #3662e), Reporting on this measure is applicable to a broad population: over a third

of acute care hospitalizations involved at least one CT scan (Vance et al., 2013).

Summary: What is the potential impact of this measure on quality of care for patients?

This is an intermediate-outcome electronic clinical quality measure (eCQM) using electronic health data at the clinician and clinician group level that provides a standardized method for monitoring the performance of diagnostic CT scan radiation doses, a risk factor for cancer, while preserving image quality. According to evidence provided in the developer's submission for consensus-based entity (CBE) endorsement in 2021, CT scans are used in most acute care facilities, and statistical inference suggests these scans cause approximately 2 percent of all new U.S. cancers diagnoses every year (Berrington de Gonzalez et al, 2009). The developer cites a retrospective cohort study finding a threefold increase in leukemia and brain cancer for pediatric patients who were CT scanned (Pierce et al., 2012).

The developer notes that doses used for CT vary substantially across imaging facilities for patients imaged for the same clinical indication. Specifically, the developer notes a study of 151 imaging facilities and hospitals where, after adjusting for patient characteristics, abdominal CT exams had a four-fold range in mean effective radiation dose and a 17-fold range in the proportion of high dose exams (Smith-Bindman et al., 2019). The developer cites a randomized clinical trial of two interventions designed to reduce CT doses, finding "detailed feedback on CT radiation dose combined with actionable suggestions and quality improvement education significantly reduced doses, particularly organ doses" (Smith-Bindman et al., 2020).

Preliminary Analysis – MUC2022-014 Ambulatory palliative care patients' experience of feeling heard and understood

Measure Description:

The percentage of top-box responses among patients aged 18 years and older who had an ambulatory palliative care visit and report feeling heard and understood by their palliative care provider and team within 2 months (60 days) of the ambulatory palliative care visit.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC) addresses patient experience and is a patient-reported outcome performance measure (PRO-PM), both of which are high priority areas for future measure consideration for the Merit-based Incentive Payment System (MIPS). It is also patient-reported, which is a priority for the Centers for Medicare & Medicaid Services (CMS) Meaningful Measures 2.0 initiative

Currently, no other measure focuses on this specific setting and clinical domain. Furthermore, the measure distinguishes differences in quality and is meaningful to patients/consumers and providers.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is a patient-reported outcome measure (PRO-PM) and has a scientific evidence-base and rationale for assessing the influence of healthcare processes or structures on the outcome. The developer provided a robust literature review documenting the importance of communication between patients and palliative care providers and impact on improved experience and quality of palliative care.

Palliative care has expanded rapidly in recent years, yet the measure developer cites studies showing that seriously ill persons often report feeling silenced, ignored, and misunderstood in medical institutions. Feeling heard and understood is essential in the process of patient-centered decision-making, which reinforces dignity, and is one of the key factors in patient-reported quality care. There is also growing consensus within the provider community regarding the need to measure the quality of palliative care.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: The quality of palliative care received in ambulatory clinics differs substantially from palliative care received in other settings, due in part to the interdisciplinary clinical team structure, patients' and families' limited access to palliative care services and difficulty managing and accepting their illness and trajectory. This variability in the patient experience of palliative care raises important measurement challenges. The developer cites evidence from the literature indicating the current level of data collected suggests there is room for improvement. In addition, the developer's national beta field test shows significant variability in performance rates across palliative care physicians with high, medium, and low performance. Across the 229 clinicians in the developer's sample, adjusted clinician

scores range from 42.0 to 90.9 with an average measure score of 71.0. The standard deviation in average clinician scores was 12.1. In addition, a clinician at the bottom 10th percentile of the ranking (e.g., the 10th lowest ranked clinician in 100 clinicians) would need a 22.0-point increase in measure score to improve to the median. This indicates that there is room for improvement and a gap in care.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The measure is conceptually aligned with, but not duplicative of the Hospice CAHPS Survey: Communication with Family (setting: hospice), which asks bereaved family caregivers of hospice patients how often they were kept informed, in laymen terms, and were listened to by healthcare providers. In contrast, this measure under consideration (MUC) asks patients directly the degree to which they felt heard or understood during their ambulatory care visit. Furthermore, the developer attests that the measure is harmonized with similar measures, and an environmental scan revealed no competing measures.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: Some data elements for this measure are in defined fields in electronic sources. According to the developer, visit information for patient eligibility, patient contact information for survey fielding, as well as patient age and gender for measure analyses will be pulled from the electronic health record. All other data elements for the measure are collected via the survey instrument. The patient experience survey developed for this measure is meant to be completed via web survey, on paper or over telephone in English. Patients who have already completed the survey are not asked to complete it again in order to avoid concerns with recall and other response biases and to minimize patient burden.

The developer conducted a national beta field test which indicated the feasibility of identifying eligible patients using administrative data and using a survey vendor to support survey administration and data collection.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This MUC is fully developed and endorsed by a consensus-based entity (CBE) (National Quality Forum #3665). The endorsed measure assesses patients' experience within 3-months (90 days) of a visit, while this measure submitted for consideration in MIPS assesses patient experience within 2 months (60 days) of an ambulatory palliative care visit. The measure submitted for MIPS consideration is intended to be used by providers eligible for MIPS, who provide palliative care services to their patients, so that the patient experience of core components of high-quality palliative care can be attributed and used to incentivize quality improvement.

The developer conducted several different tests of reliability and validity. Signal-to-noise reliability testing resulted in an estimate of the adjusted intraclass coefficient (ICC) of approximately 0.150 (95% confidence interval: 0.105 to 0.204), and a median adjusted ICC is 0.148. The developer also computed estimates of individual clinician specific reliability using a method similar to the approach utilized in

Adams (2009), estimating a posterior distribution for the overall variability of the risk-adjusted clinician scores and estimating a posterior distribution of the variance of each within-clinician score as specified in Adams (2009). The average reliability across clinicians was approximately r=0.647, and the median was r=0.698.

To evaluate the validity of the MUC, the developer examined the association of the measure score with the Receiving Desired Help for Pain measure score, the CAHPS communication measure score, and individual's overall rating of their palliative care provider and team. Interpretation of correlations followed standard conventions for small, medium, and large associations (i.e., 0.10, 0.30, 0.50) (Rosnow and Rosenthal, 1989). As hypothesized, the MUC was significantly and positively associated with the CAHPS communication performance measure (r = 0.635, p=0.011), the Receiving Desired Help for Pain performance measure (r = 0.496, p<.001) and the overall rating of the palliative care provider and team (r= 0.768, p=<.001).

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: This MUC is not yet in use. During development, the developer explored potential negative unintended consequences. The developer queried providers during alpha and beta testing about possible unintended consequences and they noted that there may be challenges with comparison across palliative care settings, especially if patient populations have differences in disease trajectories that may impact communication. Another challenge identified during testing was the perception that patients may have of palliative caregivers as bearers of bad news, which could result in skewed negative survey responses. Providers recommended strategies to mitigate these perceptions including clearly differentiating survey questions that assessed the types of news or information shared with patients from those that assessed the treatment patients received in the ambulatory care setting. Some expressed concern that surveys may be sent to deceased patients inadvertently, adding to families' distress. The recommended approach to address this concern involved mailing out notification of the upcoming survey in advance with a stamped postcard that can be returned in the event of a patient's death or an address change.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Support for Rulemaking

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) addresses patient experience and is a patient-reported outcome performance measure (PRO-PM), both of which are high priority areas for future measure consideration for the Merit-based Incentive Payment System. It is also patient-reported, which is a priority for the Centers for Medicare & Medicaid Services (CMS) Meaningful Measures 2.0 initiative. Currently, no other measure focuses on this specific setting and clinical domain. Furthermore, the measure distinguishes differences in quality and is meaningful to patients/consumers and providers.

The measure is conceptually aligned with, but not duplicative of Hospice CAHPS Survey: Communication with Family (setting: hospice), which asks bereaved family caregivers of hospice patients how often they were kept informed, in laymen terms, and were listened to by healthcare providers. In contrast, this MUC asks patients directly the degree to which they felt heard or understood during their ambulatory care visit.

Summary: What is the potential impact of this measure on quality of care for patients?

The quality of palliative care received in ambulatory clinics differs substantially from palliative care received in other settings, due in part to the interdisciplinary clinical team structure, patients and families limited access to palliative care services and difficulty managing and accepting their illness and trajectory. This variability in the patient experience of palliative care raises important measurement challenges. The developer cites evidence from the literature indicating the current level of data collected suggests there is room for improvement.

Palliative care has expanded rapidly in recent years, yet the measure developer cites studies showing that seriously ill persons often report feeling silenced, ignored, and misunderstood in medical institutions. Feeling heard and understood is essential in the process of patient-centered decision-making, which reinforces dignity, and is one of the key factors in patient-reported quality care. There is also growing consensus within the provider community regarding the need to measure the quality of palliative care.

Across the 229 clinicians in the developer's sample, adjusted clinician scores range from 42.0 to 90.9 with an average measure score of 71.0. The standard deviation in average clinician scores was 12.1.

Preliminary Analysis – MUC2022-048 CVD Risk Assessment Measure - Proportion of pregnant/postpartum patients that receive CVD Risk Assessment with a standardized instrument.

Measure Description:

This measure determines the percentage of pregnant or postpartum patients at a clinic who received a CVD risk assessment with a standardized instrument, such as the CVD risk assessment algorithm developed by the California Maternal Quality Care Collaborative (CMQCC). Aim is that 100 percent of eligible pregnant/postpartum patients undergo CVD risk assessment using a standardized tool. Every patient should be assessed for CVD risk at least once during the and, as needed, additional times when symptoms present during the pregnancy postpartum period. The measure can be calculated on a quarterly or annual basis.

Does the measure address a critical quality objective not currently ad equately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC), which assesses the proportion of pregnant or post-partum who are evaluated for cardiovascular disease, is a screening measure for a high-risk condition. This measure addresses the Meaningful Measures 2.0 domain of Wellness and Prevention, and both Maternal Health and Chronic Conditions, high-priority areas for future measure consideration identified by the Merit-Based Incentive Payment System program.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The American College of Obstetricians and Gynecologists has published a Practice Bulletin on pregnancy and cardiovascular disease, providing guidance for early antepartum and postpartum risk factor identification, among other recommendations. The screening algorithm used to identify cardiovascular disease in pregnant women was developed and tested by a research team with the California Maternal Quality Care Collaborative (CMQCC) (Hameed et al., 2017).

The developer presented evidence that cardiovascular conditions and cardiomyopathy together account for 26.5 percent of all maternal deaths, with hypertension a further 7.4 percent (Creanga et al., 2017). A retrospective evaluation of cardiovascular pregnancy-related deaths attributed 61 percent of these to delays in evaluation and treatment by healthcare providers (Hameed et al., 2015); a screening measure would directly contribute to improved clinical outcomes.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: A retrospective evaluation of cardiovascular pregnancy-related deaths attributed 61 percent of these to delays in evaluation and treatment by healthcare providers (Hameed et al., 2015). Likewise, a study of serious cardiac events in pregnant women with heart disease found that 49 percent of these events were preventable, mostly due to lapses in provider management inconsistent with standards of care, including the failure to identify the underlying condition prior to pregnancy (Pfaller et al., 2020). Researchers propose universal screening for cardiovascular disease

using the CMQCC algorithm, to address rising maternal morbidity and mortality, as well as substantial socioeconomic and racial disparities In health outcomes (Chambers et al., 2022). In testing, the developer found a mean screening rate of 19.2 percent, with a standard deviation of 22.7 percent, indicating a wide range of performance and opportunity for improvement.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: Though there are other measures of maternity care in the MIPS program (e.g., Postpartum follow-up and care coordination (CMIT 01958-C-MIPS)), and other measures that screen for conditions such as cervical cancer (CMIT 05778-E-MIPS), no other measures address screening for this particular condition. In testing at academic medical centers in California and New York, researchers found 8 percent of patients screened positively, and of these, 30 percent were found on follow-up to have a confirmed diagnosis of cardiovascular disease (Blumenthal et al., 2020). Extrapolated over the 3.6 million annual births in the U.S. suggests applicability to a broad population of hundreds of thousands of women.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: The developer reports that all data elements are in defined fields in electronic sources. The developer provided examples of integrations of the screening tool into EPIC and Cerner electronic health record (EHR) systems, and notes a paper version is also available as an alternative. The measure has been successfully tested with a sample of 169 clinicians.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The measure is fully developed, and specified for clinicians and clinician groups. The measure was not tested for reliability at the clinician level, only at the facility level. However, testing at the facility level demonstrated strong reliability, with a median reliability score of 0.992 using a signal-to-noise approach. At the patient or encounter level, the developer conducted a Kappa analysis between extracted EHR data and a manual review of the medical record across 2,535 charts, finding perfect agreement between the two.

A technical expert panel convened by the developer unanimously agreed that the performance measure scores can be used to distinguish good from poor quality care.

The measure has been submitted for endorsement to a consensus-based entity (CBE) in August 2022 but has not yet been evaluated.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The developer identifies overuse of resources as a result of implementing a new screening measure as a possible unintended consequence, but did not observe this in testing or

implementation at sites to date.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support for this measure is conditional on endorsement by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC), which assesses the proportion of pregnant or post-partum who are evaluated for cardiovascular disease, is a screening measure for a high-risk condition. This measure addresses the Meaningful Measures 2.0 domain of Wellness and Prevention, and both Maternal Health and Chronic Conditions, high-priority areas for future measure consideration identified by the Merit-Based Incentive Payment System program.

Though there are other measures of maternity care in the MIPS program (e.g., Postpartum follow-up and care coordination (CMIT 01958-C-MIPS)), and other measures that screen for conditions such as cervical cancer (CMIT 05778-E-MIPS), no other measures address screening for this particular condition. In testing at academic medical centers in California and New York, researchers found 8 percent of patients screened positively, and of these, 30 percent were found on follow-up to have a confirmed diagnosis of cardiovascular disease (Blumenthal et al., 2020). Extrapolated over the 3.6 million annual births in the U.S. suggests applicability to a broad population of hundreds of thousands of women.

Summary: What is the potential impact of this measure on quality of care for patients?

The developer presented evidence that cardiovascular conditions and cardiomyopathy together account for 26.5 percent of all maternal deaths, with hypertension a further 7.4 percent (Creanga et al., 2017). A retrospective evaluation of cardiovascular pregnancy-related deaths attributed 61 percent of these to delays in evaluation and treatment by healthcare providers (Hameed et al., 2015); a screening measure would directly contribute to improved clinical outcomes.

A study of serious cardiac events in pregnant women with heart disease found that 49 percent of these

were preventable, mostly due to lapses in provider management inconsistent with standards of care, including the failure to identify the underlying condition prior to pregnancy (Pfaller et al., 2020). Researchers propose universal screening for cardiovascular disease using the California Maternal Quality Care Collaborative (CMQCC) algorithm, to address rising maternal morbidity and mortality, as well as substantial socioeconomic and racial disparities In health outcomes (Chambers et al., 2022). In testing, the developer found a mean screening rate of 19.2 percent, with a standard deviation of 22.7 percent, indicating a wide range of performance.

Preliminary Analysis – MUC2022-052 Adult COVID-19 Vaccination Status

Measure Description:

Percentage of patients aged 18 years and older seen for a visit during the performance period who have ever completed or reported having ever completed a COVID-19 vaccination series and one booster dose

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC), assessing clinicians based on the COVID-19 vaccination rate (complete vaccination and at least one booster dose) of their patients, is a specific high priority area for future measurement identified by CMS for the Merit-based Incentive Payment System Program (MIPS). Although other measures included in the program address immunizations, none are specific to COVID-19.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The CDC's Advisory Committee on Immunization Practices (ACIP) have issued a guideline concluding that the COVID-19 vaccine is safe and highly effective at preventing symptomatic COVID-19 in adults, and recommending all adults follow a full course of vaccination. A recent study across nearly 12,000 U.S. adults found a full vaccination course was no less than 86 percent effective in reducing hospitalizations from the later delta and omicron variants (Lauring et al., 2022). Moreover, these hospitalizations are costly; one analysis by the Kaiser Family Foundation found that vaccine-preventable hospitalizations of adults cost an estimated $14 billion over six months (June-November 2021).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: According to the CDC, 111,848,839 people in the U.S. received at least one booster dose, or 49.3% of all those who have completed the initial primary series of vaccination. In testing, the developer found that the 25th percentile of performance was 28.3 percent of patients vaccinated with one booster, and the 75th percentile was 58% of patients vaccinated with one booster. Taken together, these data indicated that the performance gap assessed by this measure represents nearly half of U.S. adults.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: As the COVID-19 virus could potentially (re)infect nearly every U.S. adult, with potentially devastating consequences for an older population, the potential impact of a widespread increase in vaccination rates across all clinicians that report to MIPS is difficult to overstate. However, three of nine patients surveyed did not agree that this measure result would be useful to patients

making decisions about their healthcare, suggesting some limits to the utility of publicly reporting the measure result.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: The developer assessed the feasibility of implementing the measure at four clinical sites, finding that each systematically collected the needed data elements in an EHR in a structured field as part of their ordinary workflow, and would be able to report the measure if implemented. However, one registry that participated in the assessment indicated some small and rural practices do not collect these data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: No

Justification and Notes: The measure has been specified for the appropriate setting, level of analysis, and population, and is currently being trialed in the field as a beta test. In conducting empirical reliability testing, the developer found a median signal-to-noise ratio of 0.986. Face validity testing of the measure was conducted, and agreement among the six expert respondents was poor. Just three experts agreed the measure, as specified, would reflect the quality of care provided by clinicians. Five of the six recommended removing the booster dose from the measure numerator, and three recommended adding an exclusion for patient refusal. Moreover, the measure has not been evaluated by the consensus-based entity (CBE) for endorsement.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: Yes

Justification and Notes: The developer raises several potential unintended consequences, including patient selection based on vaccination status, or vaccination of patients despite possible contraindications if these are not specified. The paramount unintended consequence raised by the developer is the potential to penalize clinicians who disproportionately care for patient groups who may be less likely to be vaccinated. Indeed, the U.S. states show pronounced variation in vaccination and booster rates, ranging from 63.9 percent in Vermont to 29.6 percent in North Carolina. Other studies have shown substantial differences in vaccination rates based on patient characteristics, such as age (<29 years at 38.3 percent compared to 80 percent for 65, Diesel et al., 2021), and high-risk condition (63.8 percent for those with such conditions and 41.5 percent for those without, Pingali et al., 2021). Finally, the Kaiser Family Foundation has noted racial disparities in vaccination, particularly much higher rates among Asian people. Taken together, these data suggest that clinicians' performance may be disproportionately affected by demographic and cultural circumstances outside their control.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Do Not Support for Rulemaking with Potential for Mitigation

The potential mitigation is a re-specification of the measure to address concerns raised during the expert panel interviews, and endorsement by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) directly addresses a leading priority, measures to address the COVID-19 pandemic, for the Centers for Medicare and Medicaid Services (CMS) for the Merit-based Incentive Payment System Program (MIPS). Although other measures included in the program address immunizations, none are specific to COVID-19. However, three of nine patients surveyed did not agree that this measure result would be useful to patients making decisions about their healthcare, suggesting some limits to the utility of publicly reporting the measure result.

Summary: What is the potential impact of this measure on quality of care for patients?

Clinical guidelines and systematic reviews are in universal agreement that the COVID-19 vaccines are safe and effective at preventing costly and harmful hospitalizations. As the COVID-19 virus could potentially (re)infect nearly every U.S. adult, with potentially devastating consequences for an older population, the potential impact of a widespread increase in vaccination rates across all clinicians that report to MIPS is difficult to overstate. The CDC reports that less than half of people who completed an initial series of vaccination have received a booster. In testing, the developer found that even the 75th percentile of performance had only 58 percent of a clinician's patients vaccinated with one booster.

However, in a face validity test, half of the clinicians surveyed expressed concerns with the ability of the measure result to distinguish quality care, and suggested extensive updates to the inclusion and exclusion criteria for the numerator. The potential unintended consequences suggest a re-examination of the specifications and further expert panel review are appropriate at this stage of development and testing.

Preliminary Analysis – MUC2022-060 First Year Standardized Waitlist Ratio (FYSWR)

Measure Description:

The FYSWR measure tracks the number of incident patients in a practitioner (inclusive of physicians and advanced practice providers) group who are under the age of 75 and were listed on the kidney or kidney-pancreas transplant waitlist or received a living donor transplant within the first year of initiating dialysis. For this measure, patients are assigned to the practitioner group based on the National Provider Identifier (NPI)/Unique Physician Identifier Number (UPIN) information entered on the CMS Medical Evidence 2728 form.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC) targets Chronic Conditions, which is a stated high-priority area for future measure consideration for the Merit-based Incentive Payment System (MIPS). One electronic survey of 409 patients, the majority of whom were kidney transplant recipients, with the remainder chronic kidney disease patients or patients on dialysis, found that waitlisting was the most prioritized item from those surveyed, and was twice as likely to be ranked as the most important factor in choosing a dialysis center (Husain et al., 2018).

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: No

Justification and Notes: The developer convened two technical expert panels that both favored developing measures to evaluate quality of care in the area of waitlisting. Adding patients to the waitlist is a precursor to obtaining a transplanted kidney, which is very likely to improve clinical outcomes: a systematic review of the benefits of transplantation found higher self-reported quality of life and lower mortality in transplant recipients compared with those who remained on dialysis only (Tonelli et al., 2011). However, in the spring 2022 evaluation of this measure by the consensus-based entity (CBE) (National Quality Forum (NQF) #3689), the Renal Standing Committee did not reach consensus on the evidence base of the measure, and the measure was not recommended for endorsement. The Standing Committee questioned the attribution of the measure result to the care given by a nephrologist, as opposed to the practices of the transplant center.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: In the spring 2022 CBE evaluation of this measure, NQF's Renal Standing Committee was in near-universal agreement there was at least a moderate performance gap. In testing, the developer found the bottom quartile of 2,168 group practices had a 46 percent lower rate of waitlisting or living-donor transplantation than the mean rate of 1.01, whereas the top quartile had a rate that was 33 percent higher than the mean.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: No

Justification and Notes: Dialysis and renal transplants are the result of chronic kidney disease, which according to the CDC affects 37 million people in the United States. In testing, the developer found that over 281,000 patients were included in the measure calculation. There are no other measures currently in the MIPS program that capture the same concepts as this MUC. However, there is another MUC proposed in the 2022-2023 cycle, MUC2022-063 Percentage of Prevalent Patients Waitlisted (PPPW) and Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW), which tracks the percentage of patients in each dialysis practitioner group practice who are on the kidney or kidney-pancreas transplant waitlist (all patients or patients in active status), inviting a possible duplication of effort or even confusion by providers if both measures were incorporated into the MIPS program. The developer notes that this MUC distinguishes itself from MUC2022-063 by focusing on patients in their first year of dialysis, instead of measuring all patients in a dialysis practitioner group practice on the waitlist.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the developer, all data elements are currently collected in defined fields in electronic health records and other electronic sources of data, and are generated or collected by and used by healthcare personnel during provision of care.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: No

Justification and Notes: The MUC was not recommended for endorsement by NQF's Renal Standing Committee in a 2022 evaluation. The majority of the Committee voted either Low or Insufficient in the evaluation of validity. The Standing Committee expressed concerns with the exclusions (e.g., patients who are added to the wait list prior to starting dialysis), and the data source for classifying whether providers could report this measure. In addition, as part of the evidence review, the Standing Committee did not reach consensus on the attribution of the measure result to the care given by a nephrologist, as opposed to the practices of the transplant center.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The developer did not raise any potential unintended consequences. In the 2022 evaluation of the measure by the Renal Standing Committee, they raised a potential unintended consequence where patients might be coerced into obtaining a transplant when they may not want one.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Leave blank for now.]

MAP Health Equity Advisory Group Input:

Votes: [Leave blank for now.]

Recommendation

Preliminary Analysis Recommendation:

Do Not Support for Rulemaking with Potential for Mitigation

The potential mitigation for this measure would be to address the concerns raised by the Renal Standing Committee regarding the evidence base and specifications, and resubmit the measure for endorsement by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) targets Chronic Conditions, which is a stated high-priority area for future measure consideration for the Merit-based Incentive Payment System (MIPS). One electronic survey of 409 patients, the majority of whom were kidney transplant recipients, with the remainder chronic kidney disease patients or patients on dialysis, found that waitlisting was the most prioritized item from those surveyed, and was twice as likely to be ranked as the most important factor in choosing a dialysis center (Husain et al., 2018). There are no other measures currently in the MIPS program that capture the same concepts as this MUC.

However, there is another MUC proposed in the 2022-2023 cycle, MUC2022-063 Percentage of Prevalent Patients Waitlisted (PPPW) and Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW), which tracks the percentage of patients in each dialysis practitioner group practice who were on the kidney or kidney-pancreas transplant waitlist (all patients or patients in active status). In effect, this MUC tracks only new patients – MUC2022-063 tracks how well a group/practice is able to maintain patients on the waitlist.

Unlike this MUC, MUC2022-063 was recommended for endorsement by the National Quality Forum's Renal Standing Committee in the spring 2022 evaluation. Therefore, to prevent duplication while recognizing this high-value area, MUC2022-063 should be prioritized for inclusion in federal rulemaking above this MUC.

Summary: What is the potential impact of this measure on quality of care for patients?

Dialysis and renal transplants are the result of chronic kidney disease, which according to the CDC affects 37 million people in the United States. In testing, the developer found that over 281,000 patients were included in the measure calculation. Adding patients to the waitlist is a precursor to obtaining a transplanted kidney, which is very likely to improve clinical outcomes. Nevertheless, there is a substantial performance gap in the area of getting eligible patients onto the kidney transplant wait list. In testing, the developer found the lowest 25 percent of practices had a 46 percent lower rate of waitlisting or living-donor transplantation than the average.

However, in the spring 2022 evaluation of this measure by the consensus-based entity (National Quality Forum (NQF) #3689), the Renal Standing Committee did not recommend the measure for endorsement. The Standing Committee questioned the attribution of the measure result to the care given by a nephrologist, as opposed to the practices of the transplant center. The Standing Committee also did not pass the measure on the validity criterion, finding concerns with the exclusions (e.g., patients who are added to the wait list prior to starting dialysis), and the data source for classifying whether providers could report this measure.

Preliminary Analysis – MUC2022-063 Percentage of Prevalent Patients Waitlisted (PPPW) and Percentage of Prevalent Patients Waitlisted in Active Status (aPPPW)

Measure Description:

This measure tracks the percentage of patients in each dialysis practitioner group practice who were on the kidney or kidney-pancreas transplant waitlist (all patients or patients in active status). Results are averaged across patients prevalent on the last day of each month during the reporting year. The proposed measure is a directly standardized percentage, which is adjusted for covariates (e.g. age and risk factors).

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC) targets Chronic Conditions, which is a stated high-priority area for future measure consideration for the Merit-based Incentive Payment System (MIPS). One electronic survey of 409 patients, the majority of whom were kidney transplant recipients, with the remainder chronic kidney disease patients or patients on dialysis found that waitlisting was the most prioritized item from those patients, and was twice as likely to be ranked as the most important factor in choosing a dialysis center (Husain et al., 2018).

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The developer convened two technical expert panels that both favored developing measures to evaluate quality of care in the area of waitlisting. Adding patients to the waitlist is a precursor to obtaining a transplanted kidney, which is very likely to improve clinical outcomes: a systematic review of the benefits of transplantation found higher self-reported quality of life and lower mortality in transplant recipients compared with those who remained on dialysis only (Tonelli et al., 2011). In the Spring 2022 Consensus Development Process evaluation of this measure by the consensus-based entity (CBE), the Renal Standing Committee passed both the PPPW (National Quality Forum (NQF) #3695) and aPPPW (NQF #3694) measures on evidence.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: In the Spring 2022 Consensus Development Process evaluation of both the PPPW and aPPPW measures, the Renal Standing Committee was in near-universal agreement there was at least a moderate performance gap. In testing, the developer found a mean performance score of 19.1 percent with a standard deviation of 8.1 percent for the PPPW measure, and 11.9 percent with a standard deviation of 6.9 percent for the aPPPW measure.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: Dialysis and renal transplants are the result of chronic kidney disease which according to the Centers for Disease Control and Prevention (CDC) affects 37 million people in the United States. In testing, the developer found that over 281,000 patients were included in the measure calculation. There are no other measures currently in the MIPS program that capture the same concepts as this MUC. However, note that this MUC includes two different calculations, and implementation would lead to two different measure results. However, only the PPPW (NQF #3695) rate was recommended for endorsement by the consensus-based entity.

In addition, there is another measure under consideration for the 2022-2023 MUC cycle, MUC2022-060 entitled First Year Standardized Waitlist Ratio (FYSWR), which tracks the number of incident patients in a practitioner (inclusive of physicians and advanced practice providers) group who are under the age of 75 and were listed on the kidney or kidney-pancreas transplant waitlist or received a living donor transplant within the first year of initiating dialysis, inviting a possible duplication of effort or even confusion by providers if both rates were incorporated into the MIPS program. This MUC is a preferable alternative to MUC2022-060, as one of the measure constructs incorporated in this MUC was recommended for endorsement by the Renal Standing Committee (MUC2022-060 did not receive endorsement from NQF). The developer notes that this MUC distinguishes itself from MUC2022-060 by focusing on all patients in a dialysis practitioner group practice on the waitlist, instead of patients in their first year of dialysis.

The measure was previously submitted to the End-Stage Renal Disease Quality Improvement Program (ESRD QIP) in 2017, and conditionally supported by the Measure Applications Partnership (MAP), pending NQF endorsement. However, the measure reviewed by MAP took a slightly different form than the measure under consideration, being specified for and submitted for a program that covered dialysis facilities, as opposed to clinicians or practice groups.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the developer, all data elements are currently collected in defined fields in electronic health records and other electronic sources of data, and are generated or collected by and used by healthcare personnel during provision of care. In the Spring 2022 Renal Standing Committee evaluation, the PPPW measure passed the feasibility criteria. The aPPPW was not assessed for feasibility, as evaluation did not continue after the measure did not pass the evaluation of validity.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: No

Justification and Notes: One rate included in this MUC, the aPPPW (NQF #3694) was not recommended for endorsement by the Renal Standing Committee in a 2022 evaluation. The majority of the Committee voted either Low or Insufficient in the evaluation of validity. The Standing Committee expressed concerns with the use of social determinants of health in the risk adjustment model. However, the other rate included in this MUC, the PPPW (NQF #3695), was recommended for endorsement by the Standing Committee.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The developer did not raise any potential unintended consequences. In the 2022 evaluation of the measure by the Renal Standing Committee, they raised a potential unintended consequence where patients might be directed by their group/practice to transplant centers that are more likely to waitlist them, constraining the patient's choice of transplant center. The developer has indicated that the calculation of the measure result accounts for patients that go to a transplant center outside of their ZIP code.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support for this measure is conditional on updating the measure specifications to include only the PPPW (National Quality Forum (NQF) #3695) rate that was recommended for endorsement by NQF's Renal Standing Committee.

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) targets Chronic Conditions, which is a stated high-priority area for future measure consideration for the Merit-based Incentive Payment System (MIPS). One electronic survey of 409 patients, the majority of whom were kidney transplant recipients, with the remainder chronic kidney disease patients or patients on dialysis found that waitlisting was the most prioritized item from those surveyed, and was twice as likely to be ranked as the most important factor in choosing a dialysis center (Husain et al., 2018). However, note that this MUC includes two different calculations, and implementation would lead to two different results. For this reason, this measure should be updated to select only one rate.

In addition, there is another MUC proposed for the 2022-2023 MUC cycle, MUC2022-060 entitled First Year Standardized Waitlist Ratio (FYSWR), which tracks the number of incident patients in a practitioner (inclusive of physicians and advanced practice providers) group who are under the age of 75 and were listed on the kidney or kidney-pancreas transplant waitlist or received a living donor transplant within the first year of initiating dialysis, inviting a possible duplication of effort or even confusion by providers if both measures were incorporated into the MIPS program. This MUC is a preferable alternative to

MUC2022-060, as one of the measure constructs incorporated was recommended for endorsement by the Renal Standing Committee (MUC2022-060 did not receive endorsement from NQF). The developer notes that this MUC distinguishes itself from MUC2022-060 by focusing on all patients in a dialysis practitioner group practice on the waitlist, instead of patients in their first year of dialysis.

Summary: What is the potential impact of this measure on quality of care for patients?

Dialysis and renal transplants are the result of chronic kidney disease which according to the Centers for Disease Control and Prevention (CDC) affects 37 million people in the United States. In testing, the developer found that over 281,000 patients were included in the measure calculation. Adding patients to the waitlist is a precursor to obtaining a transplanted kidney, which is very likely to improve clinical outcomes. In the Spring 2022 Consensus Development Process evaluation of both the PPPW and aPPPW measures by the National Quality Forum (NQF), the Renal Standing Committee was in near-universal agreement there was at least a moderate performance gap. In testing, the developer found a mean performance score of 19.1 percent with a standard deviation of 8.1 percent for the PPPW measure, and 11.9 percent with a standard deviation of 6.9 percent for the aPPPW measure.

Preliminary Analysis – MUC2022-065 Preventive Care and Wellness (composite)

Measure Description:

Percentage of patients who received age- and sex-appropriate preventive screenings and wellness services. This measure is a denominator-weighted composite of seven component measures that are based on recommendations for preventive care by the U.S. Preventive Services Task Force (USPSTF), Advisory Committee on Immunization Practices (ACIP), American Association of Clinical Endocrinology (AACE), and American College of Endocrinology (ACE).

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC) does not address a specific priority for the Merit-based Incentive Payment System (MIPS). However, the developer notes that the Centers for Medicare & Medicaid Services (CMS) prioritized this composite measure for development because, as a composite measure, it can more easily summarize a clinician's performance across these related quality concepts, it is useful for public reporting and other accountability applications, and it provides a more reliable measure result. In addition, the composite is composed of measures already in the MIPS measure set.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The measure consists of seven preventive care and screening processes that are consistent with guidelines from the United States Preventive Services Task Force (USPSTF), the Advisory Committee on Immunization Practices (ACIP), the American Association of Clinical Endocrinology (AACE), and the American College of Endocrinology (ACE), including influenza immunization, pneumococcal immunization, breast and colorectal cancer screening, body mass index screening, tobacco use screening and cessation intervention, and screening for high blood pressure with follow-up. Each recommendation is at least "Strong" or equivalent. The developer identified a study by Fox and Shaw (2015) that highlighted the lifesaving potential of preventive services, though these are not necessarily specific to concepts captured in this MUC. The developer also provided at least one citation for each clinical outcome/condition associated with the preventive care process identified in the MUC, finding that the resulting episodes of care or annual chronic conditions costs were quite high. For example, an average episode of care for pneumonia costs the U.S. health care system $3,151 (Tong et al, 2018).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: In testing, the developer identified a performance gap, where median performance on the measure was 52.7 percent, with a standard deviation of 11.2 percent. However, one measure, Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan (CMIT: 05835-E-MIPS is identified by the MIPS Historical Quality Benchmarks in 2022 as being topped out. Note that when the Measure Applications Partnership (MAP) reviewed the measure during the 2020-2021 pre-rulemaking cycle, this MUC (MUC20-0043) was recommended as "Do Not Support with Potential for

Mitigation", in part owing to concern from the MAP Coordinating Committee that some of the measure components are topped out. Although the specifications have not changed, the developer has provided additional performance data for the MUC not available in the prior review.

The developer conducted a topped out analysis using data obtained from four primary care and specialty clinical sites, which were located in a mix of rural and urban areas and which used two different electronic health record (EHR) systems. Using a sample of 89 clinicians from four clinician networks who saw at least 11 unique patients and had eligible cases for at least two component measures, the developer reported the difference between the 75th and 90th percentiles (5.4) was greater than two times the standard deviation for the 90th percentile (2.2). According to the developer, these results indicated that the Preventive Care and Wellness composite was not topped out.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: Each component measure is already included in the MIPS program. However, a valid composite measure presents some advantages over individual measures: a composite can more easily summarize a clinician's performance, is useful for public reporting and other accountability applications, as well as provides a more reliable measure result.

Note that when MAP reviewed this measure during the 2020-2021 pre-rulemaking cycle, this MUC (MUC20-0043), was recommended as "Do Not Support with Potential for Mitigation," partly owing to concern from the MAP Coordinating Committee that the MUC would replace the individual measures already in the program, when individual rates would still be useful for interpreting the results of the overall composite. In this current pre-rulemaking cycle, the MUC has been submitted again as a fully developed measure with additional performance data. In this submission, the developer notes the potential for removal of the individual measures, consistent with CMS's program requirements to eliminate duplicate measures. However, the developer presented results of clinician interviews that found that they universally would need the individual component measure scores reported separately, and noted widespread support for retaining the individual component measures from a Technical Expert Panel, a patient and family caregiver workgroup, and public comments.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: All seven components of this composite measure are currently individually reported separately in the MIPS program.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The MUC has not yet been submitted for endorsement; however, it is fully specified for the appropriate level of analysis (clinicians). The developer conducted reliability testing finding high reliability of the measure result in both signal-to-noise testing (0.977) and test-retest (Spearman rho was 0.971, mean intra-class correlation was 0.967). In validity testing, the developer found a strong relationship between the component scores, lending support to the construct validity of the composite. As well, the developer found moderate to strong effects of patient characteristics on the

measure result consistent with their hypothesis that performance would be lower for male, non-white, self-pay and Medicaid patients. Four of the seven component measures are endorsed by the consensus-based entity (National Quality Forum (NQF) #0041, 2372, 0034, 0028); however, there are differences between the endorsed versions of the measures and the components. For example, two of the measures are endorsed at the health plan and integrated system delivery level, but not the clinician level, and do not have encounter requirements in the denominator (which the component measures do). All seven individual components are already in use in the MIPS program.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The developer reported potential unintended consequences raised during by their Technical Expert Panel, including a potential for gaming the measure result through patient selection, and potential inaccuracies in the measure result based on uneven documentation from electronic health record systems. However, these concerns would be equally reflected in the component measures that are already included in the MIPS program, and no evidence was offered that these concerns have materialized thus far.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Support for Rulemaking

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) does not address a specific priority for the Merit-based Incentive Payment System (MIPS). However, the developer notes that the Centers for Medicare & Medicaid Services (CMS) prioritized this composite measure for development because, as a composite measure, it can more easily summarize a clinician's performance across these related quality concepts, it is useful for public reporting and other accountability applications, and it provides a more reliable measure result. In addition, the composite is composed of measures already in the MIPS measure set.

The Measure Applications Partnership (MAP) reviewed this measure (MUC20-0043) during the 2020-2021 pre-rulemaking cycle, giving it a "Do Not Support with Potential for Mitigation", partly owing to concern from the MAP Coordinating Committee that the MUC would replace the individual measures already in the program, when individual rates would still be useful for interpreting the results of the overall composite. In this submission, the developer again notes the potential for removal of the individual measures, consistent with CMS's program requirements to eliminate duplicate measures. However, the developer presented results of clinician interviews that found that they universally would need the individual component measure scores reported separately, and noted widespread support for retaining the individual component measures from a Technical Expert Panel, and patient and family caregiver workgroup, and public comments.

Summary: What is the potential impact of this measure on quality of care for patients?

The measure consists of seven preventive care and screening processes that are consistent with guidelines from the United States Preventive Services Task Force (USPSTF), the Advisory Committee on Immunization Practices (ACIP), the American Association of Clinical Endocrinology (AACE), and the American College of Endocrinology (ACE), including influenza immunization, pneumococcal immunization, breast and colorectal cancer screening, body mass index screening, tobacco use screening and cessation intervention, and screening for high blood pressure with follow-up. Each recommendation is at least "Strong" or equivalent.

In testing, the developer identified a performance gap, where median performance on the measure was 52.7 percent, with a standard deviation of 11.2 percent. Note that when the MAP reviewed the measure during the 2020-2021 pre-rulemaking cycle, the MAP Coordinating Committee had concerns that some of the measure components are topped out. Although the specifications have not changed, the developer has provided additional performance data for the MUC not available in the prior review which indicates the composite measure is not topped out.

Preliminary Analysis – MUC2022-098 Connection to Community Service Provider

Measure Description:

Percent of patients 18 years or older who screen positive for one or more of the following health related social needs (HRSNs): food insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety; and had contact with a Community Service Provider (CSP) for at least 1 of their HRSNs within 60 days after screening.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC) addresses health equity, a high priority area for measures within the Merit-based Incentive Payment System (MIPS) program, and a CMS Meaningful Measure 2.0 priority. Advancing health equity is also a goal of the CMS National Quality Strategy.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The measure developer indicates that screening for HRSNs is supported by the U.S. Preventive Services Task Force (USPSTF). In a USPSTF technical brief, they note that social risk factors are mentioned in two-thirds of USPSTF recommendation statements, and six other professional medical organizations explicitly promote clinician engagement in social risk screening and referrals (Eder et al., 2021). The measure developer indicates that the USPSTF has a recommendation related to screening for an HRSN that is somewhat similar to one of the HRSNs in the denominator; however, the recommendation is for a narrower population than the measure and is related to only one of the HRSNs in the measure denominator.

The Accountable Health Communities (AHC) Model evaluated whether connecting Medicare and Medicaid beneficiaries to community resources can improve health outcomes using universal screening. Findings from the first evaluation report of the Accountable Health Communities (AHC) Model found promising results indicating that 74 percent of eligible beneficiaries accepted navigation related to their HRSNs (Accountable Health Communities (AHC) Model Evaluation, 2020), and, of those, 14 percent reported that at least one of their HRSNs was resolved (Johnson et al., 2022). A systematic review indicated that screening and referral programs positively affect outcomes relating to experience of care and population health (Escobar et al., 2021), though did not draw definitive conclusions owing to potential biases in the research.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: CMS identifies "measures that reflect social and economic determinants" as a key measure gap. Studies indicate that only 24 percent of hospitals and 16 percent of physician practices are screening for all five Driver of Health (DOH) domains (food insecurity, housing instability, transportation problems, utility assistance needs, and interpersonal safety) (Fraze et al., 2022).

While the developer did not provide evidence of variation in care at the clinician-level, the Accountable Healthcare Communities (AHC) model found that nearly 60 percent of patients eligible for navigation or referral services had two or more HRSNs and 74 percent of patients accepted navigation related to their HRSN, highlighting the need for measures to address screening. Of those, 14 percent of patients reported at least one of their HRSNs was resolved (Johnson et al., 2022).

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: A screening measure related to screening for health-related social needs (HRSNs), MUC2021-136 Screening for Social Drivers of Health, was considered by the Measure Applications Partnership (MAP) as part of the 2021-2022 pre-rulemaking process, where it received Conditional Support for Rulemaking; it was then finalized for adoption in MIPS in the Calendar Year (CY) 2023 Physician Fee Schedule (PFS) Final Rule. The MUC builds on the Screening for Social Drivers of Health measure by assessing whether patients who screened positive for at least one HRSN had contact with a community service provider (CSP) for at least one of their HRSNs within 60 days after screening. Another measure, MUC2022-111 Resolution of At Least 1 Health-Related Social Need, is proposed for the Merit-based Incentive Payment System and submitted by the same measure developer, however, it focuses on the resolution of the patient's HRSN within 12 months after screening.

The goal of this MUC is to ultimately make sure that the patient is connected to resources via the provider or entity for the identified HRSN and not just screened for HRSN. It builds on the screening measure by assessing the first step in resolving the HRSN and provides a middle ground between the screening measure and the resolution measure.

Can the measure be feasibly reported?

Yes/No: No

Justification and Notes: For this measure, some data elements are in defined fields in electronic sources. The measure developer notes that various data sources are used (administrative data, electronic clinical data, electronic health record, standardized patient assessments, and patient reported data and surveys).

The measure uses patient reported data and standardized assessments to determine if patients match the denominator (i.e., patients who have a positive result for at least one HRSN). The measure uses electronic health record (EHR)-and non-EHR electronic clinical data, as well as patient reported data, to determine whether the patient had contact with a CSP. The developer cites the current implementation of this measure in the CMS Accountable Healthcare Communities Program as evidence that data collection is feasible, but notes that a workflow analysis found that provider workflow would have to be modified to capture the data elements needed to capture the performance measure.

The developer does not provide information on feasibility outside demonstrations/Center for Medicare & Medicaid Innovation (CMMI) models or community health centers, so it is unknown whether this measure will be feasible for clinicians not participating in those programs.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: No

Justification and Notes: This measure is not endorsed by a consensus-based entity (CBE), and the current state of development as indicated by the measure developer is field (beta) testing. The measure developer provides reliability and validity results from Accountable Health Communities (AHC) testing of their validated screening tool, although no information is provided on the reliability or validity of the measure result itself.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure developer notes that one potential unintended consequence of the measure is that health systems and hospitals will not be equipped to act on it due, in part, to the lack of community resources. The Accountable Health Communities evaluation noted that some communities worked to address these institutional and structural barriers by considering eligibility requirements when making referrals (Accountable Health Communities (AHC) Model Evaluation, 2020).

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on testing indicating the measure is reliable, valid, and feasible, and endorsement by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) addresses health equity, a high priority area for measures within the Merit-based Incentive Payment System (MIPS) program, and a CMS Meaningful Measure 2.0 priority. Advancing health equity is also a goal of the CMS National Quality Strategy.

This MUC builds on a measure assessing screening for health-related social needs (HRSNs) (MUC2021-136 Screening for Social Drivers of Health) that received Conditional Supporting for Rulemaking from the Measure Applications Partnership (MAP) during the 2021-2022 pre-rulemaking process and that was

subsequently finalized for adoption in MIPS in the Calendar Year (CY) 2023 Physician Fee Schedule (PFS) Final Rule.

The goal of this MUC is to ultimately make sure that the patient is connected to resources via the provider or entity for the identified HRSN and not just screened for HRSN. It builds on the screening measure by assessing the first step in resolving the HRSN and provides a middle ground between the screening measure and a resolution of the HSRN measure (MUC2022-111), also submitted for the 2022-2023 pre-rulemaking cycle.

Summary: What is the potential impact of this measure on quality of care for patients?

Studies indicate that only 24 percent of hospitals and 16 percent of physician practices are screening for all five Driver of Health (DOH) domains (food insecurity, housing instability, transportation problems, utility assistance needs, and interpersonal safety) (Fraze et al., 2022).

While the developer did not provide evidence of variation in care at the clinician-level, the Center for Medicare & Medicaid Innovation (CMMI) Accountable Healthcare Communities (AHC) model found that nearly 60 percent of patients eligible for navigation or referral services had two or more HRSNs and 74 percent of patients accepted navigation related to their HRSN, highlighting the need for measures to address screening. Of those, 14 percent of patients reported at least one of their HRSNs was resolved (Johnson et al., 2022). A systematic review indicated that screening and referral programs positively affect outcomes relating to experience of care and population health (Escobar et al., 2021), though did not draw definitive conclusions owing to potential biases in the research.

While the measure is not strongly supported by the evidence and there is a lack of information about the measure's reliability, validity, and feasibility, addressing disparities in care is a high priority for the healthcare system. This measure will contribute to the connection of patients with community service providers for their health-related social needs, which will help to impact disparities.

Preliminary Analysis – MUC2022-111 Resolution of At Least 1 Health-Related Social Need

Measure Description:

Percent of patients 18 years or older who screen positive for one or more of the following HRSNs: food insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety; and report that at least 1 of their HRSNs was resolved within 12 months after screening.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This outcome measure addresses health equity, a high priority area for measures within the Merit-based Incentive Payment System (MIPS) and the Centers for Medicare & Medicare Services (CMS) Meaningful Measure 2.0. Advancing health equity is also a goal of the CMS National Quality Strategy.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: No

Justification and Notes: The measure developer indicates that screening for HRSNs is supported by the U.S. Preventive Services Task Force (USPSTF). In a USPSTF technical brief, they note that social risk factors are mentioned in two-thirds of USPSTF recommendation statements, and six other professional medical organizations explicitly promote clinician engagement in social risk screening and referrals (Eder et al., 2021). The measure developer indicates that the USPSTF has a recommendation related to screening for an HRSN that is somewhat similar to one of the HRSNs in the denominator; however, the recommendation is for a narrower population than the measure and is related to only one of the HRSNs in the measure denominator.

While the developer did not specifically describe interventions that could improve this particular outcome (resolving a HRSN) at the clinician level, they do reference several Center for Medicare & Medicaid Innovation (CMMI) models that have incorporated Driver of Health (DOH) screening and navigation data into their quality frameworks and care management plans for beneficiaries. They also reference the Accountable Healthcare Communities (AHC), which is using two tracks to test interventions to address HRSNs: (1) the Assistance Track, which tests universal screening to identify Medicare and Medicaid beneficiaries with HRSNs and provision of navigation assistance to connect navigation-eligible beneficiaries with the community services they need; and (2) the Alignment Track, which tests universal screening, referral, and navigation combined with engaging key stakeholders in community-level continuous quality improvement to align community service capacity with the community's service needs.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: CMS identifies "measures that reflect social and economic determinants" as a key measure gap. Studies indicate that only 24 percent of hospitals and 16 percent of physician practices are screening for all five Driver of Health (DOH) domains (food insecurity, housing instability,

transportation problems, utility assistance needs, and interpersonal safety) (Fraze et al., 2022).

While the developer did not provide evidence of variation in care at the clinician-level, the Accountable Healthcare Communities (AHC) model found that only 14 percent of those who completed a full year of navigation had any HRSNs documented as resolved and an additional four percent had been connected with a Community Service Provider (CSP) but had not resolved any HRSNs (Accountable Health Communities (AHC) Model Evaluation, 2020).

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: A screening measure related to screening for health-related social needs (HRSNs), MUC2021-136 Screening for Social Drivers of Health, was considered by the Measure Applications Partnership (MAP) as part of the 2021-2022 pre-rulemaking process, where it received Conditional Support for Rulemaking; it was then finalized for adoption in MIPS in the Calendar Year (CY) 2023 Physician Fee Schedule (PFS) Final Rule. This measure under consideration (MUC) builds on the Screening for Social Drivers of Health measure by assessing whether patients who screened positive for at least one HRSN had at least one of their HRSNs resolved within 60 days after screening. Another measure, MUC2022-111 Connection to Community Service Provider, is proposed for the Merit-based Incentive Payment System (MIPS) and submitted by the same measure developer, however, it focuses on whether patients with at least one HRSN are connected to a community service provider for their HRSN.

The goal of this measure is to focus on the resolution of HRSNs, therefore, it is not duplicative of any other measure proposed for MIPS. It builds on the screening measure by assessing the outcome associated with screening – whether the HRSN is resolved.

Can the measure be feasibly reported?

Yes/No: No

Justification and Notes: For this measure, some data elements are in defined fields in electronic sources. The measure developer notes that various data sources are used (administrative data, electronic clinical data, electronic health record, standardized patient assessments, and patient reported data and surveys). The measure uses patient reported data and standardized assessments to determine if patients match the denominator (i.e., patients who have a positive result for at least one HRSN). The measure developer notes that the measure, by design, is not prescriptive about the use of specific tools to establish patients who have been screened for HRSNs listed in the denominator.

The developer cites the current implementation of this measure in the CMS Accountable Healthcare Communities program as evidence that data collection is feasible, but also notes that a workflow analysis found that provider workflow would have to be modified to capture the data elements needed to capture the performance measure. The developer does not provide information on feasibility outside demonstrations/CMMI models or community health centers, so it is unknown whether this measure will be feasible for clinicians not participating in those programs.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: No

Justification and Notes: This measure is not endorsed by a consensus-based entity (CBE), and the current state of development as indicated by the measure developer is field (beta) testing. The measure developer provides reliability and validity results from Accountable Health Communities (AHC) testing of their validated screening tool, although no information is provided on the reliability or validity of the measure result itself.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure developer notes that one potential unintended consequence of the measure is that health systems and hospitals will not be equipped to act on it due, in part, to the lack of community resources. The Accountable Health Communities evaluation noted that some communities worked to address these institutional and structural barriers by considering eligibility requirements when making referrals (Accountable Health Communities (AHC) Model Evaluation, 2020).

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on testing indicating the measure is reliable, valid, and feasible, and endorsement by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This outcome measure addresses health equity, a high priority area for measures within the Merit-based Incentive Payment System (MIPS) and the Centers for Medicare & Medicare Services (CMS) Meaningful Measure 2.0. Advancing health equity is also a goal of the CMS National Quality Strategy.

This measure under consideration (MUC) builds on a measure assessing screening for health-related social needs (HRSNs) (MUC2021-136 Screening for Social Drivers of Health) that received Conditional Supporting for Rulemaking from the Measure Applications Partnership (MAP) during the 2021-2022 pre-

rulemaking process and that was subsequently finalized for adoption in MIPS in the Calendar Year (CY) 2023 Physician Fee Schedule (PFS) Final Rule. The goal of this measure is to focus on the resolution of HRSNs, therefore, it is not duplicative of any other measure proposed for MIPS. It builds on the screening measure by assessing the outcome associated with screening – whether the HRSN is resolved.

Summary: What is the potential impact of this measure on quality of care for patients?

Studies indicate that only 24 percent of hospitals and 16 percent of physician practices are screening for all five Driver of Health (DOH) domains (food insecurity, housing instability, transportation problems, utility assistance needs, and interpersonal safety) (Fraze et al., 2022). In addition, an evaluation of the Center for Medicare & Medicaid Innovation (CMMI) Accountable Health Communities Model found that only 14 percent of patients who completed a full year of navigation had any HRSNs documented as resolved (Accountable Health Communities (AHC) Model Evaluation, 2020).

While the measure is not strongly supported by evidence indicating that interventions can improve this particular outcome (resolving a HRSN) at the clinician level, and there is a lack of information about the measure's reliability, validity, and feasibility, addressing disparities in care is a high priority for the healthcare system. This measure will contribute to the resolution of health-related social needs, which will help to impact disparities.

Preliminary Analysis – MUC2022-114 Appropriate screening and plan of care for elevated intraocular pressure following intravitreal or periocular steroid therapy

Measure Description:

Percentage of patients without a diagnosis of glaucoma who had an intravitreal or periocular corticosteroid injection (e.g., triamcinolone, preservative-free triamcinolone, dexamethasone, dexamethasone intravitreal implant, or fluocinolone intravitreal implant) who, within seven (7) weeks following the date of injection, are screened for elevated intraocular pressure (IOP) with tonometry with documented IOP =<25 mm Hg for injected eye OR if the IOP was >25 mm Hg, a plan of care was documented.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure under consideration (MUC) addresses the Merit-based Incentive Payment System (MIPS) priority area of patient safety and the Meaningful Measure domain of Preventable Healthcare Harm. There are currently no other measures in the program set which directly measure intraocular pressure (IOP) after corticosteroid injections.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The developer states that current clinical guidelines do not address the need to assess for elevated IOP following corticosteroid injection; however, they presented data that demonstrate patients treated with corticosteroid therapy are at increased risk for elevated IOP leading to steroid induced glaucoma, visual impairment and overall poor quality of life (Breusegem, 2009; Haller, 2011). Several randomized clinical trials and a systematic review identified that IOPs typically peak around seven to nine weeks (Haller, 2010; Kiddee, 2013; Aref, 2015).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: Data collected from two practices representing 19 retina specialists from the calendar year 2021 showed performance varied from 25-100 percent with a mean performance score of 71.38 percent. The steward suggests further performance variation and gaps in care would be demonstrated by testing at non-specialist practice sites (i.e., non-fellowship general ophthalmologists), but this data has not been collected. A systematic review reported 10.9-79.0 percent of patients develop clinically significant IOP elevations following corticosteroid injection. Variations in IOP elevations are based on the type and dose of steroid as well as various patient risk factors. IOP elevation can also be affected by concurrent use of topical or oral steroids. The MUC is intended to reduce negative outcomes associated with intravitreal or periocular steroid therapy.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: Currently there are no measures in MIPS related to the screening and plan of care for elevated IOP following intravitreal or periocular steroid therapy. However, there are 15 measures applicable to ophthalmologists in the 2022 MIPS measure set, indicating that there may already be enough measures for that specialty in MIPS. While the developer did not provide data indicating that patients find this measure to be valuable, this measure may be important for retina specialists.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the measure developer, some data elements are in defined fields in electronic sources. The developer previously assessed the feasibility of collecting the required data elements of a similar measure across three practices with two different electronic health records (EHRs). The majority of the required data elements for this measure were found to be. Additional testing of this updated measure further demonstrated that two practices were able to collect and report the required data elements.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The MUC has not yet been submitted to a consensus-based entity (CBE) for endorsement. The MUC is specified for the individual clinician level of analysis. The developer identified it as in the beta testing state of development. Measure score reliability testing using signal-to-noise testing was completed across 17 physicians at two retina specialty practices using data from calendar year 2021. The total number of patients included in the analysis was 556. The developer reported a median reliability of 0.828 across the 17 physicians with five or more eligible cases suggesting good reliability.

The developer conducted a face validity assessment with 15 physicians. The developer calculated a mean score across four questions related to face validity. Across the four questions, the mean score ranged from 4.33-4.53, indicating the physicians in the assessment supported the face validity of the measure. Previous testing of a similar measure provided an assessment of the overall reliability of the electronic health record (EHR) extract versus manual abstraction, which resulted in a prevalence adjusted kappa of 0.81 (95% CI: 0.66 –0.96) for the denominator and a prevalence adjusted kappa of 0.77 (95% CI: 0.44-0.79) for the numerator. The developer has not yet validated the data elements for a plan of care in the measure numerator.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The MUC is not currently in use. The developer did not identify any unintended consequences during testing of the measure.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure under consideration (MUC) addresses the Merit-based Incentive Payment System (MIPS) priority area of patient safety and the Meaningful Measure domain of Preventable Healthcare Harm. There are currently no other measures in the program set which directly measure intraocular pressure (IOP) after corticosteroid injections.

Currently there are no measures in MIPS related to the screening and plan of care for elevated IOP following intravitreal or periocular steroid therapy. However, there are 15 measures applicable to ophthalmologists in the 2022 MIPS measure set, indicating that there may already be enough measures for that specialty in MIPS. While the developer did not provide data indicating that patients find this measure to be valuable, this measure may be important for retina specialists.

Summary: What is the potential impact of this measure on quality of care for patients?

The measure developer states that current clinical guidelines do not address the need to assess for elevated IOP following corticosteroid injection; however, they presented data that demonstrate patients treated with corticosteroid therapy are at increased risk for elevated IOP leading to steroid induced glaucoma, visual impairment and overall poor quality of life (Breusegem, 2009; Haller, 2011). Several randomized clinical trials and a systematic review identified that IOPs typically peak around seven to nine weeks (Haller, 2010; Kiddee, 2013; Aref, 2015).

Preliminary Analysis – MUC2022-115 Acute posterior vitreous detachment appropriate examination and follow-up

Measure Description:

Percentage of patients with a diagnosis of acute posterior vitreous detachment (PVD) in either eye who were appropriately evaluated during the initial exam and were re-evaluated no later than 8 weeks

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: The measure under consideration (MUC) addresses the Merit-based Incentive Payment System (MIPS) priority area of patient safety and the Meaningful Measure domain of Preventable Healthcare Harm.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: According to the measure developer, retinal tears, which most often occur in the setting of an acute PVD, are less likely to result in detachment if treated promptly (AAO, 2019, ASRS, 2016). Prompt treatment may minimize complications, such as retinal detachment, and improve a patient's quality of life (AAO, 2019).

The current guideline published by the American Academy of Ophthalmology on posterior vitreous detachment (PVD) and retinal breaks supports this MUC. The guideline states, "selected patients, particularly those with any degree of vitreous pigment, vitreous or retinal hemorrhage, or visible vitreoretinal traction, should be asked to return for a second examination promptly if they have new symptoms or within 6 weeks following the onset of PVD symptoms (AAO, 2019)." The developer notes the grading of evidence was strong for the initial exam recommendations but was discretionary for the follow-up examination recommendations (based on evidence presented in the 2014 AAO Preferred Practice Pattern Guidelines; evidence citations remained the same in the 2019 release).

Does the measure address a quality challenge?

Yes/No: No

Justification and Notes: This MUC is a process measure addressing the risk of retinal complications associated with patients diagnosed with acute PVD. In some cases, untreated retinal tears can lead to retinal detachment and/or vision loss. Data from the developer's performance gap testing on 19 physicians across two practices showed performance varied from 0 percent to 5.31 percent, representing a limited degree of performance variation.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: Currently there are no measures in MIPS which address the appropriate screening and follow-up for patients with PVD and at risk of retinal tears. There are two related outcome

measures for adults who have undergone rhegmatogenous retinal detachment surgery—Adult Primary Rhegmatogenous Retinal Detachment Surgery: No Return to the Operating Room Within 90 Days of Surgery (CMIT 02537-C-MIPS) and Adult Primary Rhegmatogenous Retinal Detachment Surgery: Visual Acuity Improvement Within 90 Days of Surgery (CMIT 02381-C-MIPS). The developer also submitted MUC2022-116 during the 2022-2023 pre-rulemaking cycle which assesses the percentage of patients with a diagnosis of acute PVD and acute vitreous hemorrhage in either eye who were appropriately evaluated during the initial eye exam and re-evaluated no later than 2 weeks.

There are 15 measures applicable to ophthalmologists in the 2022 MIPS measure set, indicating that there may already be enough measures for that specialty in MIPS. However, this measure may be important for retina specialists.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the measure developer, some data elements for this measure are in defined fields in electronic sources. The developer previously assessed the feasibility of collecting the required data elements for this measure across three practices with two different EHRs (electronic health records). The majority of the required data elements for this measure were found to be feasible. Additional testing of this updated measure further demonstrated that two practices were able to collect and report the required data elements.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The MUC has not yet been submitted to a consensus-based entity (CBE) for endorsement. The MUC is specified for the individual clinician level of analysis. The developer identified it as in the beta testing state of development. The developer completed signal-to-noise reliability testing using a beta-binomial model at the clinician level. Testing included physicians with at least five eligible cases, which resulted in a sample size of 19 physicians at two retina specialist practices. The total number of patients included in the analysis was 6,609. Using a beta-binomial method, the developer reported a median reliability of 0.978.

The developer conducted a face validity assessment with 15 physicians. The developer calculated a mean score across four questions related to face validity. Across the four questions, the mean score ranged from 4.33-4.47, indicating the physicians in the assessment supported the face validity of the measure. Previous testing of a similar measure provided an assessment of the overall reliability of the electronic health record (EHR) extract versus manual abstraction, which resulted in a prevalence adjusted kappa of 1.0 (95% CI: n/a) for the denominator, a prevalence adjusted kappa of 0.87 (95% CI: 0.58-0.95) for numerator 1, and a prevalence adjusted kappa of 0.81 (95%CI: 0.58-0.96) for numerator 2.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The MUC is not currently in use. The developer did not identify any negative unintended consequences during testing.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Do Not Support for Rulemaking

Summary: What is the potential value to the program measure set?

The measure under consideration (MUC) addresses the Merit-based Incentive Payment System (MIPS) priority area of patient safety and the Meaningful Measure domain of Preventable Healthcare Harm.

Currently there are no measures in MIPS which address the appropriate screening and follow-up for patients with PVD and at risk of retinal tears. There are two related outcome measures for adults who have undergone rhegmatogenous retinal detachment surgery—Adult Primary Rhegmatogenous Retinal Detachment Surgery: No Return to the Operating Room Within 90 Days of Surgery (CMIT 02537-C-MIPS) and Adult Primary Rhegmatogenous Retinal Detachment Surgery: Visual Acuity Improvement Within 90 Days of Surgery (CMIT 02381-C-MIPS). The developer also submitted MUC2022-116 during the 2022-2023 pre-rulemaking cycle which assesses the percentage of patients with a diagnosis of acute PVD and acute vitreous hemorrhage in either eye who were appropriately evaluated during the initial eye exam and re-evaluated no later than 2 weeks.

There are 15 measures applicable to ophthalmologists in the 2022 MIPS measure set, indicating that there may already be enough measures for that specialty in MIPS. However, this measure may be important for retina specialists.

Summary: What is the potential impact of this measure on quality of care for patients?

According to the measure developer, retinal tears, which most often occur in the setting of an acute PVD, are less likely to result in detachment if treated promptly (AAO, 2019, ASRS, 2016). Prompt treatment may minimize complications, such as retinal detachment, and improve a patient's quality of life (AAO, 2019).

The current guideline published by the American Academy of Ophthalmology on posterior vitreous detachment (PVD) and retinal breaks supports this MUC. The guideline states, "selected patients, particularly those with any degree of vitreous pigment, vitreous or retinal hemorrhage, or visible

vitreoretinal traction, should be asked to return for a second examination promptly if they have new symptoms or within 6 weeks following the onset of PVD symptoms (AAO, 2019)." The developer notes the grading of evidence was strong for the initial exam recommendations but was discretionary for the follow-up examination recommendations (based on evidence presented in the 2014 AAO Preferred Practice Pattern Guidelines; evidence citations remained the same in the 2019 release).

Data from the developer's performance gap testing on 19 physicians across two practices showed performance varied from 0 percent to 5.31 percent, representing a limited degree of performance variation.

Preliminary Analysis – MUC2022-116 Acute posterior vitreous detachment and acute vitreous hemorrhage appropriate examination and follow-up

Measure Description:

Percentage of patients with a diagnosis of acute posterior vitreous detachment (PVD) and acute vitreous hemorrhage in either eye who were appropriately evaluated during the initial exam and were re-evaluated no later than 2 weeks

Does the measure address a critical quality objective not currently adequately addressed by th e measures in the program set?

Yes/No: Yes

Justification and Notes: The measure under consideration (MUC) addresses the Merit-based Incentive Payment System (MIPS) priority area of patient safety and the Meaningful Measure domain of Preventable Healthcare Harm. There are no measures currently in the measure set which address prevention of retinal tears or complications of acute posterior vitreous detachment (PVD).

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: According to the measure developer, retinal tears, which most often occur in the setting of an acute PVD, are less likely to result in detachment if treated promptly (AAO, 2019, ASRS, 2016). Prompt treatment may minimize complications, such as retinal detachment, and improve a patient's quality of life (AAO, 2019).

The current guideline published by the American Academy of Ophthalmology on posterior vitreous detachment (PVD) and retinal breaks supports this MUC. The guideline states, "selected patients, particularly those with any degree of vitreous pigment, vitreous or retinal hemorrhage, or visible vitreoretinal traction, should be asked to return for a second examination promptly if they have new symptoms or within 6 weeks following the onset of PVD symptoms (AAO, 2019)." The developer notes the grading of evidence was strong for the initial exam recommendations but was discretionary for the follow-up examination recommendations (based on evidence presented in the 2014 AAO Preferred Practice Pattern Guidelines; evidence citations remained the same in the 2019 release).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: This MUC is a process measure addressing the risk of retinal complications associated with patients diagnosed with acute PVD with vitreous hemorrhage. In some cases, untreated retinal tears can lead to retinal detachment and/or vision loss. Data from the developer's performance gap testing on 19 physicians across two practices showed performance varied from 0 to 38.10 percent, indicating variation and a gap in performance.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: Currently there are no measures in MIPS which address the appropriate screening and follow-up for patients with PVD and at risk of retinal tears. There are two related outcome measures for adults who have undergone rhegmatogenous retinal detachment surgery—Adult Primary Rhegmatogenous Retinal Detachment Surgery: No Return to the Operating Room Within 90 Days of Surgery (CMIT 02537-C-MIPS) and Adult Primary Rhegmatogenous Retinal Detachment Surgery: Visual Acuity Improvement Within 90 Days of Surgery (CMIT 02381-C-MIPS). The developer also submitted MUC2022-115 during the 2022-2023 pre-rulemaking cycle which assesses the percentage of patients with a diagnosis of acute PVD in either eye who were appropriately evaluated during the initial exam and were re-evaluated no later than 8 weeks.

There are 15 measures applicable to ophthalmologists in the 2022 MIPS measure set, indicating that ophthalmologists may have a sufficient number of there may already be enough measures for that specialty in MIPS. However, this measure may be important for retina specialists.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: According to the measure developer, some data elements for this measure are in defined fields in electronic sources. The developer previously assessed the feasibility of collecting the required data elements for this measure across three practices with two different EHRs (electronic health records). The majority of the required data elements for this measure were found to be feasible. Additional testing of this updated measure further demonstrated that two practices were able to collect and report the required data elements.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The MUC has not yet been submitted to a consensus-based entity for endorsement. The MUC is specified for the individual clinician level of analysis and the developer noted that the measure is in beta testing. The developer completed signal-to-noise reliability testing using a beta-binomial model at the clinical level. Testing included physicians with at least five eligible cases, which resulted in a sample size of 18 physicians at two retina specialist practices. The total number of patients included in the analysis was 455. Using a beta-binomial method, the developer reported a median reliability of 0.973.

The developer conducted a face validity assessment with 15 physicians. The developer calculated a mean score across four questions related to face validity. Across the four questions, the mean score ranged from 4.33 to 4.40, indicating the physicians in the assessment supported the face validity of the measure. Previous testing of a similar measure provided an assessment of the overall reliability of the electronic health record (EHR) extract versus manual abstraction, which resulted in a prevalence adjusted kappa of 0.88 (95% CI: 0.72-0.90) for the denominator, a prevalence adjusted kappa of 0.87 (95% CI: 0.58-0.95) for numerator 1, and a prevalence adjusted kappa of 0.79(95%CI: 0.68-0.89) for numerator 2.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The MUC is not currently in use. The developer did not identify any negative unintended consequences during testing.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

The measure under consideration (MUC) addresses the Merit-based Incentive Payment System (MIPS) priority area of patient safety and the Meaningful Measure domain of Preventable Healthcare Harm. There are no measures currently in the measure set which address prevention of retinal tears or complications of acute posterior vitreous detachment (PVD).

Currently there are no measures in MIPS which address the appropriate screening and follow-up for patients with PVD and at risk of retinal tears. There are two related outcome measures for adults who have undergone rhegmatogenous retinal detachment surgery—Adult Primary Rhegmatogenous Retinal Detachment Surgery: No Return to the Operating Room Within 90 Days of Surgery (CMIT 02537-C-MIPS) and Adult Primary Rhegmatogenous Retinal Detachment Surgery: Visual Acuity Improvement Within 90 Days of Surgery (CMIT 02381-C-MIPS). The developer also submitted MUC2022-115 during the 2022-2023 pre-rulemaking cycle which assesses the percentage of patients with a diagnosis of acute PVD in either eye who were appropriately evaluated during the initial exam and were re-evaluated no later than 8 weeks.

There are 15 measures applicable to ophthalmologists in the 2022 MIPS measure set, indicating that ophthalmologists may have a sufficient number of there may already be enough measures for that specialty in MIPS. However, this measure may be important for retina specialists.

Summary: What is the potential impact of this measure on quality of care for patients?

According to the measure developer, retinal tears, which most often occur in the setting of an acute PVD, are less likely to result in detachment if treated promptly (AAO, 2019, ASRS, 2016). Prompt treatment may minimize complications, such as retinal detachment, and improve a patient's quality of life (AAO, 2019).

The current guideline published by the American Academy of Ophthalmology on posterior vitreous detachment (PVD) and retinal breaks supports this MUC. The guideline states, "selected patients, particularly those with any degree of vitreous pigment, vitreous or retinal hemorrhage, or visible vitreoretinal traction, should be asked to return for a second examination promptly if they have new symptoms or within 6 weeks following the onset of PVD symptoms (AAO, 2019)." The developer notes the grading of evidence was strong for the initial exam recommendations but was discretionary for the follow-up examination recommendations (based on evidence presented in the 2014 AAO Preferred Practice Pattern Guidelines; evidence citations remained the same in the 2019 release).

Preliminary Analysis – MUC2022-122 Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder

Measure Description:

The percentage of individuals aged 18 and older with a mental and/or substance use disorder who demonstrated improvement or maintenance of functioning based on results from the 12-item World Health Organization Disability Assessment Schedule (WHODAS 2.0) or Sheehan Disability Index (SDS) 30 to 180 days after an index assessment.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: The measure under consideration (MUC) is a patient-reported outcome performance measure (PRO-PM), which addresses a high priority for the Merit-based Incentive Payment System (MIPS). It is also a behavioral health measure, which addresses a high priority specialty and high priority clinical condition for MIPS.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The developer notes that numerous studies have shown that patient functioning, among other outcomes, can be improved through implementation of measurement-based care (i.e., systematic assessment using standardized tools and use of feedback to inform clinical decision-making) and use of collaborative care models, or the integration of behavioral health and general medical services to provide evidence-based, goal-oriented treatment. The developer also cites several guidelines for certain mental health conditions which mention functional status as a component of the initial evaluation.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: Nineteen percent of U.S. adults (46.6 million individuals aged 18 and older) have a mental illness and 7.6 percent (18.7 million individuals aged 18 and older) have a substance use disorder (McCance-Katz, 2017). Individuals with mental disorders are more likely to report severe impairment in functioning compared to those with chronic medical conditions (Druss et al., 2008). Furthermore, the developed notes the level of functional impairment associated with mental or substance use disorders as well as reduction in impairment over time vary across gender and race/ethnicity (Moitra et al., 2014; Sheehan et al., 2015).

The developer performed alpha and beta testing from September 2019 to August 2021. Alpha testing involved six sites (and internal users); three utilized electronic medical records and the other three did not. A minimum of 10 patients was the inclusion threshold for providers and sites. Measure performance testing on 48 providers resulted in a mean performance score of 28.1 percent. The developer further examined the distribution of score change and length of time between index and follow-up assessments. Mean difference in score was 11.4 and mean length of time between

assessments was 62.9. The developer noted a significant disparity between the two data sources used. PsychPro data displayed a mean provider performance rate of 7.3 percent while the DSM-5 Field Trial data demonstrated a mean performance score of 36.2 percent.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The MUC is not duplicative of any existing measures in the program. Only one measure in MIPS, CMIT 01080-C-MIPS, relates to functional assessments for patients with a mental health disorder; however, it is for dementia patients exclusively. This MUC is a comprehensive measure inclusive of broad mental and/or substance use disorder and uses a measurement-based care framework for implementation across various settings and populations.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: The developer conducted alpha testing across six unique sites. Additionally, the developer gathered feedback on the use of the WHODAS 2.0, patients' and providers' response to the assessments and workflows, as well as the use of the PsychPRO registry. While 71 percent of alpha testing respondents reported that the ease of access to the functional assessment tools was "moderately to extremely easy" and 64 percent of respondents reported it was "moderately to extremely easy" for patients to complete the functional assessment tools, the developer identified significant challenges with feasibility. Challenges included the burden of manual clinical data entry for providers not using an electronic health record (EHR); the difficulty implementing measurement-based care in some clinicians' practices that may require considerable resource investment in time and training, and changes in workflow; and patients' low response rate to distributed assessments. Additionally, the developer noted that the separation of clinical and billing information in EHRs may result in limited comprehensive diagnostic data.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The MUC has not yet been submitted to a consensus-based entity (CBE) for endorsement, however, the steward attests that it is fully developed. The MUC is specified for individual clinicians. The developer completed signal-to-noise reliability testing using a beta-binomial model at the clinician level, with a sample size of 48 and a mean signal-to-noise reliability of 0.82.

The developer also completed a face validity assessment with a 12-member Consumer Family Panel (CFP) and a 19-member Technical Expert Panel (TEP). There were 10 CFP responses and 18 TEP responses (out of 12 and 19 total panelists, respectively). The CFP participants agreed that the measure is of very high importance (average rating of 92 percent) and usability (average rating of 91 percent). The TEP members agreed that the measure is of high importance (average rating of 85 percent) and moderate to high usability (average rating of 75 percent).

The developer also compared provider and site performance on the MUC to the provider and site-level performance on a conceptually related measure (Depression Remission at Six Months (National Quality Forum (NQF) #0711). Improved/maintained functioning rates are expected to be positively correlated

with depressive remission rates. The developer found the MUC is strongly positively correlated with the depression remission measure at the provider and site level (Spearman's Rho = 0.44). However, the developer noted caution is needed when interpreting these results given the preponderance of zero performance rates among providers and sites for both measures.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The MUC is not currently in use and the developer has not identified any potential negative unintended consequences of measure use.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

The measure under consideration (MUC) is a patient-reported outcome performance measure (PRO-PM), which addresses a high priority for the Merit-based Incentive Payment System (MIPS). It is also a behavioral health measure, which addresses a high priority specialty and high priority clinical condition for MIPS. The MUC is not duplicative of any existing measures in the program. Only one measure in MIPS, CMIT 01080-C-MIPS, relates to functional assessments for patients with a mental health disorder; however, it is for dementia patients exclusively. This MUC is a comprehensive measure inclusive of broad mental and/or substance use disorder and uses a measurement-based care framework for implementation across various settings and populations.

Summary: What is the potential impact of this measure on quality of care for patients?

Nineteen percent of U.S. adults (46.6 million individuals aged 18 and older) have a mental illness and 7.6

percent (18.7 million individuals aged 18 and older) have a substance use disorder (McCance-Katz, 2017). Individuals with mental disorders are more likely to report severe impairment in functioning compared to those with chronic medical conditions (Druss et al., 2008). Furthermore, the developed notes the level of functional impairment associated with mental or substance use disorders as well as reduction in impairment over time vary across gender and race/ethnicity (Moitra et al., 2014; Sheehan et al., 2015).

The developer notes that numerous studies have shown that patient functioning, among other outcomes, can be improved through implementation of measurement-based care (i.e., systematic assessment using standardized tools and use of feedback to inform clinical decision-making) and use of collaborative care models, or the integration of behavioral health and general medical services to provide evidence-based, goal-oriented treatment. The developer also cites several guidelines for certain mental health conditions which mention functional status as a component of the initial evaluation.

Preliminary Analysis – MUC2022-125 Gains in Patient Activation Measure (PAM) Scores at 12 Months (MIPS-Quality)

Measure Description:

The Patient Activation Measure (PAM) (Registered Trademark) is a 10- or 13- item questionnaire that assesses an individual's knowledge, skills, and confidence for managing their health and health care. The measure assesses individuals on a 0-100 scale that converts to one of four levels of activation, from low (1) to high (4). The PAM performance measure (PAM-PM) is the change in score on the PAM from baseline to follow-up measurement. A positive change would mean the patient is gaining in their ability to manage their health. The measure is not disease specific but has been successfully used with a wide variety of chronic conditions, as well as with people with no medical diagnosis.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This patient-reported outcome performance measure (PRO-PM) addresses Chronic Conditions and outcomes, both of which are high priority areas for future measure consideration for the Merit-based Incentive Payment System (MIPS). As a PRO-PM, it contributes to patient-centered care and focuses on the patient voice. The Patient Activation Measure (PAM) survey collects information directly from patients regarding their knowledge, skill, and confidence for managing their health and healthcare.

It is difficult to assess what rate from the measure numerator will be reported by clinicians participating in MIPS. The developer mentions several options: the aggregate of differences between Baseline PAM score and a second score (a continuous variable measure), the proportion of eligible patients who achieved a net increase in PAM score of at least 3 points in a 6-12 month period (passing), and the proportion of eligible patients who achieved a net increase in PAM score of at least 6 points in a 6-12 month period (excellent). Clarity around the measure rate is requested in order to fully understand how the measure would be implemented in MIPS.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: The developer's submission for the consensus-based entity (CBE) endorsement in 2015 describes the logic model for how provider interventions can improve this outcome. Specifically, assessing patient activation will drive targeted coaching and support by the clinical team, which in turn can increase patient activation and improve health outcomes. Overall, provider interventions that tailor support to the person's level of activation, build skills and confidence, use peer support, and change the social environment have a positive impact on this activation measure as well as other outcomes (Hibbard et a., 2013).

The developer shared in their 2015 submission to the CBE that over 240 articles have been published regarding the Patient Activation Measure (PAM). Of these studies, at least 85 percent show a statistically significant relationship between PAM scores and positive health actions, including getting preventive screening tests, immunizations, and health checkups. Many of these studies indicate that the higher the PAM score, the better health and clinical outcomes for the patient.

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: In the developer's submission for the CBE endorsement in 2015, they demonstrated a mean performance of 57.4 to 68.2 depending on the study. The standard deviation was 9.9 to 13.3 for U.S.-based studies. While these data demonstrate significant variation and gaps in care that are indicative of a quality challenge, it is unclear if this range of performance represents performance at the clinician or clinician-group level, and for which of the rates presented by the developer this represents.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: The measure is similar to two measures used in post-acute care/long-term care settings. These measures also estimate members' ability to self-manage their conditions and effectively participate in care activities. The measure under consideration (MUC) differs from the other measures in that it is broadly applicable to various patients with different diseases and needs. It is also proposed for different programs (MIPS and the End-Stage Renal Disease Quality Incentive Program (ESRD QIP)). The measure is in use in two Center for Medicare & Medicaid Innovation (CMMI) programs: Kidney Care Choices (2022) and Maternal Opioid Misuse (2021-2022).

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: The measure is fully developed and operationalized electronically. Data can be collected at the point of care (in-person), via IVR, through the patient portal, or via mail. Most electronic health records (EHRs) accommodate PAM data, when needed. PAM questions and scoring have been integrated into various electronic medical records (e.g., Epic, eClinicalWorks), and care management software (e.g., CaseTrakker, McKesson CCR/Vitals). As of September 2022, PAM scores have been collected from 71,790 people across 67 practices, which demonstrates that feasibility can be achieved by collecting baseline data and following up as necessary.

The developer states the survey instrument used to collect the data informing the proposed measure will be provided to the Centers for Medicare & Medicaid Services (CMS), and will be publicly available at no charge.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: The measure is fully developed, endorsed by a consensus-based entity (CBE) (National Quality Forum (NQF) #2483), and measure testing has demonstrated reliability and validity. The measure was scored for reliability using Cronbach's alpha analysis. This approach to reliability testing evaluates whether the PAM items (questions) all measure the same construct and if they do so across different subsamples of respondents. Cronbach's alpha for the PAM, across numerous populations, ranges from the high 0.8s to low 0.9s, indicating the PAM is reliable.

The developer also provided results from several types of validity testing. For example, in one published

paper, top-performing clinicians (i.e., those who evidenced the most change in patient PAM scores over time and higher PAM-PM scores) were more likely to use five key strategies that had been hypothesized based on expert consensus to increase patient activation. Bottom-performing clinicians reported using far fewer of these strategies, suggesting that PAM-PM is valid at the clinician level because measure scores can distinguish between clinicians who are more effectively promoting activation and their peers who are not.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: Feedback from end users has not identified any negative unintended consequences to patients or any unreasonable implementation issues that outweigh the benefits of the measure.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Support for Rulemaking

Summary: What is the potential value to the program measure set?

This measure addresses Chronic Conditions and outcomes, both of which are high priority areas for future measure consideration for the Merit-based Incentive Payment System (MIPS). It is a patient-reported outcome performance measure (PRO-PM). The Patient Activation Measure (PAM) survey collects information directly from patients regarding their knowledge, skill, and confidence for managing their health and healthcare. It is not disease specific and has been used with a wide variety of chronic conditions, as well as with people with no medical diagnosis. As a PRO-PM, it contributes to patient-centered care and focuses on the patient voice.

Summary: What is the potential impact of this measure on quality of care for patients?

This PRO-PM provides a standardized method for clinicians to assess patient activation through the

continuum of care. The developer's submission for the consensus-based entity (CBE) endorsement in 2015 highlighted the impact of targeted interventions on increased patient engagement, activation and improved outcomes. The PAM score (and changes in PAM scores) are predictive of health behavior, clinical outcomes, and costs, and can indicate the degree to which these interventions are occurring. The underlying assumption is patients that receive high-quality care, including interventions such as coaching and support, will increase their activation (ability to manage their disease), and improve their ability to self-manage over time. The measure is endorsed by the consensus-based entity (National Quality Forum (NQF) #2483). The developer states the survey instrument used to collect the data informing the proposed measure will be provided to the Centers for Medicare & Medicaid Services (CMS), and will be publicly available at no charge.

Preliminary Analysis – MUC2022-127 Initiation, Review, And/or Update to Suicide Safety Plan For Individuals With Suicidal Thoughts, Behavior, Or Suicide Risk

Measure Description:

This measure assesses the percentage of adult aged 18 and older with suicidal ideation or behavior symptoms (based on results of a standardized assessment tool) or increased suicide risk (based on the clinician's evaluation) for whom a suicide safety plan is initiated, reviewed, and/or updated in collaboration between the patient and their clinician.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure, which focuses on a process supporting the reduction of suicidal ideation, conceptually addresses behavioral health, a Meaningful Measures 2.0 domain and a high-priority area for the Merit-based Incentive Payment System (MIPS) program. According to the Centers for Disease Control and Prevention (CDC), suicide is a leading cause of death in the United States, and over 12 million people seriously think about committing suicide every year.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is a process measure, where initiating and reviewing a suicide safety plan with a patient at risk of suicide is a proxy for the clinical outcome of a reduction in suicides, suicide attempts, and suicidal ideation. The developer presents evidence that clinical interventions aimed at suicide prevention are effective in reducing suicidal behavior: one cohort comparison study found a suicide safety planning intervention with structured follow-up, administered in the emergency department to patients who presented with suicide-related concerns, led to a 45 percent reduction in suicidal behavior over six months, and a doubling of the likelihood of receiving outpatient mental health services (Stanley et al., 2018). The National Action Alliance for Suicide Prevention has presented a multilevel framework for clinical practices to reduce suicides, including evidence-based interventions, such as the safety plan intervention and follow-up and monitoring (Brodsky et al., 2018).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: Suicide is highly prevalent: according to the Centers for Disease Control and Prevention (CDC), suicide is a leading cause of death in the United States, and over 12 million people seriously think about committing suicide every year. Over the last two years, suicide rates are rising in nearly every state (CDC Vital Signs).

The developer reports that in a testing sample of 817 providers, 635 (77.7 percent) did not have enough patients to calculate a score. Of the remainder, 13.8 percent did not have statistically distinguishable performance from the mean score, 2.2 percent performed better than average, and 6.2 percent performed worse. These results indicate that there may be some limitation in the ability of the measure to distinguish quality across a wide range of clinical providers. However, in testing clinical sites,

differentiation improved: only 29.7 percent did not have enough patients to calculate a result, 48.4 percent were no different than the mean, 7.7 percent were better than the mean, and 14.3 percent were worse than the mean.

All eleven members of a technical expert panel convened by the developer agreed that the measure could be used to distinguish good from poor quality care.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: This is a new measure never reviewed by a consensus-based entity (CBE), or by the Measure Applications Partnership (MAP). There is one existing measure in the MIPS program that focuses on suicide: Adult Major Depressive Disorder (MDD): Suicide Risk Assessment (CMIT Ref No. 05813-E-MIPS). Though conceptually related, the measure under consideration (MUC) distinguishes itself by focusing on a care process that is directly designed to mitigate suicide risk, as opposed to screening for it. Taken together, these measures are complementary. There is another outcome measure in MIPS focused on a related mental health area: Depression Remission at Twelve Months (CMIT Ref No. 01741-C-MIPS, 05811-E-MIPS). The instrument used to assess remission, the PHQ-9, does include one question about self-harm. However, the existing MIPS measure is condition-specific to depression; the MUC would include patients with other behavioral health conditions who are at risk of suicide.

This MUC is concurrently submitted for pre-rulemaking in MIPS with MUC2022-131, Reduction in Suicidal Ideation or Behavior Symptoms, a clinical outcome measure that addresses a conceptually similar area. However, as these measures capture different quality constructs (i.e., a process and an outcome), both could be successfully implemented in the MIPS program.

Can the measure be feasibly reported?

Yes/No: No

Justification and Notes: The developer reports that all data elements are in defined fields in electronic sources. The developer reports a potential challenge of missing data in the electronic health record, but notes that in validity testing they were able to conclude that missing data had a negligible impact on performance on the measure.

The developer also reports that clinicians may need to make significant changes to their workflow in order to integrate routine use of assessment instruments. They hypothesize that as this type measurement-based care becomes more widely adopted as part of routine clinical practice, data collection difficulties are expected to become less of a barrier to implementation. During alpha testing with 33 clinicians, the developer reports participating sites did not find the use of the suicide safety plan to be easy. In addition, when answering open-ended questions, the clinicians thought that the suicide safety plan is useful for identifying triggers prompting suicidal ideation/behavior. They also mentioned that they will draw on their clinical judgment to determine whether it is necessary to complete an SSP for a given patient reporting suicidal ideation/behavior. These results indicate there could be some feasibility challenges with this measure.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This fully developed measure has not been reviewed by a consensus-based entity (CBE). However, the developer reported results of reliability and validity testing. The developer conducted a signal-to-noise reliability test, finding a mean score of 0.85, indicating strong reliability. The developer assessed the convergent validity of this measure and Adult Major Depressive Disorder (MDD): Suicide Risk Assessment (National Quality Forum (NQF) #0104), finding moderate correlations at both the provider (0.22 Spearman's r) and the site (0.34) levels. Finally, in a face validity assessment, all 11 members of a technical expert panel convened by the developer agreed that the measure could be used to distinguish good from poor quality care.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: No unintended consequences were identified by the developer.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure, which focuses on a process supporting the reduction of suicidal ideation, conceptually addresses behavioral health, a Meaningful Measures 2.0 domain and a high-priority area for the Merit-based Incentive Payment System (MIPS) program. According to the Centers for Disease Control and Prevention (CDC), suicide is a leading cause of death in the United States, and over 12 million people seriously think about committing suicide every year.

This is a new measure never reviewed by a consensus-based entity (CBE), or by the Measure

Applications Partnership (MAP). There is one existing measure in the MIPS program that focuses on suicide: Adult Major Depressive Disorder (MDD): Suicide Risk Assessment (CMIT Ref No. 05813-E-MIPS). Though conceptually related, this measure distinguishes itself by focusing on a care process that is directly designed to mitigate suicide risk, as opposed to screening for it. Taken together, these measures are complementary. There is another outcome measure in MIPS focused on a related mental health area: Depression Remission at Twelve Months (CMIT Ref No. 01741-C-MIPS, 05811-E-MIPS). The instrument used to assess remission, the PHQ-9, does include one question about self-harm. However, the existing MIPS measure is condition-specific to depression; the MUC would include patients with other behavioral health conditions who are at risk of suicide.

This MUC is concurrently submitted for pre-rulemaking in MIPS with MUC2022-131, Reduction in Suicidal Ideation or Behavior Symptoms, a clinical outcome measure that addresses a conceptually similar area. However, as these measures capture different quality constructs (i.e., a process and an outcome), both could be successfully implemented in the MIPS program.

Summary: What is the potential impact of this measure on quality of care for patients?

This is a process measure, where initiating and reviewing a suicide safety plan with a patient at risk of suicide is a proxy for the clinical outcome of a reduction in suicides, suicide attempts, and suicidal ideation. The developer presents evidence that clinical interventions aimed at suicide prevention are effective in reducing suicidal behavior: one cohort comparison study found a suicide safety planning intervention with structured follow-up, administered in the emergency department to patients who presented with suicide-related concerns, led to a 45 percent reduction in suicidal behavior over six months, and a doubling of the likelihood of receiving outpatient mental health services (Stanley et al., 2018). The National Action Alliance for Suicide Prevention has presented a multilevel framework for clinical practices to reduce suicides, including evidence-based interventions, such as the safety plan intervention and follow-up and monitoring (Brodsky et al., 2018).

The developer reports that in a testing sample of 817 providers, 635 (77.7 percent) did not have enough patients to calculate a score. Of the remainder, 13.8 percent did not have statistically distinguishable performance from the mean score, 2.2 percent performed better than average, and 6.2 percent performed worse. These results indicate that there may be some limitation in the ability of the measure to distinguish quality across a wide range of clinical providers. However, in testing clinical sites, differentiation improved: only 29.7 percent did not have enough patients to calculate a result, 48.4 percent were no different than the mean, 7.7 percent were better than the mean, and 14.3 percent were worse than the mean. All eleven members of a technical expert panel convened by the developer agreed that the measure could be used to distinguish good from poor quality care.

Preliminary Analysis – MUC2022-131 Reduction in Suicidal Ideation or Behavior Symptoms

Measure Description:

The percentage of individuals aged 18 and older with a mental and/or substance us disorder who demonstrated a reduction in suicidal ideation and/or behavior symptoms based on results from the Columbia-Suicide Severity Rating Scale 'Screen Version' or 'Since Last Visit' (CSSRS), within 120 days after an index assessment.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This measure, which focuses on the reduction of suicidal ideation, conceptually addresses behavioral health, a Meaningful Measures 2.0 domain and a high-priority area for the Merit-based Incentive Payment System (MIPS) program. According to the Centers for Disease Control and Prevention (CDC), suicide is a leading cause of death in the United States, and over 12 million people seriously think about committing suicide every year.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This is a clinical outcome measure, where reductions in suicidal ideation as measured would in turn lead to reductions in suicides and suicide attempts. The developer presents evidence that clinical interventions aimed at suicide prevention are effective in reducing suicidal behavior: one cohort comparison study found a suicide safety planning intervention with structured follow-up, administered in the emergency department to patients who presented with suicide-related concerns, led to a 45 percent reduction in suicidal behavior over six months, and a doubling of the likelihood of receiving outpatient mental health services (Stanley et al., 2018). The National Action Alliance for Suicide Prevention has presented a multilevel framework for clinical practices to reduce suicides, including evidence-based interventions, such as the safety plan intervention and follow-up and monitoring (Brodsky et al., 2018).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: Suicide is highly prevalent: according to the Centers for Disease Control and Prevention (CDC), suicide is a leading cause of death in the United States, and over 12 million people seriously think about committing suicide every year. Over the last two years, suicide rates are rising in nearly every state (CDC Vital Signs).

The developer reports that in a testing sample of 754 providers, 583 (77.3 percent) did not have enough patients to calculate a score. Of the remainder, 15.3 percent did not have statistically distinguishable performance from the mean score, 2.7 percent performed better than average, and 4.8 percent performed worse. These results indicate that there may be some limitation in the ability of the measure to distinguish quality across a wide range of clinical providers. However, in testing clinical sites, differentiation improved: only 30.7 percent did not have enough patients to calculate a result, 35.2

percent were no different than the mean, 11 percent were better than the mean, and 23.1 percent were worse than the mean.

Nine of ten members of a technical expert panel convened by the developer agreed that the measure could be used to distinguish good from poor quality care.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: This is a new measure never reviewed by a consensus-based entity (CBE), or by the Measure Applications Partnership (MAP). There is one existing measure in the MIPS program that focuses on suicide: Adult Major Depressive Disorder (MDD): Suicide Risk Assessment (CMIT Ref No. 05813-E-MIPS). Though conceptually related, this measure distinguishes itself by focusing on the relevant clinical outcome. There is another outcome measure in MIPS focused on a related mental health area: Depression Remission at Twelve Months (CMIT Ref No. 01741-C-MIPS, 05811-E-MIPS). The instrument used to assess remission, the PHQ-9, does include one question about self-harm. However, the existing MIPS measure is condition-specific to depression; the MUC would include patients with other behavioral health conditions who are at risk of suicide.

This MUC is concurrently submitted for pre-rulemaking in MIPS with MUC2022-127, Initiation, Review, And/Or Update To Suicide Safety Plan For Individuals With Suicidal Thoughts, Behavior, Or Suicide Risk, a process measure that addresses a conceptually similar area. However, as these measures capture different quality constructs (i.e., a process and an outcome), both could be successfully implemented in the MIPS program.

Can the measure be feasibly reported?

Yes/No: No

Justification and Notes: The developer reports that all data elements are in defined fields in electronic sources. The developer reports a potential challenge of missing data in the electronic health record, but notes that in validity testing they were able to conclude that missing data had a negligible impact on performance on the measure.

During alpha testing with 33 clinicians, the developer reports participating sites did not find the use of the assessment tool to be easy. In addition, when answering open-ended questions, the clinicians suggested that patients should have the option to complete the assessment at home. They agreed that the measure helps them in determining the frequency of suicidal ideation/behavior.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This fully developed measure has not been reviewed by a consensus-based entity. However, the developer reported results of reliability and validity testing. The developer conducted a signal-to-noise reliability test, finding a mean score of 0.77, indicating strong reliability. The developer assessed the convergent validity of MUC 2022-131 Reduction in Suicidal Ideation or Behavior Symptoms and Depression Remission at Six Months (NQF #0711), finding a weak correlation at the provider level but a strong (0.56 Spearman's r) at the site level. Finally, in a face validity assessment,

nine of ten members of a technical expert panel convened by the developer agreed that the measure could be used to distinguish good from poor quality care.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: No unintended consequences were identified by the developer.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support of this measure is conditional on endorsement of the measure by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

This measure, which focuses on the reduction of suicidal ideation, conceptually addresses behavioral health, a Meaningful Measures 2.0 domain and a high-priority area for the Merit-based Incentive Payment System (MIPS) program. According to the Centers for Disease Control and Prevention (CDC), suicide is a leading cause of death in the United States, and over 12 million people seriously think about committing suicide every year.

This is a new measure never reviewed by a consensus-based entity (CBE), or by the Measure Applications Partnership (MAP). There is one existing measure in the MIPS program that focuses on suicide: Adult Major Depressive Disorder (MDD): Suicide Risk Assessment (MIT Ref No. 05813-E-MIPS). Though conceptually related, this measure distinguishes itself by focusing on the relevant clinical outcome. There is another outcome measure in MIPS focused on a related mental health area: Depression Remission at Twelve Months (CMIT Ref No. 01741-C-MIPS, 05811-E-MIPS). The instrument used to assess remission, the PHQ-9, does include one question about self-harm. However, the existing MIPS measure is condition-specific to depression; the MUC would include patients with other behavioral

health conditions who are at risk of suicide. This MUC is concurrently submitted for pre-rulemaking in MIPS with MUC2022-127, Initiation, Review, And/Or Update To Suicide Safety Plan For Individuals With Suicidal Thoughts, Behavior, Or Suicide Risk, a process measure that addresses a conceptually similar area. However, as these measures capture different quality constructs (i.e., a process and an outcome), both could be successfully implemented in the MIPS program.

Summary: What is the potential impact of this measure on quality of care for patients?

This is a clinical outcome measure, where reductions in suicidal ideation as measured would in turn lead to reductions in suicides and suicide attempts. The developer presents evidence that clinical interventions aimed at suicide prevention are effective in reducing suicidal behavior: one cohort comparison study found a suicide safety planning intervention with structured follow-up, administered in the emergency department to patients who presented with suicide-related concerns, led to a 45 percent reduction in suicidal behavior over six months, and a doubling of the likelihood of receiving outpatient mental health services (Stanley et al., 2018). The National Action Alliance for Suicide Prevention has presented a multilevel framework for clinical practices to reduce suicides, including evidence-based interventions, such as the safety plan intervention and follow-up and monitoring (Brodsky et al., 2018).

The developer reports that in a testing sample of 754 providers, 583 (77.3 percent) did not have enough patients to calculate a score. Of the remainder, 15.3 percent did not have statistically distinguishable performance from the mean score, 2.7 percent performed better than average, and 4.8 percent performed worse. These results indicate that there may be some limitation in the ability of the measure to distinguish quality across a wide range of clinical providers. However, nine of ten members of a technical expert panel convened by the developer agreed that the measure could be used to distinguish good from poor quality care.

Part C & D Star Rating [Medicare]

Preliminary Analysis – MUC2022-043 Kidney Health Evaluation for Patients with Diabetes (KED) - Health Plans

Measure Description:

This measure assesses the percentage of members 18-85 years of age with diabetes (type 1 and type 2) who received a kidney health evaluation, defined by an estimated glomerular filtration rate (eGFR) and a urine albumin-creatinine ration (uACR), during the measurement year.

Does the measure address a critical quality objective not currently adequately addressed by the measures in the program set?

Yes/No: Yes

Justification and Notes: This Kidney Health Evaluation for Patients with Diabetes (KED) - Health Plans measure promotes the effective prevention and treatment of chronic conditions, a high priority area for the Part C & D Star Ratings program and addressing the Chronic Conditions priority of CMS' Meaningful Measures 2.0 initiative. This measure is specifically listed as a high-priority area for future measure consideration for the Part C Star Ratings Program. The intent of this measure is to improve rates of kidney health evaluation in patients with diabetes, a high-risk population.

Is the measure evidence-based and either strongly linked to outcomes or an outcome measure?

Yes/No: Yes

Justification and Notes: This process measure aligns with the clinical guidelines of the American Diabetes Association's (ADA) Standards of Medical Care in Diabetes. The guideline states that at least annually, urinary albumin (e.g., spot urinary albumin-to-creatinine ratio) and estimated glomerular filtration rate (eGFR) should be assessed in patients with type 1 diabetes with duration of ≥ 5 years and in all patients with type 2 diabetes regardless of treatment (American Diabetes Association, 2021). The guideline is listed as having a "B" level of evidence, meaning "Supportive evidence from well-conducted cohort studies". According to the ADA, both albuminuria and eGFR should be monitored annually to enable timely diagnosis of chronic kidney disease (CKD), monitor progression of CKD, detect superimposed kidney diseases including acute kidney injury (AKI), assess risk of CKD complications, dose drugs appropriately, and determine whether nephrology referral is needed (American Diabetes Association, 2021). The National Kidney Foundation and the Endocrine Society have published similar guidelines. Evidence from clinical trials indicate progressive kidney disease and cardiovascular outcomes can be slowed or prevented by treatment; thus, early detection and intervention for CKD is important (Skolnik & Style, 2021).

Does the measure address a quality challenge?

Yes/No: Yes

Justification and Notes: Up to 40 percent of those individuals with diabetes also develop CKD, which increases their risk of developing cardiovascular disease (Skolnik & Style, 2021). The presence of CKD can also progress to end-stage renal disease (ESRD) which requires dialysis or kidney transplantation. (American Diabetes Association, 2021). The developer indicated for this measure that higher scores indicate better performance. The measure has a mean performance score of 40.0 percent demonstrating an opportunity for improvement broadly. Average plan performance was lowest among the 18-64 age group and varied across geographic regions, highlighting differences in measure

performance and evidence of performance gaps. The measure developer indicated that the estimated annual denominator size across accountable entities eligible to report this measure is 5,000,000 and that fewer than 50% of adults with diabetes receive annual kidney health evaluation (Saran et al., 2020). This measure can address a gap in care for the surveillance of CKD among diabetic patients enrolled in Medicare Advantage plans.

Does the measure contribute to efficient use of measurement resources and/or support alignment of measurement across programs?

Yes/No: Yes

Justification and Notes: There are 40 measures (38 unique measures) in Part C & D Star Ratings program, 12 of which address the meaningful measure area of chronic conditions. This proposed measure is similar to 08075-E-MIPS Kidney Health Evaluation which was specified, proposed, and finalized for the MIPS program. In 2021, MAP evaluated and conditionally supported 08075-E-MIPS Kidney Health Evaluation. This new proposed measure MUC 2022-043 Kidney Health Evaluation for Patients with Diabetes (KED) - Health Plans has similar specifications as 08075-E-MIPS Kidney Health Evaluation, however this proposed measure is specified for the health plan level.

Related, the program currently has a Diabetes Care Kidney Disease Monitoring measure in the program (CMIT: 04021-C-PARTC). That measure assesses the percentage of patients with diabetes who had a kidney function test during the measurement year. NCQA proposed retiring Diabetes Care Kidney Disease Monitoring from the HEDIS measure set due to the measure not being precise enough to meet the needs of kidney health evaluations. NCQA proposes the inclusion of this measure Kidney Health Evaluation for Patients with Diabetes (KED) - Health Plans in this program, as this measure provides a more accurate evaluation of kidney health by measuring both eGFR and uACR rates.

Can the measure be feasibly reported?

Yes/No: Yes

Justification and Notes: The measure developer notes that institutional, professional, and pharmacy claims data are used to identify members with diabetes (denominator). Procedural codes on claims are used to identify members receiving both the eGFR and uACR services (numerator) for this measure. Thus, all data elements are in defined fields in electronic sources.

Is the measure applicable to and appropriately specified for the program's intended care setting(s), level(s) of analysis, and population(s)?

Yes/No: Yes

Justification and Notes: This measure has not been submitted to a consensus-based entity (CBE). The fully developed measure was tested in the ambulatory/office-based care setting and the level of analysis is at the health plan level. A signal-to-noise reliability test provided by the measure developer using a beta-binomial model yielded a reliability result of 0.995, indicating strong reliability. Construct validity testing indicated a mild to moderate correlation with other similar measures of monitoring and treatment for patient with diabetes, including the correlations of 0.315 with Hemoglobin A1c Control, -0.296 with Hemoglobin A1c Poor Control, and 0.201 with Eye Exam (Retinal) Performed. The measure developer notes that results are not unexpected due to the measure being new and developed to address a gap in care. The measure was additionally assessed for face validity and 17 members on NCQA's Committee on Performance Measurement unanimously voted to approve the measure for public reporting based on first year results of the measure.

If the measure is in current use, have negative unintended issues to the patient been identified? Have implementation challenges outweighing the benefits of the measure been identified?

Yes/No: No

Justification and Notes: The measure developer noted no unintended consequences and providers from multiple panels did not indicate significant challenges for implementation.

PAC/LTC Core Concept?

Yes/No: N/A

Impact Act Domain

Yes/No: N/A

Hospice High Priority Areas

Yes/No: N/A

MAP Rural Health Advisory Group Input:

Votes: [Not yet available.]

MAP Health Equity Advisory Group Input:

Votes: [Not yet available.]

Recommendation

Preliminary Analysis Recommendation:

Conditional Support for Rulemaking

Support for this measure is conditional on endorsement by a consensus-based entity (CBE).

Summary: What is the potential value to the program measure set?

The newly developed measure addresses the Meaningful Measures 2.0 Chronic Conditions priority, and the measure itself is specifically cited as a high priority for the Part C & D Star Ratings program. If this measure is implemented into the proposed program, the measure will focus on the appropriate identification, monitoring, and treatment of chronic kidney disease (CKD) among Medicare Advantage members with diabetes. This measure is aligned with current American Diabetes Association clinical guidelines and addresses a high-risk population. In 2021, a similar measure was proposed for the MIPS program and conditionally supported by the Measure Applications Partnership.

Summary: What is the potential impact of this measure on quality of care for patients?

The National Kidney Foundation indicated that fewer than 50 percent of adults with diabetes receive annual kidney health evaluation; thus, this measure has the potential to narrow the gap of care among Medicare Advantage members with diabetes. In testing, the measure was found to be both reliable and valid. Screening for CKD is recommended per guidelines set by the American Diabetes Association, the Endocrine Society, and the National Kidney Foundation, and enables early CKD diagnosis, counseling, pharmacologic intervention, and possibly referral to a nephrologist (Skolnik & Style, 2021). Kidney Health Evaluation for Patients with Diabetes (KED) - Health Plans has the potential to prevent CKD among a population at increased risk.