NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Applications Partnership (MAP) Clinician: 2022-2023 Measures Under Consideration (MUC) Cycle Measure Specifications

MANUAL

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

MUC2022-097 Low Back Pain

Program

Merit-based Incentive Payment System-Cost

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all Low Back Pain episodes attributed to a clinician, where each ratio is weighted by the number of days in each episode assigned to a clinician. This ratio is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.

Numerator Exclusions

N/A

Denominator

The measure denominator is the total number of days from Low Back Pain episodes assigned to the clinician across all patients.

Denominator Exclusions

The following populations are excluded from the measure to ensure data completeness:

- Patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the episode window.
- Patient was not enrolled in Medicare Parts A/B for the entirety of the 120-day lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.
- Patient was not found in the Medicare Enrollment Database.
- Death date of the patient occurred before the episode end date.
- Patient has an episode window shorter than 120 days.

Exclusions specific to the Low Back Pain measure are developed with input from the Low Back Pain Clinician Expert Workgroup. They include cauda equina syndrome, osteoporotic compression fracture, spinal infection, spinal neoplasms, myelopathy, and trauma. Additionally, episodes are excluded if a spinal surgery occurs within 60 days of the initial trigger service to minimize the risk of capturing episodes that are purely consultative or pre-operative.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Chiropractic medicine, Physical Therapy, Internal Medicine

Measure Type

Cost/ReSource: Use

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory surgery center;Ambulatory/office-based care;Hospital outpatient department (HOD);Hospital inpatient acute care facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Affordability and Efficiency

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

Section 101(f) of MACRA

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Claims

Stratification

No

Feasibility of Data Elements

data elements are in defined fields in electronic

Feasibility Assessment

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.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Analysis of all TIN-NPIs with at least 20 attributed episodes for the 2019 performance period shows a large range of provider scores on Low Back Pain measure. The measure score has the following distributional characteristics:

Mean: $1,712, standard deviation: $518

Median: $1,640

Min: $395, max: $10,179

Interquartile range is $617

Coefficient of variation: 0.30

The score decile distribution for the 2019 performance period is:

10th: $1,146

20th: $1,300

30th: $1,420

40th: $1,531

50th: $1,640

60th: $1,758

70th: $1,896

80th: $2,074

90th: $2,349

Unintended Consequences

No unintended consequences to individuals or populations have been identified during testing.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence o r level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

Please see evidence attachment.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

538,186 total clinicians participated in MIPS as either a clinician group or individuals in 2019. We estimate a similar number of clinician groups and individual clinicians will continue to participate in MIPS.

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Other (enter here):: Medicare Part D enrollment status, provider specialty

Patient-level demographics: please select all that apply:

Age

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Health behaviors/health choices

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

To determine whether the risk model adequately accounts for confounding factors, we assessed two factors: discrimination and calibration. In this case, discrimination is the ability to explain the variance in cost of individual episodes. The amount of variance explained is estimated by the R-squared metric with the range between 0 and 1. The R-square for the measure is 0.532, and 0.531 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 0.1% of the variation in the actual observed cost of episodes is explained by the risk adjustment model and sub-group stratification. The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is not essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians. Consequently, results should also be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted. Calibration evaluates the consistency of the measure in estimating episode cost across the full range of reSource: use patterns in the population. It is estimated by the average predictive ratios across groups within the population. We calculated the predictive ratio using the formula of average expected cost / average observed cost for all episodes in each decile. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. Below is the predictive ratio by decile of predicted episode cost: Decile 1: 0.95 Decile 2: 1.03 Decile 3: 1.03 Decile 4: 0.98 Decile 5: 0.98 Decile 6: 0.99 Decile 7: 1.00 Decile 8: 1.01 Decile 9: 1.01 Decile 10: 1.00 This demonstrates that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, ranging between 0.95 and 1.03. Overall, the risk adjustment model does not over- or under-predict cost across the full range of resource use patterns in the population.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

To assess the impact on cost this measure may have, we examine the share of Medicare Parts A&B spending that this measure covers during the performance year 2019, assuming a volume threshold of 20 episodes. At the TIN-NPI reporting level, this measure covers 1.55% of Medicare Parts A&B spending. This measure can help to encourage more cost efficient care related to Low Back Pain chronic care.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Weighted Heteroscedastic Within-Group Variance Estimation

Signal-to-Noise: Sample size

46,326

Signal-to-Noise: Statistical result

0.761

Signal-to-Noise: Interpretation of results

We conducted reliability testing of measures for clinicians (TIN-NPIs), constructed using episodes ending between January 1, 2019 and December 31, 2019. Reliability refers to the extent to which a measure reflects true variation between risk-adjusted episode spending of clinicians, as opposed to random variation. The reliability metric specifically captures how much of the variance in a measure is due to systematic differences in episode spending between clinicians, rather than differences in episode spending within a clinicians set of episodes. A measure with high reliability suggests that comparisons of performance across clinicians can be expected to better reflect systematic differences in actual performance. In the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171), CMS identified reliability levels between 0.4 to 0.7 as moderate and reliability levels above 0.7 as high. In the CY 2017 Quality Payment Program final rule, CMS also identified a threshold of 0.4 for mean reliability to be applied for measures and this was reiterated as the threshold in the CY 2022 Physician Fee Schedule Final Rule (86 FR 64996). Our testing results indicate that this measure has high reliability for clinicians across a range of volume thresholds.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Validity is tested empirically by examining the association between the measure score and high-cost events that drive the measure score, such as downstream complications and consequences of care.

Empiric Validity: Sample size

46,326

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

All providers:

All Mean score: $1,712 (SD $518)

Acute Inpatient Stay:

0% Mean score: $1,653 (SD $526)

Q1 (0.1-1.5%) Mean Score: $1,695 (SD $461)

Q2 (1.5-2.5%) Mean Score: $1,764 (SD $475)

Q3 (2.5-3.9%) Mean Score: $1,819 (SD $490)

Q4 (3.9-47.6%) Mean Score: $1,931 (SD $493)

Post-Acute Care:

0% Mean score: $1,586 (SD $498)

Q1 (0.2-2.4%) Mean Score: $1,635 (SD $450)

Q2 (2.4-4.2%) Mean Score: $1,723 (SD $491)

Q3 (4.2-7.9%) Mean Score: $1,780 (SD $468)

Q4 (7.9-96.9%) Mean Score: $1,979 (SD $574)

Imaging:

0% Mean score: $1,360 (SD $580)

Q1 (2.2-23.6%) Mean Score: $1,514 (SD $480)

Q2 (23.6-32.1%) Mean Score: $1,627 (SD $464)

Q3 (32.1-47.5%) Mean Score: $1,768 (SD $520)

Q4 (47.5-100%) Mean Score: $1,934 (SD $502)

Spine Injections:

0% Mean score: $1,493 (SD $557)

Q1 (0.5-5.9 %) Mean Score: $1,564 (SD $505)

Q2 (5.9-10.6%) Mean Score: $1,639 (SD $488)

Q3 (10.6-22.2%) Mean Score: $1,752 (SD $492)

Q4 (22.3-100%) Mean Score: $1,943 (SD $477)

Empiric Validity: Methods and findings: This analysis examines the cost pattern when there is a concurrent high cost. High cost events are expensive services/facility stays that are grouped to the episode through service assignment or auto-grouping, such as acute inpatient and post-acute care. These metrics show a specific subset of potentially high-costs events that could influence performance on the Low Back Pain cost measure, as well as the mean score and standard deviation associated with how frequently each type of high-cost events occurs with a 20-episode volume threshold applied. We would expect to see that providers who have more instances of high-cost events have higher mean scores. Across all events shown, providers with low frequencies of high cost events had minimal or no increase in mean score compared to the mean score for all providers. As expected, higher frequencies of high-costs events are associated with higher scores. The increases in mean scores are most pronounced among TIN-NPIs with higher frequencies of imaging.

Empiric Validity: Interpretation of results

Yes

Face Validity

No

Face Validity: Number of voting experts and patients/caregivers

N/A

Face Validity: Result

N/A

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Other: The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians.

Measure Performance Score Interpretation

Lower score is better

Mean performance score

1,712

Median performance score

1,640

Minimum performance score

395

Maximum performance score

10,179

Standard deviation of performance scores

518

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare & Medicaid Services

Measure Steward Contact Information

Donta Henson
Center for Clinical Standards and Quality
7500 Security Boulevard
Baltimore, MD 21244
Donta.Henson1@cms.hhs.gov
(410-786-1947

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Joyce Lam
Acumen, LLC
500 Airport Blvd. Suite 100
Burlingame, CA 94010
ccsq-macra-support@acumenllc.com
(650) 558-8882

Secondary Submitter Contact Information

Anastasiia Biriuchinskaia
Acumen, LLC
500 Airport Blvd. Suite 100
Burlingame, CA 94010
ccsq-macra-support@acumenllc.com
(650) 558-8882

Submitter Comments

Please note that we selected "other" to describe the type of measure performance score that the measure uses, but the description was not visible. The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians. Please also note that since we were not able to enter the empiric validity results in the "Empiric Validity: Statistical result" field, we added those results in the "Empiric Validity: Methods and findings" field, before summarizing the empiric validity methods and findings. Please let us know if you have any questions.

MUC2022-100 Emergency Medicine

Program

Merit-based Incentive Payment System-Cost

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

The cost measure numerator is the sum of the ratio of observed to expected payment-standardized cost to Medicare for all Emergency Medicine episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.

Numerator Exclusions

N/A

Denominator

The cost measure denominator is the total number of episodes from the Emergency Medicine episode group attributed to a clinician.

Denominator Exclusions

The following populations are excluded from the measure to ensure data completeness:

Patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the episode window.

Patient was not enrolled in Medicare Parts A/B for the entirety of the 120-day lookback period plus 14-day episode window, or was enrolled in Part C for any part of the lookback plus episode window.

Patient's date of birth is missing

Patient's death date occurred before the episode end date.

Exclusions specific to the Emergency Medicine measure are developed with input from the Emergency Medicine Clinician Expert Workgroup.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Emergency medicine

Measure Type

Cost/ReSource: Use

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Emergency department

Multiple Scores

No

What one healthcare domain applies to this measure?

Affordability and Efficiency

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

Section 101(f) of MACRA

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Claims

Stratification

No

Feasibility of Data Elements

data elements are in defined fields in electronic

Feasibility Assessment

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.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Analysis of all clinicians with at least 20 attributed episodes for the 2019 performance period shows a large range of provider scores on Emergency Medicine measure. The measure score has the following distributional characteristics:

Mean: $5,058, standard deviation: $831

Median: $5,101

Min: $1,135, max: $84,903

Interquartile range is $837

Coefficient of variation: 0.16

The score decile distribution for the 2019 performance period is:

10th: $4,101

20th: $4,511

30th: $4,760

40th: $4,945

50th: $5,101

60th: $5,247

70th: $5,397

80th: $5,578

90th: $5,868

Unintended Consequences

No unintended consequences to individuals or populations have been identified during testing.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

Please see evidence attachment.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

538,186 total clinicians participated in MIPS as either a clinician group or individuals in 2019. We estimate a similar number of clinician groups and individual clinicians will continue to participate in MIPS.

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Proxy social risk factors

Patient-level demographics: please select all that apply:

Age

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Health behaviors/health choices

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

Dual Eligibility for Medicare and Medicaid

Patient community characteristic: please select all that apply:

N/A

Risk model performance

To determine whether the risk model adequately accounts for confounding factors, we assessed two factors: discrimination and calibration. In this case, discrimination is the ability to explain the variance in

cost of individual episodes. The amount of var iance explained is estimated by the R-squared metric with the range between 0 and 1. The R-square for the measure is 0.607, and 0.607 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 60.7% of the variation in the actual observed cost of episodes is explained by the risk adjustment model and sub-group stratification. The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is not essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians. Consequently, results should also be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted. Calibration evaluates the consistency of the measure in estimating episode cost across the full range of reSource: use patterns in the population. It is estimated by the average predictive ratios across groups within the population. We calculated the predictive ratio using the formula of average expected cost / average observed cost for all episodes in each decile. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. Below is the predictive ratio by decile of predicted episode cost: Decile 1: 1.01 Decile 2: 1.01 Decile 3: 1.00 Decile 4: 0.99 Decile 5: 1.00 Decile 6: 0.99 Decile 7:1.00 Decile 8: 1.00 Decile 9: 1.00 Decile 10: 1.00 This demonstrates that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, ranging between 0.99 and 1.01. Overall, the risk adjustment model does not over- or under-predict cost across the full range of resource use patterns in the population.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

To assess the impact on cost this measure may have, we examine the share of Medicare Parts A&B spending that this measure covers during the performance year 2019, assuming a volume threshold of 20 episodes. At the TIN-NPI reporting level, this measure covers 17.38% of Medicare Parts A&B spending. This measure can help to encourage more cost efficient care related to emergency medicine care.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Homoscedastic Mean of Individual Variance Estimation

Signal-to-Noise: Sample size

79,787

Signal-to-Noise: Statistical result

0.836

Signal-to-Noise: Interpretation of results

We conducted reliability testing of measures for clinicians (TIN-NPIs), constructed using episodes ending between January 1, 2019 and December 31, 2019. Reliability refers to the extent to which a measure reflects true variation between risk-adjusted episode spending of clinicians , as opposed to random variation. The reliability metric specifically captures how much of the variance in a measure is due to systematic differences in episode spending between clinicians, rather than differences in episode spending within the set of episodes of a clinician. A measure with high reliability suggests that comparisons of performance across clinicians can be expected to better reflect systematic differences in actual performance. In the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171), CMS identified reliability levels between 0.4 to 0.7 as moderate and reliability levels above 0.7 as high. In the CY 2017 Quality Payment Program final rule, CMS also identified a threshold of 0.4 for mean reliability to be applied for measures and this was reiterated as the threshold in the CY 2022 Physician Fee Schedule Final Rule (86 FR 64996). Our testing results indicate that this measure has high reliability for clinicians across a range of volume thresholds.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Validity is tested empirically by examining the association between the measure score and high-cost events that drive the measure score, such as downstream complications and consequences of care.

Empiric Validity: Sample size

79,787

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

All episodes:

Mean risk-adjusted cost: $5,012 (SD $4,843)

Median: $4,316

Emergency Department Revisit

Mean risk-adjusted cost: $8,498 (SD $7,622)

Median: $6,025

Inpatient Stay Readmission

Mean risk-adjusted cost: $17,541 (SD $11,184)

Median: $14,039

This analysis examines the cost pattern when there is a concurrent high cost event to demonstrate validity. High cost events are expensive services/facility stays that are grouped to the episode through service assignment or auto-grouping, such as emergency department revisits. These metrics show a specific subset of potentially high-costs events that could influence performance on the Emergency Medicine cost measure. The results show that episodes with an emergency department re-visit or inpatient readmissions after fully discharged from the initial visit have higher mean scores than the overall population in the measure. The cost measure is thus able to differentiate the cost efficiency of episodes based on high-cost events.

Empiric Validity: Interpretation of results

Yes

Face Validity

No

Face Validity: Number of voting experts and patients/caregivers

N/A

Face Validity: Result

N/A

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Other: The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians.

Measure Performance Score Interpretation

Lower score is better

Mean performance score

5,058

Median performance score

5,101

Minimum performance score

1,135

Maximum performance score

84,903

Standard deviation of performance scores

831

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare & Medicaid Services

Measure Steward Contact Information

Donta Henson

Center for Clinical Standards and Quality

7500 Security Boulevard

Baltimore, MD 21244

Donta.Henson1@cms.hhs.gov

410-786-1947

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Joyce Lam

Acumen, LLC

500 Airport Blvd. Suite 100

Burlingame, CA 94010

ccsq-macra-support@acumenllc.com

650-558-8882

Secondary Submitter Contact Information

Anastasiia Biriuchinskaia

Acumen, LLC

500 Airport Blvd. Suite 100

Burlingame, CA 94010

ccsq-macra-support@acumenllc.com

650-558-8882

Submitter Comments

Please note that we selected "other" to describe the type of measure performance score that the measure uses, but the description was not visible. The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians. Please also note that since we were not able to enter the empiric validity results in the "Empiric Validity: Statistical result" field, we added those results in the "Empiric Validity: Methods and findings" field, before summarizing the empiric validity methods and findings. Please let us know if you have any questions.

MUC2022-101 Depression

Program

Merit-based Incentive Payment System-Cost

Section 1: Measure Information

Measure Specifications and Endorsement Status

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 clinicians role in managing care during a Depression episode.

Numerator

The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all Depression episodes attributed to a clinician, where each ratio is weighted by each episodes number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.

Numerator Exclusions

N/A

Denominator

The measure denominator is the total number of days from Depression episodes assigned to the clinician across all patients.

Denominator Exclusions

The following populations are excluded from the measure to ensure data completeness:

Patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the episode window.

Patient was not enrolled in Medicare Parts A/B for the entirety of the 120-day lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.

Patient was not found in the Medicare Enrollment Database.

Patients death date occurred before the episode end date.

Patient has an episode window shorter than one year.

Exclusions specific to the Depression measure are developed with input from the Depression Clinician Expert Workgroup and include bipolar disorder, schizophrenia, and drug or alcohol psychosis.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Internal medicine, Family Practice, Psychiatry

Measure Type

Cost/Resource: Use

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care;Hospital outpatient department (HOD);Hospital inpatient acute care facility;Nursing home;Skilled nursing facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Affordability and Efficiency

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

Section 101(f) of MACRA

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Claims

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources

Feasibility Assessment

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.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Analysis of all TIN-NPIs with at least 20 attributed episodes for the 2019 performance period shows a large range of provider scores on the Depression measure. The measure score has the following distributional characteristics:

Mean: $1,429, standard deviation: $539

Median: $1,333

Min: $231, max: $8,212

Interquartile range is $575

Coefficient of variation: 0.38

The score decile distribution for the 2019 performance period is:

10th: $897

20th: $1,027

30th: $1,134

40th: $1,231

50th: $1,333

60th: $1,445

70th: $1,573

80th: $1,744

90th: $2,038

Unintended Consequences

No unintended consequences to individuals or populations have been identified during testing.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

Please see evidence attachment.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

538,186 total clinicians participated in MIPS as either a clinician group or individuals in 2019. We estimate a similar number of clinician groups and individual clinicians will continue to participate in MIPS.

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Proxy social risk factors;Other (enter here):: Medicare Part D enrollment status; provider specialty

Patient-level demographics: please select all that apply:

Age

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Health behaviors/health choices

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

Dual Eligibility for Medicare and Medicaid

Patient community characteristic: please select all that apply:

N/A

Risk model performance

To determine whether the risk model adequately accounts for confounding factors, we assessed two factors: discrimination and calibration. In this case, discrimination is the ability to explain the variance in cost of individual episodes. The amount of variance explained is estimated by the R-squared metric with the range between 0 and 1. The R-square for the measure is 0.18, and 0.18 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 18% of the variation in the actual observed cost of episodes is explained by the risk adjustment model and sub-group stratification.

The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is not essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians. Consequently, results should also be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted.

Calibration evaluates the consistency of the measure in estimating episode cost across the full range of reSource: use patterns in the population. It is estimated by the average predictive ratios across groups within the population. We calculated the predictive ratio using the formula of average expected cost / average observed cost for all episodes in each decile. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. Below is the predictive ratio by decile of predicted episode cost:

Decile 1: 0.95

Decile 2: 0.99

Decile 3: 0.99

Decile 4: 0.99

Decile 5: 1.01

Decile 6: 1.00

Decile 7: 1.03

Decile 8: 1.02

Decile 9: 1.00

Decile 10: 0.99

This demonstrates that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, ranging between 0.95 and 1.03. Overall, the risk adjustment model does not over- or under-predict cost across the full range of reSource: use patterns in the population.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

To assess the impact on cost this measure may have, we examine the share of Medicare Parts A&B spending that this measure covers during the performance year 2019, assuming a volume threshold of 20 episodes. At the TIN-NPI reporting level, this measure covers 0.21% of Medicare Parts A&B spending. This measure can help to encourage more cost efficient care related to Depression.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Weighted Heteroscedastic Within-Group Variance Estimation

Signal-to-Noise: Sample size

21,802

Signal-to-Noise: Statistical result

0.835

Signal-to-Noise: Interpretation of results

We conducted reliability testing of measures for clinicians (TIN-NPIs), constructed using episodes ending between January 1, 2019 and December 31, 2019. Reliability refers to the extent to which a measure reflects true variation between risk-adjusted episode spending of clinicians, as opposed to random variation. The reliability metric specifically captures how much of the variance in a measure is due to systematic differences in episode spending between clinicians, rather than differences in episode spending within the set of episodes of the clinician. A measure with high reliability suggests that comparisons of performance across clinicians can be expected to better reflect systematic differences in actual performance. In the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171), CMS identified reliability levels between 0.4 to 0.7 as moderate and reliability levels above 0.7 as high. In the CY 2017 Quality Payment Program final rule, CMS also identified a threshold of 0.4 for mean reliability to be applied for measures and this was reiterated as the threshold in the CY 2022 Physician Fee Schedule Final Rule (86 FR 64996). Our testing results indicate that this measure has high reliability for clinicians across a range of volume thresholds.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Validity is tested empirically by examining the association between the measure score and high-cost events that drive the measure score, such as downstream complications and consequences of care.

Empiric Validity: Sample size

21,802

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

All providers:

All Mean score: $1,429 (SD $539)

Acute Inpatient Stay:

0% Mean score: $1,413 (SD $539)

Q1 (0.1-1.9%) Mean Score: $1,423 (SD $426)

Q2 (1.9-2.9%) Mean Score: $1,563 (SD $528)

Q3 (2.9-4.0%) Mean Score: $1,608 (SD $464)

Q4 (4.1 -18.2%) Mean Score: $1,696 (SD $614)

Emergency Department:

0% Mean Score: $1,271 (SD $567)

Q1 (0.4-4.6%) Mean Score: $1,354 (SD $538)

Q2 (4.7-8.2%) Mean Score: $1,417 (SD $515)

Q3 (8.2-12.5%) Mean Score: $1,462 (SD $495)

Q4 (12.6-53.1%) Mean Score: $1,596 (SD $540)

Empiric Validity: Methods and findings: This analysis examines the cost pattern when there is a concurrent high cost event to demonstrate validity. High cost events are expensive services/facility stays that are grouped to the episode through service assignment or auto-grouping, such as acute inpatient and post-acute care. These metrics show a specific subset of potentially high-cost events that could influence performance on the Depression cost measure, as well as the mean score and standard deviation associated with how frequently each type of high-cost event occurs with a 20-episode volume threshold applied. We would expect to see that providers who have more high-cost events have higher mean scores. Across all events shown, providers with low frequencies of high cost events had minimal or no increase in mean score compared to the mean score for all providers. As expected, higher frequencies of high-costs events are associated with higher scores. The increases in mean scores are most pronounced among TIN-NPIs with higher frequencies of acute inpatient stays.

Empiric Validity: Interpretation of results

Yes

Face Validity

No

Face Validity: Number of voting experts and patients/caregivers

N/A

Face Validity: Result

N/A

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Other: The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians.

Measure Performance Score Interpretation

Lower score is better

Mean performance score

1,429

Median performance score

1,333

Minimum performance score

231

Maximum performance score

8,212

Standard deviation of performance scores

539

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare & Medicaid Services

Measure Steward Contact Information

Donta Henson

Center for Clinical Standards and Quality

7500 Security Boulevard

Baltimore, MD 21244

Donta.Henson1@cms.hhs.gov

410-786-1947

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Joyce Lam

Acumen, LLC

500 Airport Blvd. Suite 100

Burlingame, CA 94010

ccsq-macra-support@acumenllc.com

650-558-8882

Secondary Submitter Contact Information

Anastasiia Biriuchinskaia

Acumen, LLC

500 Airport Blvd. Suite 100

Burlingame, CA 94010

ccsq-macra-support@acumenllc.com

650-558-8882

Submitter Comments

Please note that we selected "other" to describe the type of measure performance score that the measure uses, but the description was not visible. The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians. Please also note that since we were not able to enter the empiric validity results in the "Empiric Validity: Statistical result" field, we added those results in the "Empiric Validity: Methods and findings" field, before summarizing the empiric validity methods and findings. Please let us know if you have any questions.

MUC2022-106 Heart Failure

Program

Merit-based Incentive Payment System-Cost

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all Heart Failure episodes attributed to a clinician, where each ratio is weighted by each episodes number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.

Numerator Exclusions

N/A

Denominator

The measure denominator is the total number of days from Heart Failure episodes assigned to the clinician across all patients.

Denominator Exclusions

The following populations are excluded from the measure to ensure data completeness:

Patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the episode window.

Patient was not enrolled in Medicare Parts A/B for the entirety of the 120-day lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.

Patient was not found in the Medicare Enrollment Database.

The death date of the patient occurred before the episode end date.

Patient has an episode window shorter than one year.

Exclusions specific to the Heart Failure measure are developed with input from the Heart Failure Clinician Expert Workgroup.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Cardiovascular disease (cardiology), Internal Medicine, Family Practice

Measure Type

Cost/Resource: Use

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care;Hospital outpatient department (HOD);Skilled nursing facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Affordability and Efficiency

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

Section 101(f) of MACRA

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Claims

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

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.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Analysis of all clinicians with at least 20 attributed episodes for the 2019 performance period shows a large range of provider scores on Heart Failure measure. The measure score has the following distributional characteristics:

Mean: $12,118, standard deviation: $3,510

Median: $11,711

Min: $2,310, max: $37,010

Interquartile range is $4,381

Coefficient of variation: 0.29

The score decile distribution for the 2019 performance period is:

10th: $8,063

20th: $9,241

30th: $10,147

40th: $10,959

50th: $11,711

60th: $12,568

70th: $13,536

80th: $14,750

90th: $16,590

Unintended Consequences

No unintended consequences to individuals or populations have been identified during testing.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

Please see evidence attachment.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

538,186 total clinicians participated in MIPS as either a clinician group or individuals in 2019. We estimate a similar number of clinician groups and individual clinicians will continue to participate in MIPS.

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Proxy social risk factors;Other (enter here):: Medicare Part D enrollment status, provider specialty

Patient-level demographics: please select all that apply:

Age

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Health behaviors/health choices

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

Dual Eligibility for Medicare and Medicaid

Patient community characteristic: please select all that apply:

N/A

Risk model performance

To determine whether the risk model adequately accounts for confounding factors, we assessed two factors: discrimination and calibration. In this case, discrimination is the ability to explain the variance in cost of individual episodes. The amount of variance explained is estimated by the R-squared metric with the range between 0 and 1. The R-square for the measure is 0.132, and 0.131 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 13.1% of the variation in the actual observed cost of episodes is explained by the risk adjustment model and sub-group stratification.

The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is not essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians. Consequently, results should also be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted.

Calibration evaluates the consistency of the measure in estimating episode cost across the full range of reSource: use patterns in the population. It is estimated by the average predictive ratios across groups within the population. We calculated the predictive ratio using the formula of average expected cost / average observed cost for all episodes in each decile. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. Below is the predictive ratio by decile of predicted episode cost:

Decile 1: 0.91

Decile 2: 0.97

Decile 3: 1.00

Decile 4: 1.01

Decile 5: 1.01

Decile 6: 1.03

Decile 7: 1.02

Decile 8: 1.02

Decile 9: 1.01

Decile 10: 0.98

This demonstrates that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, ranging between 0.91 and 1.03. Overall, the risk adjustment model does not over- or under-predict cost across the full range of reSource: use patterns in the population.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

To assess the impact on cost this measure may have, we examine the share of Medicare Parts A&B spending that this measure covers during the performance year 2019, assuming a volume threshold of 20 episodes. At the TIN-NPI reporting level, this measure covers 1.88% of Medicare Parts A&B spending. This measure can help to encourage more cost efficient care related to heart failure chronic care.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Weighted Heteroscedastic Within-Group Variance Estimation

Signal-to-Noise: Sample size

19,843

Signal-to-Noise: Statistical result

0.609

Signal-to-Noise: Interpretation of results

We conducted reliability testing of measures for clinicians (TIN-NPIs), constructed using episodes ending between January 1, 2019 and December 31, 2019. Reliability refers to the extent to which a measure reflects true variation between risk-adjusted episode spending of the clinicians, as opposed to random variation. The reliability metric specifically captures how much of the variance in a measure is due to systematic differences in episode spending between clinicians, rather than differences in episode spending within the set of episodes of a clinician. A measure with high reliability suggests that comparisons of performance across clinicians can be expected to better reflect systematic differences in actual performance. In the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171), CMS identified reliability levels between 0.4 to 0.7 as moderate and reliability levels above 0.7 as high. In the CY 2017 Quality Payment Program final rule, CMS also identified a threshold of 0.4 for mean reliability to be applied for measures and this was reiterated as the threshold in the CY 2022 Physician Fee Schedule Final Rule (86 FR 64996). Our testing results indicate that this measure has moderate reliability for clinicians across a range of volume thresholds.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Validity is tested empirically by examining the association between the measure score and high-cost events that drive the measure score, such as downstream complications and consequences of care.

Empiric Validity: Sample size

19,843

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

All providers:

Mean Score: $12,118 (SD $3,510)

Acute Inpatient Stay:

No event (Frequency: 0%) Mean Score: $6,844 (SD $2,689)

Q1 (1.5-15.1%) Mean Score: $9,898 (SD $2,926)

Q2 (15.1-20.6%) Mean Score: $11,457 (SD $2,835)

Q3 (20.6-27.1%) Mean Score: $12,671 (SD $3,030)

Q4 (27.1-70.0%) Mean Score: $14,538 (SD $3,406)

Emergency Department

No event (0%) Mean Score: $5,483 (SD N/A)

Q1 (3-31.8%) Mean Score: $11,074 (SD $3,468)

Q2 (31.8-38.9%) Mean Score: $11,815 (SD $3,363)

Q3 (38.9-46.2%) Mean Score: $12,357 (SD $3,295)

Q4 (46.2-86.4%) Mean Score: $13,233 (SD $3,551)

This analysis examines the cost pattern when there is a concurrent high cost event to demonstrate validity. High cost events are expensive services/facility stays that are grouped to the episode through service assignment or auto-grouping, such as acute inpatient stays and emergency department visits. These metrics show a specific subset of potentially high-costs events that could influence performance on the Heart Failure cost measure, as well as the mean score and standard deviation associated with how frequently each type of high-cost events occurs with a 20-episode volume threshold applied. We would expect to see that providers who have more instances of high-cost events have higher mean scores. The results show that the measure score is not impacted until a provider has substantially more high-cost events than their peers. Specifically, providers with lowest frequency of high-cost events, either 0% or at Q1 or Q2, had lower mean score than the overall mean score for all providers. In other words, the measure differentiates performance based the relative frequency of high-cost events compared to peers instead of the simple presence of high-cost events during a performance period.

Empiric Validity: Interpretation of results

Yes

Face Validity

No

Face Validity: Number of voting experts and patients/caregivers

N/A

Face Validity: Result

N/A

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Other: The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians.

Measure Performance Score Interpretation

Lower score is better

Mean performance score

12,118

Median performance score

11,711

Minimum performance score

2,310

Maximum performance score

37,010

Standard deviation of performance scores

3,510

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare & Medicaid Services

Measure Steward Contact Information

Donta Henson

Center for Clinical Standards and Quality

7500 Security Boulevard

Baltimore, MD 21244

Donta.Henson1@cms.hhs.gov

410-786-1947

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Joyce Lam

Acumen, LLC

500 Airport Blvd. Suite 100

Burlingame, CA 94010

ccsq-macra-support@acumenllc.com

650-558-8882

Secondary Submitter Contact Information

Anastasiia Biriuchinskaia

Acumen, LLC

500 Airport Blvd. Suite 100

Burlingame, CA 94010

ccsq-macra-support@acumenllc.com

650-558-8882

Submitter Comments

Please note that we selected "other" to describe the type of measure performance score that the measure uses, but the description was not visible. The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians. Please also note that since we were not able to enter the empiric validity results in the "Empiric Validity: Statistical result" field, we added those results in the "Empiric Validity: Methods and findings" field, before summarizing the empiric validity methods and findings. Please let us know if you have any questions.

MUC2022-129 Psychoses and Related Conditions

Program

Merit-based Incentive Payment System-Cost

Section 1: Measure Information

Measure Specifications and Endorsement Status

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 clinicians 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 clinicians 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.

Numerator

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

Numerator Exclusions

N/A

Denominator

The measure denominator is the total number of episodes from the Psychoses/Related Conditions episode group attributed to a clinician.

Denominator Exclusions

The following populations are excluded from the measure to ensure data completeness:

- Patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the trigger day.
- Patient was not enrolled in Medicare Parts A/B for the entirety of the lookback period plus episode window, or was enrolled in Part C for any part of the lookback plus episode window.
- No clinician group (identified by TIN) is attributed the episode.
- The date of birth of the patient is missing.
- The death date of the patient occurred before the episode ended.
- The trigger IP stay has the same admission date as another IP stay.

Exclusions specific to the Psychoses/Related Conditions measure are:

- All episodes not meeting triggering logic
- Beneficiary death in episode

- Not an acute hospital or psychiatric facility
- Outlier
- No attributed TIN
- Involuntary holds at admission
- Transferred to state psychiatric hospitals
- TIN does not meet testing volume threshold
- TIN-NPI does not meet testing volume threshold.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Psychiatry, Internal Medicine, Family Practice

Measure Type

Cost/ReSource: Use

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Inpatient psychiatric facility;Hospital inpatient acute care facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Affordability and Efficiency

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

Yes

Previous Measure Information

In what prior year was this measure published?

2018

What was the MUC ID for the measure in this year?

MUC18-119

List the CMS CBE MAP workgroup(s) in this year:

MAP Clinician Workgroup, 2018

What were the programs that MAP reviewed the measure for in this year?

Merit-Based Incentive Payment System (MIPS), 2019

What was the MAP recommendation in this year?

2019, Do Not Support

Why was the measure not recommended by the MAP workgroups in this year?

In the 2019 Final Report, the MAP discussed their rationale for not supporting the Psychoses/Related Conditions measures. They stated specifically, MAP expressed concerns about the measures validity with respect to the attribution model, noting that the measure may ineffectively assess quality of care in the target population due to several factors which fall outside the clinicians locus of control. MAP noted that patients with psychosis or related conditions require community supports but the availability of such supports can vary significantly depending on where a patient resides. MAP also noted that these conditions are often accompanied by a number of physical comorbidities that are not treated by the clinician managing the mental health of the patients but which could influence the results of this measure. Finally, MAP noted that many outpatient behavioral health clinicians do not accept Medicare or Medicaid and cautioned that this measure could exacerbate access issues.

MAP report page number being referenced for this year:

MAP 2019 Considerations for Implementing Measures in Federal Programs: Merit-Based Incentive Payment System (MIPS) and Medicare Shared Savings Program (SSP); page 7-8

What is the history or background for including this measure on the new measures under consideration list?

Measure previously submitted to MAP, refined and resubmitted per MAP recommendation

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

Section 101(f) of MACRA

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Claims

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

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.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Analysis of all TIN-NPIs with at least 20 attributed episodes for the 2019 performance period shows a large range of provider scores on Psychoses/Related Conditions measure. The measure score has the following distributional characteristics:

Mean: $20,418, standard deviation: $4,549

Median: $19,876

Min: $6,950, max: $43,094

Interquartile range is $5,825

Coefficient of variation: 0.22

The score decile distribution for the 2019 performance period is:

10th: $15,172

20th: $16,632

30th: $17,685

40th: $18,702

50th: $19,876

60th: $21,011

70th: $22,230

80th: $23,973

90th: $26,678

Unintended Consequences

No unintended consequences to individuals or populations have been identified during testing.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

Please see evidence attachment.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

538,186 total clinicians participated in MIPS as either a clinician group or individuals in 2019. We estimate a similar number of clinician groups and individual clinicians will continue to participate in MIPS.

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Other (enter here):: Medicare Part D enrollment status

Patient-level demographics: please select all that apply:

Age

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Health behaviors/health choices

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

To determine whether the risk model adequately accounts for confounding factors, we assessed two factors: discrimination and calibration. In this case, discrimination is the ability to explain the variance in cost of individual episodes. The amount of variance explained is estimated by the R-squared metric with the range between 0 and 1. The R-square for the measure is 0.082, and 0.078 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 7.8% of the variation in the actual observed cost of episodes is explained by the risk adjustment model and sub-group stratification.

The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is not essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians. Consequently, results should also be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted.

Calibration evaluates the consistency of the measure in estimating episode cost across the full range of reSource: use patterns in the population. It is estimated by the average predictive ratios across groups within the population. We calculated the predictive ratio using the formula of average expected cost / average observed cost for all episodes in each decile. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. Below is the predictive ratio by decile of predicted episode cost:

Decile 1: 0.97

Decile 2: 0.99

Decile 3: 1.01

Decile 4: 1.02

Decile 5: 1.02

Decile 6: 1.01

Decile 7: 1.01

Decile 8: 0.99

Decile 9: 0.99

Decile 10: 1.00

This demonstrates that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, ranging between 0.97 and 1.02. Overall, the risk adjustment model does not over- or under-predict cost across the full range of reSource: use patterns in the population.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

To assess the impact on cost this measure may have, we examine the share of Medicare Parts A&B spending that this measure covers during the performance year 2019, assuming a volume threshold of 20 episodes. At the TIN-NPI reporting level, this measure covers 0.45% of all Medicare Parts A&B spending. This measure can help to encourage more cost efficient care related to Psychoses/Related Conditions acute care.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Homoscedastic Mean of Individual Variance Estimation

Signal-to-Noise: Sample size

5,129

Signal-to-Noise: Statistical result

0.854

Signal-to-Noise: Interpretation of results

We conducted reliability testing of measures for clinicians (TIN-NPIs), constructed using episodes ending between January 1, 2019 and December 31, 2019. Reliability refers to the extent to which a measure reflects true variation between risk-adjusted episode spending of clinicians, as opposed to random variation. The reliability metric specifically captures how much of the variance in a measure is due to systematic differences in episode spending between clinicians, rather than differences in episode spending within the set of episodes of a clinician . A measure with high reliability suggests that comparisons of performance across clinicians can be expected to better reflect systematic differences in

actual performance. In the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171), CMS identified reliability levels between 0.4 to 0.7 as moderate and reliability levels above 0.7 as high. In the CY 2017 Quality Payment Program final rule, CMS also identified a threshold of 0.4 for mean reliability to be applied for measures and this was reiterated as the threshold in the CY 2022 Physician Fee Schedule Final Rule (86 FR 64996). Our testing results indicate that this measure has high reliability for clinicians across a range of volume thresholds.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Validity is tested empirically by examining the association between the measure score and high-cost events that drive the measure score, such as downstream complications and consequences of care.

Empiric Validity: Sample size

5,129

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

Observed Cost:

All Episodes

Mean: $17,136 (SD $11,849)

Median: $13,384

Inpatient Stay Readmission:

Mean: $24,764 (SD $11,364)

Median: $22,195

Emergency Department Visit:

Mean: $18,597 (SD $11,197)

Median: $16,080

Risk-Adjusted Cost:

All Episodes

Mean: $17,127 (SD $11,230)

Median: $13,571

Inpatient Stay Readmission:

Mean: $25,613 (SD $11,536)

Median: $23,237

Emergency Department Visit:

Mean: $19,396 (SD $11,465)

Median: $16,544

This analysis examines the cost pattern when there is a concurrent high cost event. High cost events are expensive services/facility stays that are grouped to the episode through service assignment or auto-grouping, such as acute inpatient and post-acute care. These metrics show a specific subset of potentially high-costs events that could influence performance on the Psychoses/Related Conditions cost measure. The results show that episodes with a readmission or an emergency department visit during the episode window of the cost measure have higher mean observed and risk-adjusted costs than the overall population of episodes included in the measure. The cost measure is able to differentiate the cost efficiency of episodes based on high-cost events.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

5

Face Validity: Result

5

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Other: The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians.

Measure Performance Score Interpretation

Lower score is better

Mean performance score

20,418

Median performance score

19,876

Minimum performance score

6,950

Maximum performance score

43,094

Standard deviation of performance scores

4,549

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare & Medicaid Services

Measure Steward Contact Information

Donta Henson

Center for Clinical Standards and Quality

7500 Security Boulevard

Baltimore, MD 21244

Donta.Henson1@cms.hhs.gov

410-786-1947

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Joyce Lam
Acumen, LLC
500 Airport Blvd. Suite 100
Burlingame, CA 94010
ccsq-macra-support@acumenllc.com
(650) 558-8882

Secondary Submitter Contact Information

Anastasiia Biriuchinskaia
Acumen, LLC
500 Airport Blvd. Suite 100
Burlingame, CA 94010
ccsq-macra-support@acumenllc.com
(650) 558-8882

Submitter Comments

Please note that we selected "other" to describe the type of measure performance score that the measure uses, but the description was not visible. The score is a ratio multiplied by the national average cost for an episode to create a dollar figure that may be more meaningful to clinicians. Please also note that since we were not able to enter the empiric validity results in the "Empiric Validity: Statistical result" field, we added those results in the "Empiric Validity: Methods and findings" field, before summarizing the empiric validity methods and findings. Please let us know if you have any questions.

Merit-based Incentive Payment System-Quality

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a noise value greater than a threshold specific to the CT Category.

Numerator Exclusions

None

Denominator

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

Denominator Exclusions

Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing noise. These are technical exclusions ("missing data") from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

Denominator Exceptions

None

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

All payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Diagnostic radiology

Measure Type

Intermediate Outcome

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Clinical Data (non-EHR);Electronic Health Record

If applicable, specify the data source

N/A

Description of parts related to these sources

(1) The measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems, including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS). These are labeled A and B below. (2) Primary imaging data stored in structured fields in the radiology electronic clinical data systems have been historically inaccessible using the existing eCQM framework. (3) Thus, the eCQM cannot consume CT images and Radiation Dose Structured Reports (RDSR, which contain the radiation dose) in their original DICOM formats. These primary data, listed below, must be transformed into calculated data elements that can then be ingested by the eCQM. (4) This is described in the feasibility attachment. The measure developers have created software (available for free to reporting entities) to transform primary data elements from these electronic systems to generate variables that the eCQM uses to calculate the measure score. These electronic systems include (A) EHR: The measure characterizes CT exams based on the type of exam performed (derived from procedure (CPT) codes associated with the exam bill), and the reason for study (derived from diagnosis (ICD-10-CM) codes associated with the exam order and bill). (Data element Diagnostic study, performed: CT Studies) During transformation, a validated algorithm uses combinations of CPT and ICD-10-CM codes to generate the CT Dose and Image Quality Category (CT category, LOINC code 96914-7) that specifies the radiation dose and image quality thresholds for each CT exam. The measure also derives birth date to calculate age at the start of the measurement period, and supplemental data elements including payer, race, ethnicity, and sex. (B) RADIOLOGY ELECTRONIC CLINICAL DATA SYSTEMS (NON-EHR): The PACS stores CT exam data generated by CT machines during the ordinary course of care, including image pixel data (data element Diagnostic Study Performed: CT Studies Result attribute: Image Pixel Data) and Radiation Dose Structured Reports

(RDSR) (data element Diagnostic Study Performed: CT Studies Result attribute: Radiation Dose Structured Report (RDSR)) Both of these data are formatted and stored as DICOM structured data. These primary data elements are used for calculating inputs to the eCQM, including the Calculated CT Size-Adjusted Dose (size-adjusted dose, LOINC code 96913-9) and Calculated CT Global Noise (noise, LOINC code (96912-1), respectively.

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Ambulatory/office-based care;Community hospital;Emergency department;Hospital outpatient department (HOD);Hospital inpatient acute care facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

This measure correlates with the following two improvement activities: (1) IA_PSPA_8: Use of Patient Safety Tools. Clinicians must use tools that assist specialty practices in tracking specific measures that are meaningful to their practice. This measure allows clinicians to see their performance to guide dose optimization. And (2) IA_PSPA_19: Implementation of formal quality improvement methods, practice changes, or other practice improvement processes.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

06138

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

NQF IDs: 3633e (clinician level) and 3662e (clinician group level)

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

2022

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

Yes

If eCQM, enter Measure Authoring Tool (MAT) number

MAT eCQM identifier: 1056 (QDM version) and 1076FHIR (FHIR version).

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

Yes

If eCQM, does any electronic health record (EHR) system tested need to be modified?

Yes

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

Submitted previously but not included in MUC List

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

Three existing MIPS process measures are related (not competing) in that they address patient safety related to radiation exposure in CT imaging: (1) Optimizing Patient Exposure to Ionizing Radiation: Count

of Potential High Dose Radiation Imaging Studies: Computed Tomography (CT) and Cardiac Nuclear Medicine Studies (CMIT 2286); (2) Radiation Consideration for Adult CT: Utilization of Dose Lowering Techniques (CMIT 2570); and (3) Multi-strata weighted average for 3 CT Exam Types: Overall Percent of CT exams for which Dose Length Product is at or below the size-specific diagnostic reference level (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single phase scan and CT Head/Brain without contrast/single phase scan) (ACRAD34).

How will this measure be distinguished from other similar and/or competing measures?

See related measures attachment.

How will this measure add value to the CMS program?

(1) IT WOULD BE THE ONLY RADIOLOGY ECQM IN THE CMS MEASURES INVENTORY, aligning with CMS's goal of transitioning to all digital quality measures by 2025. Our measure is designed using both QDM and FHIR specifications, supporting CMS's stated intention of encouraging healthcare information interoperability based on standard APIs, specifically FHIR. (2) IT IS THE FIRST AND ONLY MEASURE TO ASSESS IMAGE QUALITY as a means of protecting the diagnostic value of CT imaging from unintended consequences of excessive radiation dose reduction. (3) IT ASSESSES RADIATION DOSE AND IMAGE QUALITY BASED ON THE UNDERLYING CLINICAL INDICATION – in other words, the reason the patient was imaged – and not based simply on the exam that was performed, which often results in doses higher than needed for diagnosis. The measure covers the two key process of care components that determine the radiation doses, including: (a) the choice of imaging protocol (i.e. the type of CT exam - for example, whether a patient is imaged with a single- or double-phase CT exam); and (b) decisions regarding the technical settings used for that type of CT exam, which are usually at the discretion of the technologist or medical physicist who oversees and operates the machines. Both components contribute to radiation dose, and as a result, a comprehensive quality measure must encompass both of these decision-making processes. This measure is uniquely able to encompass both components. (4) THE DENOMINATOR INCLUDES MOST DIAGNOSTIC CT EXAMS in adults, including multiphase high dose examination types. And (5) THE MEASURE ADJUSTS FOR PATIENT SIZE, an important contributor to dose.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

eCQM

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

Feasibility testing was conducted in 8 different EHR systems reflecting 606 individual clinicians and 16 clinician groups [Epic (N=5), Cerner (N=1), Allscripts (N=1), MedInformatix (N=1)], and evaluated the availability, accuracy, standardization, and workflow relative to each data element used in the measure. All data elements were found to be available and accessible, accurate, and structured in standardized vocabularies. Generating and collecting the data elements had no impact on clinician workflow. Please see feasibility attachment for more details on how feasibility was evaluated, as well as how the measure will be operationalized.

Method of Measure Calculation

eCQM

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

The measure was developed to address a considerable performance gap in the use of excessive and highly variable radiation dose in CT imaging. Doses used for CT vary substantially across imaging facilities for patients imaged for the same clinical indication. For example, (1) In a prior study of 151 imaging facilities and hospitals, even 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 2019). (2) EVIDENCE IN THE UCSF REGISTRY: When we applied the proposed measure to the UCSF International CT Dose Registry – a repository of CT data containing over 8 million exams from 161 hospitals and imaging facilities – overall 33% of CT exams were out-of-range based on radiation dose criteria. Overall, 135 facilities (84%) had out-of-range scores over 10%. (3) EVIDENCE IN THE FIELD-TESTING DATA: In the field-testing performed across 16 clinician groups and 606 individual clinicians – the rates of out-of-range?exams varied 20%-43%?by clinician group and 0-100% by individual clinician.?Virtually all of this was driven by excessive radiation doses, as extremely few CT exams?were assessed as out-of-range based on?noise: on average <1% across all reporting entities. Less than 5% of individual clinicians had an out-of-range score based on noise of 1.4% or greater, but these were uniformly clinicians with very low sample size. (4) SUMMARY: This variation in radiation dose underscores the performance gap that the measure addresses, and these outcomes indicate a considerable opportunity to reduce doses without impacting quality.

Unintended Consequences

There is a relationship between image quality and radiation dose such that, as radiation dose increases, image quality increases until a threshold is reached, at which point no further diagnostic benefit from image quality occurs. Conversely, too little radiation dose can produce inadequate image quality. Thus, image quality must remain diagnostically sufficient as excessive doses are lowered. The actual risk for this is low, as research suggests doses may be lowered between 50-90% without impacting image diagnostic utility (den Harder 2018, Rob 2017, Konda 2016, Huppertz 2015). In our field-testing data, out-of-range measure scores due to inadequate image quality (i.e. excessive noise) were exceedingly rare, with less than 1% of exams, on average, across all reporting entities. This was to some degree expected, given the results of an Image Quality Study – performed as part of measure development – in which radiologists graded 3% and 8% of exams as “poor” or “marginally acceptable” image quality,

respectively (manuscript in preparation). These findings support a considerable opportunity to reduce radiation doses without impacting quality. Given the evidence of harm from excessive radiation, and the low likelihood of deteriorating image quality to the point of rendering exams unacceptable, there is little question that the benefit outweighs the cost of dose optimization. Nevertheless, the measure steward will monitor out-of-range rates annually to determine if image quality is worsening due to declining radiation doses and determine if thresholds should be adjusted or if a subsequent radiologist satisfaction study should be repeated.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

7

Outline the clinical guidelines supporting this measure

The proposed measure aligns with numerous consensus-based clinical recommendations and guidelines asking radiologists to track, optimize, and lower the radiation doses they use for CT. These guidelines are based on evidence that radiation doses are highly variable across institutions, higher than needed for diagnosis, and can lead to excessive patient harm. These recommendations and guidelines have been written by: the American College of Radiology (Kanal 2017); a collaboration of the American College of Radiology, The American Association of Physicists in Medicine, and the Society for Pediatric Radiology (ACR-AAPM-SPR 2018); the Radiological Society of North America (Hricak 2010); the Society of Interventional Radiology (Stecker 2009); the Society of Cardiovascular CT (Halliburton 2011); Image Gently, an initiative of the American College of Radiology, the Radiological Society of North America, American Society of Radiologic Technologists, and American Association of Physicists in Medicine (Goske 2008); and the FDA (US Food and Drug Administration 2019). The most common approach advised is for physicians to collect and compare their doses to benchmarks and to reduce their doses if they are found to routinely exceed these benchmarks.

Name the guideline developer/entity

The guideline was jointly developed by the American College of Radiology (ACR), the American Association of Physicists in Medicine (AAPM), and the Society of Pediatric Radiology (SPR).

Publication year

2018

Full citation +/- URL

ACR-AAPM-SPR Practice Parameter for Diagnostic Reference Levels and Achievable Doses in Medical X-Ray Imaging. Revised October 1, 2018. <https://www.acr.org/-/media/ACR/Files/Practice-Parameters/diag-ref-levels.pdf>.

Is this an evidence-based clinical guideline?

No

Is the guideline graded?

No

List the guideline statement that most closely aligns with the measure concept.

The establishment of reference levels in diagnostic medical imaging requires close cooperation and communication between the team of physicians who are responsible for the clinical management of the patient, the Qualified Medical Physicist who is responsible for monitoring equipment and image quality and estimating patient dose, and the radiologic technologist who is responsible for adherence to protocols. Adherence to this practice parameter should help maximize the efficacy of these procedures, optimize patient radiation dose and image quality, minimize radiation dose to staff, maintain safe conditions, and ensure compliance with applicable regulations. This is particularly important for children who are more vulnerable than adults to the potential risks of ionizing radiation.

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

The establishment of reference levels in diagnostic medical imaging requires close cooperation and communication between the team of physicians who are responsible for the clinical management of the patient, the Qualified Medical Physicist who is responsible for monitoring equipment and image quality and estimating patient dose, and the radiologic technologist who is responsible for adherence to protocols. Adherence to this practice parameter should help maximize the efficacy of these procedures, optimize patient radiation dose and image quality, minimize radiation dose to staff, maintain safe conditions, and ensure compliance with applicable regulations. This is particularly important for children who are more vulnerable than adults to the potential risks of ionizing radiation.

Number of systematic reviews that inform this measure concept

3

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

Please see systematic reviews evidence attachment.

Source of empirical data

Published, peer-reviewed original research

Summarize the empirical data

(1) THERE IS EVIDENCE OF A RELATIONSHIP BETWEEN PROCESS INTERVENTIONS (SPECIFICALLY, EDUCATIONAL FEEDBACK SIMILAR TO THAT PROVIDED BY THIS MEASURE) AND THE INTERMEDIATE OUTCOME OF THIS MEASURE, RADIATION DOSE. In a randomized controlled trial involving roughly 1 million CT exams from 100 imaging facilities across 6 countries, Smith-Bindman et al. observed that multicomponent educational feedback achieved a 23-58% reductions in the proportion of high-dose exams, based on organ dose, with no observed change in image quality. (Smith-Bindman 2020) Another interventional study across the University of California system deployed radiation dose audits and best practice sharing, resulting in considerable dose reductions: a 19% and 25% decrease in mean effective dose for chest and abdomen exams, respectively, and a reduction in the number of exams exceeding allowable benchmarks by 48% and 54% for chest and abdomen, respectively. (Demb 2017). (2) THERE IS EXTENSIVE EPIDEMIOLOGICAL AND BIOLOGICAL EVIDENCE THAT SUGGESTS EXPOSURE TO RADIATION IN THE SAME RANGE AS THAT ROUTINELY DELIVERED BY CT (10-100 MILLI-SIEVERTS, MSV) INCREASES A PERSON'S RISK OF DEVELOPING CANCER (Board of Radiation Effects 2006, Pearce 2012, Pierce 2000, Preston 2007, Brenner 2003, Hong 2019). In a case-control study of over 3 million adult patients imaged between 2000-2013 in Taiwan, Shao et al. found that exposure to CT imaging was associated with elevated risk of thyroid cancer (OR = 2.55, 95% CI = 2.36 to 2.75) and leukemia (OR = 1.55, 95% CI = 1.42 to 1.68) for all patients, with higher risk in women, and for non-Hodgkin lymphoma in patients aged 45 or younger. (Shao 2019) A clear dose-response relationship was observed in patients 45 years or younger for all three cancers. (3) DESPITE THE KNOWN RISKS OF CT, ITS USE HAS GROWN SUBSTANTIALLY over the last few decades (Harvey L Neiman 2017), with 91.4 million CT exams performed in the United States in 2019 (IMV 2020), including 428 exams per 1000 patients aged 65 years and older (Smith-Bindman 2019). It was estimated in 2009 that 2% of cancers diagnosed annually are the result of CT; in 2019 that would amount to 36,000 cancers diagnosed each year due to the use of CT. (Berrington de Gonzalez 2009, NCI Cancer Statistics).

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

45,500,000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines;Peer-Reviewed Systematic Review;Empirical data

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level health status & clinical conditions

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

Other (enter here):: Patient size

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

For each CT category, the dose-length product used to classify an accountable entity as "out of range" is adjusted for patient diameter using a log-linear Gaussian mixed model that includes the dose-length product as the outcome, the patient size as the fixed effect of interest, and the institution at which the exam was performed as a confounding random effect. The adequacy of the resulting size-adjusted dose-length product was assessed using the same model, but with the outcome of (raw) dose-length product replaced with the size-adjusted dose-length product. Prior to size adjustment, the marginal R-squared relating patient diameter to dose was 0.08 for the average CT category, increasing to as high as 0.29 for the CT category (Low Dose Abdomen) with the strongest relationship between patient diameter and dose-length product. After size adjustment, the marginal R-squared relating patient diameter to dose is uniformly <0.01 for all CT categories. This suggests that the adjustment mechanism has adequately removed bias from patient diameter, a potential confounder of the relationship between dose-length product and quality of care. Please see the risk adjustment methodology attachment for further details.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

COST IMPACT: The measure is expected to result in cost savings to Medicare of $1,859,606,000 to $5,206,896,800 annually, based on an estimate of $133,000 - $372,400 per cancer avoided. Implementation costs to reporting entities are expected to be around $2600-3250 per practice annually. ASSUMPTIONS BEHIND COST SAVINGS ESTIMATE: Based on the current estimated number of CT exams performed annually in the U.S. [IMV 2020], distribution in exam types and observed doses [Demb 2017, Smith-Bindman 2019], and modelling of the cancer risk associated with CT [Berrington de Gonzalez 2009], 18,643 cancers could be prevented annually by reducing doses to the median measure score from our testing data. The majority of these cancers will be prevented among elderly adults because imaging rates are nearly five times higher in that population [Smith-Bindman 2019], and because absolute and excess cancer rates are higher among older adults compared with non-elderly adults or children [Berrington de Gonzalez 2009, Shuryak 2010]. We estimate that 75% of all cancers prevented annually (13,982) will occur among Medicare beneficiaries who undergo CT, and that approximately 3 cancers would be prevented per 10,000 Medicare patients who undergo CT (or 1 cancer per 3,254 patients). The cost avoided by the measure reflects the cost of cancer cases prevented. The cost of care for breast, colorectal, and lung cancer during the 4 years after diagnosis in 2011 was estimated at $100,000-$280,000 per case [Dieguez 2017]. This estimate was based on actual costs incurred between 2011-2014 and was not adjusted for inflation, though cancer care costs were projected to rise 27-39% between 2011 and 2020 (Mariotto 2011). Using a mean inflation rate of 33% between 2011 and 2020, this reflects a 4-year cost per cancer ranging from $133,000 to $372,400 per case avoided. Using this average cost of cancer care ($133,000-$372,400) and the number of cancers prevented annually among Medicare beneficiaries (13,982). This results in $1.86 billion to $5.21 billion annual cost savings. Furthermore, cancer patients who survive beyond the first 4 years may continue to incur high costs, especially in the last year of life. Thus, these estimates could be lower than actual savings. ASSUMPTIONS BEHIND IMPLEMENTATION/REPORTING COST ESTIMATE: For clinician and clinician group reporting, costs will be incurred at locations where clinicians work, including outpatient imaging facilities and inpatient and outpatient hospital settings. The costs derive primarily from IT assembling the data from the relevant data Sources: . We estimate the implementation costs per location at $3250 per hospital and $2600 per outpatient imaging facility. These estimates are based on the costs reported by our field-testing sites (see feasibility attachment for more information). This cost estimate is conservative for two reasons: (1) as our testing partners noted, the work of assembling the relevant data decreased over time. (2) Further, because there are shared data systems across multiple hospital and outpatient facilities, these costs will likely be lower as implementation occurs only once across multiple locations with shared data sources.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

2

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

2

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

16

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

16

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

Yes

If yes, how many sites were evaluated in the provider workflow analysis?

16

Did the provider workflow have to be modified to accommodate the new measure?

No

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Random Split-Half Correlation

Signal-to-Noise: Name of statistic

N/A

Signal-to-Noise: Sample size

N/A

Signal-to-Noise: Statistical result

N/A

Signal-to-Noise: Interpretation of results

N/A

Random Split-Half Correlation: Name of statistic

We estimated measure score reliability at the accountable entity level using the intraclass correlation coefficient (ICC), a reliability coefficient that conceptually represents the true (between-entity) variance in a measure divided by the sum of true variance and error (within-entity) variance. We used randomly split samples for each accountable entity with 1,000 repetitions, applying a one-way random effects model, assuming that both entity effects and residual effects are random, independent, and normally distributed with mean 0. This approach corresponds to Case 1 or the ICC(1) in McGraw and Wong's seminal description of ICC reliability methods. (McGraw 1996) The Spearman-Brown prophecy formula was applied, in the usual manner, to adjust reliability from one-month test samples to the anticipated 12-month sample (i.e., (12*r)/(1 + (11*r)). (Frey 2018) These ICC(1) estimates (bounded between 0 and 1) were then logit-transformed and used to model the linear relationship between entity volume and logit reliability. By ranking predicted reliabilities across the complete range of potential volumes, we estimated the volume threshold that would correspond to ICC(1)=0.9 for an accountable entity. At the individual clinician level, clinicians who read only 1 CT exam during the testing month (equivalent to 12 in a year) were excluded from reliability analysis because split half sampling was impossible.

Random Split-Half Correlation: Sample size

606

Random Split-Half Correlation: Statistical result

0.99

Random Split-Half Correlation: Interpretation of results

According to the scale developed by Koo and Li, an ICC estimate between 0.75-0.90 may be interpreted as good reliability, and an ICC estimate greater than 0.90 may be interpreted as excellent reliability (Koo 2016). Based on the mean ICC of 0.99, after Spearman-Brown adjustment to a 12-month reporting period (after excluding the 5% of clinicians who only read 1 CT scan during the testing period) the measure is reliable at the individual clinician level. Overall 8% of individual clinicians in our field-testing would not meet the minimum denominator to achieve ICC > 0.90. Please see reliability attachment for results at the clinician group level.

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

A logistic mixed model was used to determine whether a facility's proportion of radiation doses above the 75th percentile was predicted by process measures that are known to be associated with positive health outcomes. (Solberg 2020) Methods are described in the Validity Testing at the Accountable Entity Level Attachment.

Empiric Validity: Sample size

90

Empiric Validity: Statistical result

0.47

Empiric Validity: Methods and findings

Please see the Validity Testing at the Accountable Entity Level Attachment.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

17

Face Validity: Result

17

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between eCQM and manual reviewer;Agreement between other gold standard and manual reviewer

Sample Size

47,635

Statistic Name

Percent agreement

Statistical Results

0.95

Interpretation of results

See the Patient/Encounter Level Validity Testing Attachment, both for results at the clinician group level, and for more details on methods, results, and interpretation of results at the clinician and clinician group levels.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Lower score is better

Mean performance score

0.30

Median performance score

0.28

Minimum performance score

0.01

Maximum performance score

0.90

Standard deviation of performance scores

0.14

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Other: Alara Imaging, Inc. in collaboration with the University of California, San Francisco (UCSF)

Measure Steward Contact Information

Nate Mazonson
550 16th Street, Box 0560
San Francisco, CA 94044
nate@alaracare.com
(650)c520-6649

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Rebecca Smith-Bindman
550 16th Street, Box 0560
San Francisco, CA 94143
Rebecca.Smith-Bindman@ucsf.edu
(415) 377-7957

Secondary Submitter Contact Information

Carly Stewart
550 16th Street, Box 0560
San Francisco, CA 94143
carly.stewart@ucsf.ed
(954) 683-7859

Submitter Comments

EXECUTIVE SUMMARY: In the US, over 90 million CT scans are performed annually, and the radiation doses associated with these exams are a safety issue, as unnecessarily high radiation doses lead to harm by exposing patients to elevated cancer risk. Our measure fills this quality gap and is aligned with clinical recommendations, grounded in extensive epidemiologic evidence, and tested in diverse settings. The measure also supports CMS in moving from process or QCDR measures to intermediate outcome measures that focus on radiation-related risk reduction for exposed patients and populations. This measure is also the first radiology digital quality measure. Using electronic and standardized data already collected as part of routine clinical care, our measure assesses the radiation dose for every exam, taking into consideration the reason for the exam and patient size, and is coupled with an

assessment of imaging quality to ensure that efforts to reduce radiation dose do not result in poor image quality. The measure will improve patient safety, reduce population-level cancer risks, and reduce associated cancer-related morbidity, mortality, and cost. 100% of the diverse technical expert panel (TEP) members assembled for this measure's development agreed that performance on the measure as specified is a representation of quality, differentiating good from poor performance. Nearly all (16/17 of TEP members) agreed that the measure, if implemented, is likely or very likely to improve quality. The measure is also undergoing endorsement review by the National Quality Forum in the Fall 2021 cycle. The reliability and validity of the measure were considered acceptable for endorsement by the NQF Scientific Methods Panel in October 2021. Subsequently, the Patient Safety Standing Committee evaluated the measure in February 2022 and recommended NQF endorsement. In the related public commenting period, over 20 messages of support were submitted from various notable stakeholders and testing site partners. A final endorsement will be issued in July 2022.

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

The Feeling Heard and Understood measure is calculated using top-box scoring. The top-box score refers to the percentage of respondents that give the most positive response. For all four questions in this measure, the top box numerator is the number of respondents who answer "Completely true." An individual's score can be considered an average of the four top-box responses and these scores are adjusted for mode of survey administration and proxy assistance. Individual scores are combined to calculate an average score for an overall palliative care clinician or group.

Numerator note: This is a multi-item measure consisting of 4 items: Q1- "I felt heard and understood by this provider and team", Q2- "I felt this provider and team put my best interests first when making recommendations about my care", Q3- "I felt this provider and team saw me as a person, not just someone with a medical problem", Q4- "I felt this provider and team understood what is important to me in my life."

Numerator Exclusions

N/A

Denominator

All patients aged 18 years and older who had an ambulatory palliative care visit.

Denominator Criteria:

All patients aged 18 years and older on date of encounter.

AND

Ambulatory palliative care visit* defined as:

ICD-10 Z51.5 (Encounter for Palliative Care), OR

Provider Hospice and Palliative Care Specialty Code 17; AND

CPT 99201-99205 (New Office Visit); OR CPT 99211-99215 (Established Office Visit); or Place of service (POS) Code 11 - Office.

WITH

A MIPS-eligible provider.

*Telehealth visits were not included in testing.

Denominator Exclusions

- Patients who do not complete at least one of the four items in the multi-item measure.
- Patients who do not complete the patient experience survey within six months of the eligible ambulatory palliative care visit.
- Patients who respond on the patient experience survey that they did not receive care by the listed ambulatory palliative care provider in the last six months (disavowal).
- Patients who were deceased when the survey reached them.
- Patients for whom a proxy completed the entire survey on their behalf for any reason (no patient involvement).

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

All Payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Palliative care

Measure Type

Outcome - (PRO-PM)

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

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

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Person-Centered Care

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

We identified two potentially relevant Improvement Activities for MIPS quality measures: 1) Collection and follow-up on patient experience and satisfaction data on beneficiary engagement and 2) Implementation of formal quality improvement methods, practic

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

06117

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

3665

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

2022

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

Submitted previously but not included in MUC List

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

This measure is similar to Hospice CAHPS Survey: Communication with Family (setting: hospice). There are no competing measures.

How will this measure be distinguished from other similar and/or competing measures?

The Hospice CAHPS Survey: Communication with Family measure asks bereaved family caregivers of hospice patients how often the hospice team kept them informed, explained things in a way that was easy to understand, listened carefully to them, and gave them confusing or contradictory information.

The proposed Feeling Heard and Understood measure differs in key ways: 1) it captures the patient experience of care directly from the patient rather than a caregiver, which is facilitated by its use in ambulatory rather than hospice settings; and 2) it captures core interpersonal processes and relational aspects of care delivery that reflect whether a patient feels, seen, acknowledged, and respected, rather than acts and processes of communication such as information-sharing. This is particularly relevant to patients at the end of life receiving palliative care, who often report feeling silenced, ignored, and misunderstood in medical institutions.

Note: Although the proposed measure was developed in conjunction with another measure not in-use (MUC2021-092: Ambulatory palliative care patients' experience of receiving desired help for pain), the two measures differ significantly. The Feeling Heard and Understood measure assesses the extent to which patients feel seen and acknowledged by their palliative care provider and team, while the Receiving Desired Help for Pain measure assesses the extent to which patients received the help they wanted for the specific symptom of pain. Although we anticipate that performance on both measures will be informed in part by the overall strength and quality of the patient-provider relationship, the Feeling Heard and Understood measure is about interpersonal connection while the Receiving Desired Help for Pain measure is about preference-concordant pain management.

How will this measure add value to the CMS program?

In comparison to the related CAHPS communication measure designed for use in hospice settings, the proposed measure assesses feeling heard and understood, a construct that goes beyond patient-provider communication to reflect positive interpersonal relationship, adequate acknowledgement and respect, and a whole-person orientation. While existing quality measures focus on specific communication processes, systematically monitoring, reporting, and responding to how well patients feel heard and understood is crucial to creating and sustaining a health care environment that excels in caring for those who are seriously ill. Therefore, the proposed measure complements and adds an important dimension to existing quality measures of care planning and documentation of care pReferemces:. Further the proposed measure is designed for use among all patients receiving palliative care in ambulatory settings, where no other measures are in use. The proposed measure is and is harmonizable with existing performance measures.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Other: MIPS clinical quality measure

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

The proposed quality measure of ambulatory palliative care experience was rigorously tested in a multi-phase process. First, a patient experience survey instrument was created, including data elements (i.e., items) for the proposed measure. The survey instrument also included other patient-reported data elements to capture information on respondent health and sociodemographic information, use and details of proxy involvement in survey completion, perceived overall quality of care and communication, and related measure concepts such as receiving desired emotional support. These additional data elements were used for measure analyses, e.g., to explore data element-level and quality measure-level validity and quality measure score risk-adjustment. Once the instrument was drafted, testing was initiated in two sequential phases: 1) a pre-testing phase focused on cognitively testing and finalizing survey-based data elements for the proposed measures and establishing testing parameters through a small pilot, or alpha test among five palliative care groups and 2) a beta field test.

We conducted cognitive testing with patients and caregivers to evaluate comprehensibility and feasibility of administration for the data elements that comprise the proposed measure. We also conducted interviews with palliative care programs that participated in alpha and beta testing to assess perceived feasibility of the measures in clinical practice across providers and administrators, including the feasibility of identifying eligible patients using administrative data and data collection via survey vendor.

Through the beta field test, we sought to establish: 1) psychometric properties of the measure data elements; 2) the scientific acceptability of the measure specifications (their feasibility with regard to administration, mode, and calculation of the quality measures); and 3) final quality measure technical specifications, construction of the specifications regarding the numerator and denominator, and the reliability and validity of the quality measure. Results of the test are described under reliability and validity testing.

The patient experience survey developed for this measure is meant to be completed via web survey, on paper or over telephone in English. 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. Findings from the national beta field test indicate the feasibility of identifying eligible patients using administrative data and using a survey vendor to support survey administration and data collection. Details of this workflow are described in the attached Measure Information Form.

Method of Measure Calculation

Other (enter here):: Patient-reported data is collected via survey instrument. The instrument was developed for this measure and is meant to be completed via web survey, on paper or over telephone in English. Visit information for patient eligibility, pat

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Existing evidence suggests there is considerable room for improvement in the domain of communication in palliative care contexts (Frosch et al., 2012; Gramling et al., 2016; Ingersoll et al., 2018; Institute of Medicine Committee on Approaching Death, 2015). Analyses from the national beta field test demonstrate room for improvement in the Feeling Heard and Understood quality measure at the clinician level:

- The observed variability across palliative care clinicians (adjusted ICC point estimate = 0.150) supports the potential of the measure to distinguish among clinicians with high, medium, and low performance.
- Across the 229 clinicians in our 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 is 12.1. Confidence intervals for the highest and lowest clinician scores do not overlap: Lowest Clinician CI: (19.0, 67.4); Highest Clinician CI: (84.2, 95.2).
- When clinicians are ranked by their measure performance, we calculated that a clinician at the median of measure performance would need a large increase of 8.3 points in their measure score to improve to the 20th top-ranked clinician. 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.

Analyses from the national beta field test further demonstrate room for improvement in the Feeling Heard and Understood quality measure at the group level:

- The observed variability across palliative care groups (adjusted ICC point estimate = 0.052) supports the potential of the measure to distinguish among groups with high, medium, and low performance.
- Across the 44 palliative care groups in our sample, adjusted group scores range from 54.3 to 85.1 with an average adjusted measure score of 72.1. The standard deviation in average group scores is 7.1. Confidence intervals for the highest and lowest group scores do not overlap: Lowest Group CI: (42.2, 65.6); Highest Group CI: (77.2, 91.4).
- When groups are ranked by their measure performance, we calculated that a group at the median of measure performance would need a large increase of 4.19 points in their measure score to improve to the 20th top-ranked group. A group at the bottom of the ranking (e.g., the 10th lowest ranked group) would need a 7-point increase in measure score to improve to the median.

Citations:

Frosch, D. L., May, S. G., Rendle, K. A., Tietbohl, C., & Elwyn, G. (2012). Authoritarian physicians and patients' fear of being labeled "difficult" among key obstacles to shared decision making. Health Affairs, 31(5), 1030-1038.

Gramling, R., Stanek, S., Ladwig, S., Gajary-Coots, E., Cimino, J., Anderson, W., Norton, S. A., Aslakson, R. A., Ast, K., Elk, R., Garner, K. K., Gramling, R., Grudzen, C., Kamal, A. H., Lamba, S., LeBlanc, T. W., Rhodes, R. L., Roeland, E., Schulman-Green, D., & Unroe, K. T. (2016). Feeling Heard and Understood: A

Patient-Reported Quality Measure for the Inpatient Palliative Care Setting. J Pain Symptom Manage, 51(2), 150-154.

Ingersoll, L. T., Saeed, F., Ladwig, S., Norton, S. A., Anderson, W., Alexander, S. C., & Gramling, R. (2018). Feeling Heard & Understood in the Hospital Environment: Benchmarking Communication Quality Among Patients with Advanced Cancer Before and After Palliative Care Consultation. J Pain Symptom Manage, 56(2), 239-244.

Institute of Medicine Committee on Approaching Death (2015). Dying in America: Improving Quality and Honoring Individual PReferemces: Near the End of Life. National Academies Press.

Unintended Consequences

We have not encountered any unintended adverse consequences from measuring the extent to which patients feel heard and understood by providers. In qualitative interviews with palliative care groups that participated in the alpha pilot test and national beta field test, providers were asked about potential unintended consequences of the Feeling Heard and Understood measure. Providers noted that comparison across palliative care groups may be challenging if patient populations have differences in disease trajectories that impact communication. Another potential concern reported by providers was repercussions of negative feedback. There were concerns that some patients may have unrealistic expectations for palliative care, and patients whose expectations are not met may identify as not being heard and understood. Palliative care providers often have to deliver bad news to patients, which may negatively impact patient perceptions of the palliative care team. Providers recommended strategies to prevent some of these potential unintended consequences, including encouraging providers to establish expectations with patients up front and set realistic goals for palliative care. Providers also recommended framing the questions to help patients understand that the measure is useful for the clinician and group and ultimately for other patients.

In addition, it is possible that patients who have died may be contacted to complete the survey, potentially causing distress for families. Our recommended data collection approach is to first send eligible patients a letter notifying them of the upcoming survey with a stamped postcard that can be returned in the event of death or a move/new address.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

This quality measure assesses an outcome. The type of evidence referenced is the linkage between the outcome and at least one process of care.

Summarize the evidence

The importance of the proposed measure for being heard and understood is predicated on existing guidelines and conceptual models of the quality of palliative care, including the National Consensus Project Clinical Practice Guidelines for Quality Palliative Care (2018) supported by a systematic review (Ahluwalia et al., 2018), the National Quality Forum Preferred Practices of Palliative and Hospice Care (National Quality Forum, 2006) (i.e. Preferred Practice 7 and 9 and 24), a consensus building process from the National Coalition for Hospice and Palliative Care, and input from qualitative inquiry of patients and providers.

The goal of the proposed measure is to facilitate and improve effective patient-provider communication that engenders trust, acknowledgement, and a whole-person orientation to the care that is provided. The outcome that is the focus of the proposed quality measure is that the patient feels heard and understood by the ambulatory palliative care provider and team. The proposed measure is related to three NQF Preferred Practices for Palliative and Hospice Care Quality (National Quality Forum, 2006): #7 Ensure that upon transfer between healthcare settings, there is timely and thorough communication of the patient's goals, pReferemces:, values, and clinical information so that continuity of care and seamless follow-up are assured.; #9 - Patients and caregivers should be asked by palliative and hospice care programs to assess physicians'/healthcare professionals' ability to discuss hospice as an option; and #24 - Incorporate cultural assessment as a component of comprehensive palliative and hospice care assessment, including but not limited to locus of decision making, preferences regarding disclosure of information, truth telling and decision making, dietary preferences, language, family communication, desire for support measures such as palliative therapies and complementary and alternative medicine, perspectives on death, suffering, and grieving, and funeral/burial rituals.

Citations:

Ahluwalia, S. C., Chen, C., Raaen, L., Motala, A., Walling, A. M., Chamberlin, M., O'Hanlon, C., Larkin, J., Lorenz, K., Akinniranye, O., & Hempel, S. (2018). A Systematic Review in Support of the National Consensus Project Clinical Practice Guidelines for Quality Palliative Care, Fourth Edition. J Pain Symptom Manage, 56(6), 831-870.

National Consensus Project for Quality Palliative Care. Clinical Practice Guidelines for Quality Palliative Care, 4th edition. Richmond, VA: National Coalition for Hospice and Palliative Care; 2018. https://www.nationalcoalitionhpc.org/ncp

National Quality Forum. (2006). A National Framework and Preferred Practices for Palliative and Hospice Care Quality: A Consensus Report.

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

Yes

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Other (enter here):: This measure is an outcome and is linked to at least one process of care.

Is the measure risk adjusted?

Yes

Risk adjustment variables

Other (enter here):: Survey mode; proxy assistance

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

We used Kendall's tau to assess the unadjusted and adjusted scores and explore how rankings among providers change after risk adjustment. We also used statistical tests to assess the significance of covariates in the risk adjusted model and discussed thes e results with our technical expert panel. Tests using Kendall's tau compare the rank order of unadjusted scores to the order of risk-adjusted scores and assess the percentage of cases where the order has changed (i.e., 100 * (1 - tau) / 2). A statistic of 1 would imply that risk adjustment has no effect on the rank order of programs and a statistic of -1 would imply that the order is completely reversed by risk adjustment. Values of 0.8 to 0.95 are typical of those reported in NQF documentation for the CAHPS surveys (Parast et al., 2018). Clinician Model: Kendall's Tau: 0.84; Percentage of cases where rank order changes: 7.9%. Clinician Model, w/ Minimum Sample Size: Kendall's Tau: 0.84; Percentage of cases where rank order changes: 8.1%. Group Model: Kendall's Tau: 0.88; Percentage of cases where rank order changes: 5.8%. Group Model, w/ Minimum Sample Size: Kendall's Tau: 0.88; Percentage of cases where rank order changes: 5.9%. Citations: Parast, L., Haas, A., Tolpadi, A., Elliott, M. N., Teno, J., Zaslavsky, A. M., & Price, R. A. (2018). Effects of Caregiver and Decedent Characteristics on CAHPS Hospice Survey Scores. Journal of Pain and Symptom Management, 56(4), 519-529.e511.

Rationale for not using risk adjustment

N/A

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

71

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

56

Meaningful to Patients: Numbers consulted

3535

Meaningful to Patients: Number indicating survey/tool is meaningful

3535

Meaningful to Clinicians: Numbers consulted

28

Meaningful to Clinicians: Number indicating survey/tool is meaningful

28

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

28

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

28

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Construct Validity;Other (enter here):: test-retest

Testing methodology and results

Data Element Reliability: The reliability of the Feeling Heard and Understood multi-data element scale was evaluated using both internal consistency and test-retest reliability coefficients. We used Cronbach's alpha as a measure of internal consistency re

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Intraclass correlation coefficient (ICC)

Signal-to-Noise: Sample size

229

Signal-to-Noise: Statistical result

0.148

Signal-to-Noise: Interpretation of results

To assess the reliability of the quality measure, we used a traditional signal-to-noise analysis that decomposes variability in the measure score into a) between-subject variability and b) within-subject variability. If there is a large amount of between -subject variability (i.e. signal) compared to within-subject variability (i.e., noise), then there is more evidence that it is possible to discriminate performance among palliative care clinicians or groups. To evaluate quality measure reliability for clinician-level reporting, we used hierarchical generalized-linear regression models to relate our outcome measures to our providers and their covariates, where the hierarchy of data is patient observations within individual clinicians. The variance of the model can be decomposed using the adjusted intraclass correlation coefficient (ICC), which provides a summary of the reliability of the measure as tested, with higher values implying more variability between clinicians. Additionally, we incorporate risk adjustment variables into our models to provide fair comparisons among clinicians and

to provide a best effort to ensure that the observed differences among clinicians are truly from differences in performance and not due to baseline differences in risk variables (including survey mode) that represent the clinicians. The reliability from the measure test is then projected out based on observed variances and sample sizes from each clinician, using the Spearman-Brown prophecy formula. This allows us to estimate the required within-clinician sample size to achieve a desired reliability for the measure. Reliability values of approximately 0.7 were a target of an acceptable level of reliability and helped determine required sample sizes (Nunnally, 1978). To evaluate quality measure reliability for group-level reporting, we repeated these analyses, considering clustering of patient observations within palliative care groups. All analyses remain the same as at the clinician level. Group-level testing results are reported in an attachment. We note that while the clinician-level analysis informs on clinician differences and the ability to discern among clinicians, it is distinct from the analysis that compares among groups. That is, these two analyses together do not provide a clean and concise method for comparing individual clinicians against groups of clinicians. The group-level models may be more appropriate for comparing groups, where a group might be an individual clinician and compared against another group of clinicians, all residing within the same ambulatory palliative care program. Further work is needed to identify clinicians that may report measure scores independently and frame the model test as closely as possible to this implementation. Results of Clinician-Level Reliability Testing: The estimate of the adjusted ICC is approximately 0.150 (95% confidence interval: 0.105 to 0.204), and the median adjusted ICC is 0.148. We then extend our reliability results to future samples using the Spearman-Brown prophecy formula, which estimates the average number of patient respondents within clinicians to achieve a desired reliability for a given ICC. We estimate that in order to obtain a nominal reliability of 0.7, an average sample size of 14 patient respondents would be required. However, because each participant is responding to four questions, and those questions are highly correlated (estimated design effect of 3.25 from repeated measures), an average sample size of 12 participants responding to the four data elements would be required when adjusted for this design effect. This suggests a reasonable level of reliability based on the observed between-clinician variability and the within-clinician variability, given our sample sizes. Additionally, we computed estimates of individual clinician specific reliability using a method similar to the approach utilized in Adams (2009). Here, to gain consistency between the approach in Adams and our models, we use our models to estimate a posterior distribution for the overall variability of the risk-adjusted clinician scores and estimate 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. Impact of small sample size: We sought to assess sensitivity to small sample size (i.e., clinicians with low patient volume) in these estimates and address reliability estimates when imposing a minimum sample size requirement. Thus, we removed clinicians with fewer than 12 responses and estimated the measure scores for these providers and calculated their reliabilities again using a method similar to the approach utilized in Adams (2009). The average reliability in the restricted sample was approximately r=0.652, demonstrating that the minimum sample size threshold should provide adequate reliability among providers. Citations: Adams, J. L. (2009). The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/technical_reports/TR653.html. Nunnally, J. C. (1978). Psychometric Theory (2nd ed. ed.). McGraw-Hill.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Pearson's correlations to assess convergent validity

Empiric Validity: Sample size

229

Empiric Validity: Statistical result

0.635

Empiric Validity: Methods and findings

To evaluate the validity of the Feeling Heard and Understood quality measure, we examined the association of the Feeling Heard and Understood 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, with the hypothesis that scores would be positively associated. Associations between the performance measures were evaluated using bivariate correlations. Interpretation of correlations followed standard conventions for small, medium, and large associations (i.e., 0.10, 0.30, 0.50) (Rosnow and Rosenthal, 1989).

Results of validity testing at the performance measure level provide evidence supporting the use of the Feeling Heard and Understood performance measure as constructed. As hypothesized, the Feeling Heard and Understood performance measure 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).

Citations:

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

7

Face Validity: Result

7

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Internal consistency reliability (Cronbach's Alpha), test retest reliability (bivariate correlation), convergent validity. For this PRO-PM, the numerator is a patient-reported outcome collected via patient-reported survey instrument. As requested, data element testing for this patient-reported outcome is described under the Patient-Reported Data section. The denominator (i.e., all patients ages 18 years and older who had an ambulatory palliative care visit) is constructed via a programmed administrative data pull (i.e., no manual medical record or chart review) using standardized and validated codes--ICD-10 Z51.5 (Encounter for Palliative Care) OR Provider Hospice and Palliative Care Specialty Code 17, AND CPT 99201-99205 (New Office Visit), OR CPT 992211-99215 (Established Office Visit), or Place of Service (POS) Code 11 Office. Patient birthdate was pulled administratively, and only adult outpatient palliative care programs were included in the test.

Sample Size

We fielded the survey to 7,595 sampled patients. Of these, 3,356 were not returned, 1,435 were excluded from any analyses due to ineligibility for the larger study, and 2,804 were returned and included in analyses (37% raw response rate; 46% response rate

Statistic Name

Other (enter here):: Test-retest (polychoric correlation coefficient)

Statistical Results

0.85

Interpretation of results

Interpretation of denominator data element testing: As noted previously, the denominator (i.e., all patients ages 18 years and older who had an ambulatory palliative care visit) is constructed via a programmed administrative data pull (i.e., no manual medical record or chart review) using standardized and validated codes. We did conduct a denominator exclusion analysis: We considered five

exclusions from the proposed denominator of all adult patients with an ambulatory palliative care visit: (1) patients who did not complete at least one of the four data elements in the multi-data element measure (n = 26); (2) patients who did not complete and return the patient experience survey within six months of the eligible ambulatory palliative care visit (n = 3,356); (3) patients who responded on the patient experience survey that they did not receive care by the listed ambulatory palliative care provider in the past six months (disavowal; n = 146); (4) patients who were deceased when the survey reached them (n = 748); (5) patients for whom a proxy completed the entire survey on their behalf for any reason (no patient involvement; n = 435). Patients who did not respond to at least one of the four data elements in the multi-data element measure, who did not return the survey at all, who disavowed the program, or who died before the survey could be completed were necessary exclusions because no survey data for the performance measure would be available. We further excluded the small number of patients who did not return the survey within the six-month time frame because of concerns regarding recall bias and because of their likely minimal impact (n = 61 out of 3,356 nonrespondents, or 1.8 percent). We also considered the impact of proxy assistance on measure responses to determine whether to apply exclusions based on proxy assistance. Respondents were categorized into three distinct groups based on proxy assistance, as follows: respondent only (no proxy assistance at all), proxy assisted (proxy helped patient complete the survey, but patient supplied answers, e.g., proxy read questions and wrote down answers), and proxy only (proxy answered all questions, and patient was not involved). We had a total of 2,548 completed surveys by patients without proxy assistance, 224 completed by patients with proxy assistance, and 430 completed by proxies alone with no patient involvement. The mean performance score across these three groups (patient only: mean=0.71, SD=0.37; proxy-assisted: mean=0.77, SD=0.34; proxy only: mean =0.69, SD=0.37) differed significantly ($F(2, 3199) = 3.80$, $p = 0.023$), and follow-up pairwise mean comparisons revealed no difference between patient only and proxy only ($t(581) = 1.22$, $p = 0.22$). However, proxy assisted was significantly different from both patient only ($t(271) = -2.48$, $p = 0.01$) and proxy only ($t(487) = -2.86$, $p = 0.004$). Despite the lack of a significant difference in Feeling Heard and Understood performance measure score means between the proxy-only and patient-only groups, we decided for conceptual reasons to exclude surveys that were completed solely by a proxy with no patient involvement after discussing these results with our project advisory group. Because this is a patient-reported measure of palliative care experience, we wanted to ensure that at least some direct patient report was reflected in the measure response, a rationale for excluding proxy-only responses that was endorsed by the project advisory group. Further, the absence of a significant difference in responses by proxy involvement suggests minimal to no impact of this decision on measure outcomes. Interpretation of numerator data element testing: There was high internal consistency (Cronbach's alpha = 0.90), and test-retest reliability (polychoric correlation coefficient=0.85) for the Feeling Heard and Understood data element that comprises the numerator. Results of validity testing also support the use of the Feeling Heard and Understood data element in the performance measure numerator. As hypothesized, higher scores on the Feeling Heard and Understood scale were associated with higher CAHPS communication scores ($r = 0.54$, $p < 0.001$) and Receiving Desired Help for Pain ($r = 0.48$, $p < 0.001$). Taken together, these results indicate that scores obtained are reliable and support the convergent validity of the Feeling Heard and Understood scale which can, therefore, be used in the construction of the performance measure.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

71.0

Median performance score

73.2

Minimum performance score

42.0

Maximum performance score

90.9

Standard deviation of performance scores

12.1

Does the performance measure use survey or patient-reported data?

Yes

Surveys or patient-reported outcome tools

Patient-reported data is collected via survey instrument (see attachment). The instrument was developed for this measure and is meant to be completed via web survey, on paper or over telephone in English. 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 survey instrument used to collect the data informing the proposed measure will be provided to the Centers for Medicare & Medicaid Services (CMS), to be made available to CMS-approved survey vendors and palliative care clinicians and groups ("groups" refers to palliative care programs). Findings from the national beta field test indicate the feasibility of identifying eligible patients using administrative data and using a survey vendor to support survey administration and data collection. Details of this workflow are described in the attached Measure Information Form. Sampling and data collection could be completed by an authorized survey vendor to minimize bias and reduce workload burden. The survey vendor would be responsible for identifying eligible cases using electronic/automated queries, fielding the survey in the appropriate timeframes, receiving, cleaning, and summarizing survey data for quality improvement (if requested), and submitting a final clinician- or group-level data set to CMS for measure scoring. This last step may include the submission of clinician- or group-level data as well as unadjusted scores to CMS, for risk-adjustment once data are aggregated across clinicians or groups. Use of the survey instrument does not require licenses or fees.

Section 5: Measure Contact Information

Measure Steward

American Academy of Hospice and Palliative Medicine (AAHPM)

Measure Steward Contact Information

Katherine Ast

American Academy of Hospice and Palliative Medicine

8735 West Higgins Road, Suite 300

Chicago, IL 60631

kast@aahpm.org

847-375-4818

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

In qualitative interviews with palliative care groups that participated in the national beta field test, providers were asked about perceived benefits of the proposed Heard and Understood quality measure. Overall, providers were very positive about the Heard and Understood quality measure, noting its central importance to palliative care and the need to measure it. Providers described the value of this quality measure to capture patient experience, stating that this measure encapsulates "the philosophy and soul of palliative care," the "essence" or "mission" of palliative care. Providers also talked about the usefulness and appropriateness of this type of quality measure for palliative care because it does not focus on the patient's medical status or other quality metrics that are not expected in palliative care. They noted that the measure would inform quality improvement efforts to better understand potential gaps in their program and aspects of their care that may impact patients' experiences of feeling heard and understood.

We would also like to make an important point about this proposed measure (MUC2021-087). Although the proposed measure was developed in conjunction with another measure not in-use (MUC2021-092: Ambulatory palliative care patients' experience of receiving desired help for pain), the two measures differ significantly. The Feeling Heard and Understood measure assesses the extent to which patients feel seen and acknowledged by their palliative care provider and team, while the Receiving Desired Help for Pain measure assesses the extent to which patients received the help they wanted for the specific symptom of pain. Although we anticipate that performance on both measures will be informed in part by the overall strength and quality of the patient-provider relationship, the Feeling Heard and

Understood measure is about interpersonal connection while the Receiving Desired Help for Pain measure is about preference-concordant pain management.

Secondary submitter contact information: Sangeeta Ahluwalia, RAND Corporation; 310-393-0411 x7546; sahluwal@rand.org

MUC2022-048 Cardiovascular Disease (CVD) Risk Assessment Measure - Proportion of pregnant/postpartum patients that receive CVD Risk Assessment with a standardized instrument.

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Patients who are assessed for CVD risk via a standardized algorithm. A completed CVD risk assessment will have a calculated risk score and clinician signature (group E). Patients will be assessed at their first contact with the provider for pregnancy-related care (prenatal visit, L&D, postpartum visit) and have repeat assessments if they present symptoms. The measure can be calculated quarterly or annually. See attached word document [CPT-ICD 10 Code Book] for full list of CVD confirmation CPT codes.

Numerator Exclusions

A completed CVD risk assessment will have a calculated risk score and clinician signature. Any patient with an incomplete CVD risk assessment (e.g., no clinician signature) will not be included in the numerator.

Denominator

Pregnant and Postpartum Office Visit assess the CVD risk of patients who are pregnant or postpartum (group B). Any person who is pregnant or postpartum who attends a pregnant or postpartum clinic visit at any participating site should undergo a CVD risk assessment.

Denominator include Patients (a) who have an office visit for prenatal or post-partum care at the intervention site (regardless of gestational age or prior prenatal care at other sites), (b) Any age (including pregnant and postpartum minors), (c) Outpatient OB visit at hospital or in affiliated clinics; Labor and Delivery including private providers contracting with hospital for delivery.

Denominator Exclusions

[a] Patients who have another reason of visit to clinic [not prenatal or postpartum care] and have a positive pregnancy test but have not established the clinic as OB provider (plan to terminate pregnancy or seek prenatal services elsewhere).

[b] Prior history of known cardiac disease. CVD confirmation is identified if the patient has one or more ICD codes in their medical chart during the data abstraction period. If CVD confirmation falls on a date prior to CVD algorithm use with a patient who has a completed algorithm, they are considered an exclusion, and did not require CVD algorithm evaluation. See attached word document [CPT-ICD 10 Code Book] for full list of CVD confirmation CPT-ICD codes.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

OB patients: patients who have an active pregnancy or postpartum episode with at least 1 office visit.

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Obstetrics/gynecology

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record

If applicable, specify the data source

Other:

Description of parts related to these sources

Data collected from EHR:

OB Patient population: patients who have a pregnancy or postpartum episode with at least 1 visit

MRN

Visit Dates

Date of algorithm completion

Race of Mother

Ethnicity of Mother

Date of Birth of Mother

Date of Birth of Infant

Insurance Plan

Clinic site the patient was at when algorithm was completed

Data captured by algorithm: CVD algorithm items [list], algorithm signed by the clinician

Calculated risk outcome of algorithm: at risk, not at risk, possible risk

CVD Testing with dates

Confirmed CVD with dates

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Ambulatory/office-based care;Federally qualified health center (FQHC);Hospital outpatient department (HOD);Other: Labor & Delivery (inpatient)

Multiple Scores

No

What one healthcare domain applies to this measure?

Wellness and Prevention

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

Not Available

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Web interface

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

The data of the CVD risk assessment (standard vital signs, demographic data) are measured and documented in the electronic health record as routine clinical prenatal care. Healthcare providers are expected to do heart and lung examination for the completion of the physical exam and document it in the EHR. The risk assessment has been implemented in electronic health record systems in Epic and Cerner and the score is automatically calculated and provided to the clinician during the clinic visit. The risk assessment can also be calculated manually on a paper document. As the data elements are routine data that clinicians and medical assistants document during prenatal and postpartum visits, the data fields can be consistently located and abstracted from the medical charts.

Method of Measure Calculation

Other digital method;Manual abstraction

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

CVD risk assessment measures were successfully implemented at three large hospital networks in California and Tennessee. The rate for CVD risk assessment in the three hospital networks was 54.4%, 71.5% and 98.7%. Frequency of risk assessment varied widely between the networks and within each network sites based on size, specialty, and setting, ranging from 0% to 100% of patients who had a CVD risk assessment. See attachment [Figures of CVD Risk Assessment Rate by Clinical Sites].

Unintended Consequences

The consistent use of the tool has raised awareness of the importance of CVD risk assessment among obstetricians. Additionally, it has improved patient awareness of the immediate and life-time risk of developing CVD that drives changes in health behavior. Training of clinicians on how to counsel patients about their CVD risk and address potential concerns to avoid negative emotional reactions related to CVD risk with patients. We have not seen any evidence that the follow-up of patients who were deemed high risk for CVD lead to inappropriate use of reSources: .

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system u sed to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

See attachment " Summary of Empirical Data".

Name evidence type

California Maternal Quality Care Collaborative Toolkit , Alliance for Innovation on Maternal Health Cardiac Conditions in Obstetrical Care Bundle

Summarize the evidence

See attachment "Summary of Other Evidence".

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

2,700

Type of Evidence to Support the Measure

Empirical data;Other (enter here):: -California Maternal Quality Care Collaborative Toolkit

-Alliance for Innovation on Maternal Health Cardiac Conditions in Obstetrical Care Bundle

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Other (enter here):: No, it is not risk adjusted, because Black race is one of the variables that contribute towards the risk score.

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

2

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

2

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

132

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

132

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

Yes

If yes, how many sites were evaluated in the provider workflow analysis?

4

Did the provider workflow have to be modified to accommodate the new measure?

Yes

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Signal-to-Noise Reliability

Signal-to-Noise: Sample size

20

Signal-to-Noise: Statistical result

0.985

Signal-to-Noise: Interpretation of results

Signal-to-Noise close to 1 implies that all the variability is attributable to real differences in measure score in different entities.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Pearson Correlation Coefficients

Empiric Validity: Sample size

20

Empiric Validity: Statistical result

0.445

Empiric Validity: Methods and findings

The CVD risk assessment measure and percent of confirmed CVD cases were calculated for 20 entities. We hypothesized them to be positively correlated. Pearson Correlation Coefficient (r) was calculated to test the correlation between measure 1 and % of confirmed CVD cases. The r=0.445 (p-value=0.049) shows that the CVD risk assessment measure and percent of confirmed CVD cases have moderate positive correlation with a statistically significant p-value.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

10

Face Validity: Result

10

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Percent agreement between automated extracted EHR data and manual reviewer

Sample Size

2540

Statistic Name

Kappa

Statistical Results

1.0

Interpretation of results

We are confident to be able to identify prenatal and postpartum patients. We are confident to identify those who have a positive CVD risk assessment score and who are not. We checked the data extracted by the UCI IT department with results from manual review of a subset of charts. We reviewed any discrepancies to adjust the logic until the data pull was completely consistent with the gold standard. The measure could provide meaningful and actionable data on the percentage of patients who received a CVD risk assessment at each clinic site. At each hospital network, we identified low performing and high performing sites.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

61.5

Median performance score

64.8

Minimum performance score

0

Maximum performance score

100

Standard deviation of performance scores

28.2

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

University of California, Irvine

Measure Steward Contact Information

Afshan Hameed

200 S. Manchester Ave Suite 600

Orange, CA 92868

ahameed@hs.uci.edu

(714) 456-7879

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

N/A

MUC2022-052 Adult COVID-19 Vaccination Status

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement 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

Numerator

Patients who have ever completed or reported having ever completed a COVID-19 vaccination series and one booster dose

Numerator Exclusions

N/A

Denominator

All patients aged 18 years and older seen for a visit during the performance period

Denominator Exclusions

Patient received hospice services any time during the performance period

Denominator Exceptions

Complete COVID-19 vaccination series and one booster dose were not administered because patient contraindication documented by clinician

State of development

Field (Beta) Testing

State of Development Details

We have completed empirical reliability testing and face validity testing for this measure, consistent with the CMS Blueprint requirements for testing new measures. We are submitting this measure based on guidance received from CMS that for this MUC list submission cycle a new measure that relies on face validity testing would be eligible for inclusion in the MUC list, even though it is not considered fully developed.

What is the target population of the measure?

All payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Primary care

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Registries

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Wellness and Prevention

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

This measure score and the score from the Total Per Capita Cost (TPCC) cost measure, which measures the overall cost of care delivered to a beneficiary with a focus on the primary care they receive from their provider(s), could be used to assess impacts of the delivery of overall preventive care and wellness services on the cost of overall care. One Improvement Activity, not related to vaccination but treatment of COVID-19, is in MIPS: IA_ERP_3: COVID-19 Clinical Data Reporting with or without Clinical Trial.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

08063

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

Yes

Previous Measure Information

In what prior year was this measure published?

2020

What was the MUC ID for the measure in this year?

MUC20-0045

List the CMS CBE MAP workgroup(s) in this year:

Jan 2021: Rural Health Jan 2021: Clinician Jan 2021: Coordinating Committee

What were the programs that MAP reviewed the measure for in this year?

2020: Merit-based Incentive Payment System-Quality

What was the MAP recommendation in this year?

2020-2021: Conditionally Support

Why was the measure not recommended by the MAP workgroups in this year?

The MAP reviewed this measure in combination with several other de novo COVID-19 vaccination measures applicable to several settings of care. The MAP rationale was the same for all of the COVID-19 vaccination measures: "MAP offered conditional support for rulemaking contingent on CMS bringing the measures back to the MAP once the specifications are further refined, CMS considering an expedited process for the measures for both NQF and CMS, and CMS exploring the inclusion of pediatric hospitals within the COVID measures." The last statement about the inclusion of pediatric hospitals is not relevant for this measure as this measure is focused on the ambulatory setting.

MAP report page number being referenced for this year

2020-2021: Page 25. Link to MAP report: https://www.qualityforum.org/Publications/2021/03/MAP_2020-2021_Considerations_for_Implementing_Measures_Final_Report_-_Clinicians,_Hospitals,_and_PAC-LTC.aspx

What is the history or background for including this measure on the new measures under consideration list?

Measure previously submitted to MAP, refined and resubmitted per MAP recommendation

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

We are aware of another COVID-19 vaccination measure (COVID-19 Vaccination Coverage among Healthcare Personnel) that applies to nine different care settings and has therefore been included in the following nine quality reporting programs:

1. End-Stage Renal Disease Quality Incentive Program (ESRD QIP)
2. Hospital Outpatient Quality Reporting Program (Hospital OQR Program)
3. Hospital Inpatient Quality Reporting Program (Hospital IQR Program)

4. Ambulatory Surgical Center Quality Reporting Program (ASCQR)
5. Inpatient Psychiatric Facility Quality Reporting Program (IPFQR)
6. PPS-Exempt Cancer Hospital Quality Reporting Program (PCHQR)
7. Inpatient Rehabilitation Facility Quality Reporting Program (IRF QRP)
8. Long-Term Care Hospital Quality Reporting Program (LTCH QRP)
9. Skilled Nursing Facility Quality Reporting Program (SNF QRP)

This measure assesses COVID-19 vaccination coverage among healthcare personnel (HCP) for each of the nine care settings. The MIPS measure assesses vaccination rates across a broader population beyond HCP. Additionally, the HCP measure assesses receipt of a COVID-19 vaccination course, while the MIPS measure assesses receipt of a primary vaccination series and one booster dose. The HCP measure does not specify an age range, whereas the MIPS measure we developed includes all adults aged 18 years and older. The HCP measure also excludes HCP with contraindications, whereas the proposed MIPS measure excludes patients in hospice. The proposed MIPS measure also has a denominator exception for contraindications, whereas the HCP measure has no denominator exceptions.

We are also aware of another patient-level measure that is still under development but is aiming to assess whether a patient is up to date on COVID-19 vaccination based on guidance from the CDC. This measure, which assesses vaccination status at time of discharge from a post-acute care setting (e.g., skilled nursing facility), is reported by facilities. As currently specified, this measure has no denominator exclusions. The MIPS measure assesses vaccination rates across a broader population beyond patients discharged from a post-acute care setting. The patient-level measure does not specify an age range, whereas the MIPS measure we developed includes all adults aged 18 years and older. The patient-level measure also has no denominator exclusions, whereas the proposed MIPS measure excludes patients in hospice. The proposed MIPS measure also has a denominator exception for contraindications, whereas the patient-level measure has no denominator exceptions.

How will this measure be distinguished from other similar and/or competing measures?

As discussed above, the proposed MIPS COVID-19 vaccination measure differs from the HCP and patient-level post-acute care setting measures in its specification (i.e., eligible population, age range, exclusions, denominator exceptions, and numerator). It also differs in that it 1) assesses clinician performance (not health care settings or patients at discharge from post-acute care) and 2) assesses how frequently clinicians (not post-acute care settings) are vaccinating their patients (not HCP).

How will this measure add value to the CMS program?

The MIPS COVID-19 vaccination measure is proposed for a unique CMS program and will therefore assess performance of another accountable entity (i.e., eligible clinicians). It will also assess vaccination rates across a broader patient population (i.e., all adults 18 years and older).

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

We recruited two sets of stakeholders to participate in feasibility testing. We recruited four clinical sites (two primary care private clinician practices, one primary care site affiliated with a university health system, and one Federally Qualified Health Center Look-Alike [FQLA]), all of which administer the vaccine, and we recruited four registry vendors with experience reporting to MIPS. We collected data from each test site through a WebEx meeting, conducting one interview per test site. Through the workflow assessments, we asked how stakeholders engaged patients and documented receipt (including patient self-reported receipt) of vaccinations, and we determined if offering vaccinations is included in existing workflows. We focused mainly on the workflows related to the numerator criteria: patients who completed or reported having ever completed a COVID-19 vaccination series. We developed a stakeholder-specific protocol to facilitate the completion of the workflow assessments. Specifically, we gathered information on data elements required for registry reporting. For each data element, we assessed the following: (1) availability in a structured field in the clinical or billing EHR module or in free-text notes; (2) accuracy; (3) extent to which data are coded using standard terminologies; and (4) extent to which data are collected during routine care. In our interviews with registry vendors, we asked both about the feasibility of obtaining the required data from clinical sites and of storing and reporting those data to MIPS. Through our interviews, we found that all four clinical sites collected the data in the EHR in structured fields to report the measure under current workflows. One caveat to this finding is that in practice, when we requested patient-level data from our beta testing sites, one of our three beta testing sites (not included in the feasibility assessment) identified instances of COVID-19 vaccination through a combination of data from the EHR in structured fields and data obtained through a manual review of patient charts. As noted above, all four of these sites offered the vaccine directly at the time of interview. The sites reported capturing data on patient vaccination from a range of data Sources: , including their own vaccination administration data (for vaccines they delivered directly to their patients), patient self-reports, vaccination data obtained from state immunization registries, and faxed data from pharmacy chains on vaccinations delivered to the site's patients by the pharmacies. In all cases, the sites noted they take the data from these disparate sources and enter them into the EHR. All registry sites noted they could support measure submission if the measure were implemented in MIPS. However, one registry that serves small and rural practices noted that its clients, many of whom report on paper, do not collect data on COVID-19 vaccination and would face challenges in doing so. Note that we conducted our feasibility assessment interviews between August and November 2021, as booster guidance was evolving, and at that point we asked solely about the feasibility of implementing a version of the measure focused on patients completing a full course of the vaccination series, without booster doses. However, we do not expect that asking sites about boosters would have changed our feasibility findings, since capturing the boosters would simply involve tracking an additional dose of the vaccines.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

According to the CDC, as of May 18, 2022, 582,757,136 vaccine doses have been administered, 89.1% of the total U.S. adult population has received at least one vaccine dose, 76.4% are fully vaccinated, and 50.0% have received at least one booster dose (CDC COVID Data Tracker: Vaccinations in the US). In general, states in the Northeast, Northwest, and Upper Midwest are vaccinating at higher rates than other regions of the United States.

Among the 33 clinicians in our testing sample with sufficient denominator size, the minimum measure score was 0 percent, the 10th percentile of measure scores was 16.8 percent, the 25th percentile was 28.3 percent, the 75th percentile was 58.0 percent, the 90th percentile was 63.9 percent, and the maximum was 77.8 percent, indicating there is room for improvement.

Unintended Consequences

Clinicians may administer the vaccine to patients with contraindications in an attempt to achieve higher performance rates if the exclusions and exceptions are not broad enough. Clinicians may also choose their patients based on vaccination status. Patients may utilize performance scores in a way that's harmful to clinicians. For instance, if a patient is against vaccination and their clinician has a high patient vaccination rate, the patient could refuse care. Measure scores may reflect practices' ability to hunt down vaccination data rather the quality of care provided by their clinicians. Finally, the measure does not focus on disparities. If certain patient groups are more or less likely to be vaccinated, clinicians who treat those patients may be penalized or rewarded in performance.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

12

Outline the clinical guidelines supporting this measure

ACIP guidelines support the measure, all of which are evidence-based. All ACIP recommendations can be referenced at the following webpage: https://www.cdc.gov/vaccines/hcp/acip-recs/vacc-specific/covid-19.html.

The Centers for Disease Control and Prevention's (CDC) Interim COVID-19 Immunization Schedule, based on ACIP recommendations, can be referenced at the following webpage: https://www.cdc.gov/vaccines/covid-19/clinical-considerations/interim-considerations-us.html.

1. 1. On December 12, 2020, ACIP issued an interim recommendation for use of Pfizer-BioNTech COVID-19 vaccine in persons aged 16 years and older.
2. 2. On December 19, 2020, ACIP issued an interim recommendation for use of the Moderna COVID-19 vaccine in persons aged 18 years and older.
3. 3. On February 28, 2021, ACIP issued an interim recommendation for use of the Janssen COVID-19 vaccine in persons aged 18 years and older.

4. On May 12, 2021, ACIP issued an interim recommendation for use of the Pfizer-BioNTech COVID-19 vaccine in adolescents aged 12-15 years.
5. On August 30, 2021, ACIP revised its interim recommendation to a standard recommendation for use of the Pfizer-BioNTech COVID-19 vaccine in persons aged 16 years and older.
6. On October 21, 2021, ACIP recommended persons aged 18 years and older receive a booster dose at least two months after receiving Janssen COVID-19 vaccination (CDC, 2021b).
7. On November 2, 2021, ACIP issued an interim recommendation for use of the Pfizer-BioNTech COVID-19 vaccine in children aged 5-11 years.
8. On November 19, 2021, ACIP recommended persons aged 12 years and older receive a Pfizer-BioNTech booster dose at least five months after completing the primary COVID-19 vaccination series (CDC, 2021a).
9. On November 19, 2021, ACIP recommended persons aged 18 years and older receive a Moderna booster dose at least five months after completing the primary COVID-19 vaccination series (CDC, 2021a).
10. On December 16, 2021, ACIP made a preferential recommendation for the use of mRNA COVID-19 vaccines over the Janssen adenoviral-vectored COVID-19 vaccine in all persons aged 18 years and older.
11. On January 4, 2022, ACIP recommended children 5-11 years who are moderately or severely high-risk receive a Pfizer-BioNTech booster dose at least one month after completing the primary COVID-19 vaccination series (CDC, 2022).
12. On February 4, 2022, ACIP issued a standard recommendation for use of the Moderna COVID-19 vaccine in persons aged 18 years and older.

CDC, 2021a. CDC Expands Eligibility for COVID-19 Booster Shots to All Adults. https://www.cdc.gov/media/releases/2021/s1119-booster-shots.html (Accessed May 17, 2022).

CDC, 2021b. CDC Expands Eligibility for COVID-19 Booster Shots. https://www.cdc.gov/media/releases/2021/p1021-covid-booster.html (Accessed May 17, 2022).

CDC, 2022. CDC Recommends Pfizer Booster at 5 Months, Additional Primary Dose for Certain Immunocompromised Children. https://www.cdc.gov/media/releases/2022/s0104-Pfizer-Booster.html (Accessed May 17, 2022).

Name the guideline developer/entity

Advisory Committee on Immunization Practices (ACIP). This primary guideline reflects updated recommendations related to the age range (18+) currently specified in this measure.

Publication year

2022

Full citation +/- URL

Wallace M, Moulia D, Blain AE, et al. The Advisory Committee on Immunization Practices' Recommendation for Use of Moderna COVID-19 Vaccine in Adults Aged [Greater Than or Equal to] 18 Years and Considerations for Extended Intervals for Administration of Primary Series Doses of mRNA COVID-19 Vaccines - United States, February 2022. MMWR Morb Mortal Wkly Rep 2022;71:416-421. DOI: http://dx.doi.org/10.15585/mmwr.mm7111a4

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

On February 4, 2022, ACIP issued a standard recommendation for use of the Moderna COVID-19 vaccine in persons aged 18 years and older.

What evidence grading system did the guideline use to describe strength of recommendation?

Modified GRADE

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

Type 1 = high certainty

Type 2 = moderate certainty

Type 3 = low certainty

Type 4 = very low certainty

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

Other (enter here):: Type 1

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

Modified GRADE

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

Type 1 = high certainty

Type 2 = moderate certainty

Type 3 = low certainty

Type 4 = very low certainty

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

High or similar

List the guideline statement that most closely aligns with the measure concept.

On February 4, 2022, ACIP issued a standard recommendation for use of the Moderna COVID-19 vaccine in persons aged 18 years and older.

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

206,000,000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Other (enter here):: The Adult COVID-19 Vaccination Status measure is a process measure that does not need risk adjustment because the measured process is appropriate for all patients included in the denominator and the measure excludes all the patients f

Cost estimate completed

Yes

Cost estimate methods and results

No assumptions were made for the data cited. All data points were shared explicitly in the Sources: listed below. The studies cited were published between 2020 and 2021 and included 2020-2021 data. Therefore, they used 2020 dollars.

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

9

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

6

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

9

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

5

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

Yes

If yes, how many sites were evaluated in the provider workflow analysis?

4

Did the provider workflow have to be modified to accommodate the new measure?

No

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

The signal-to-noise statistic, R, summarizes the proportion of variation between clinician scores on a measure that is due to real differences in underlying characteristics (such as differences in medical care), rather than to background-level or random variation (for example, due to measurement or sampling error). We estimated signal-to-noise reliability for the measure using the beta-binomial method.

Signal-to-Noise: Sample size

33

Signal-to-Noise: Statistical result

0.986

Signal-to-Noise: Interpretation of results

Reliability was high. The 10th percentile reliability score was 0.849, the 25th percentile was 0.947, the median was 0.986, the 75th percentile was 0.994, and the 90th percentile was 0.997. Reliability coefficients above 0.70 are considered sufficient to draw conclusions about groups, and values above 0.9 are considered sufficient to draw conclusions about individuals (Adams, 2009). Citation: Adams, J. L. The reliability of provider profiling: A tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009. Available at https://www.rand.org/pubs/technical_reports/TR653.html. Accessed May 3, 2022.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

No

Empiric Validity: Statistic name

N/A

Empiric Validity: Sample size

N/A

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

N/A

Empiric Validity: Interpretation of results

N/A

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

6

Face Validity: Result

3

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

44.3

Median performance score

52.2

Minimum performance score

0

Maximum performance score

77.8

Standard deviation of performance scores

19.8

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare and Medicaid Services

Measure Steward Contact Information

Joel Andress
7500 Security Boulevard
Mailstop S3-10-26
Baltimore, MD 21244
Joel.Andress@cms.hhs.gov
(410) 786-5237

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Pam Lighter
1100 13th St NW
Third Floor
Washington, DC 20005
lighter@ncqa.org
(202) 768-8051

Secondary Submitter Contact Information

Christine Holland
1110 First St NE #1200
Washington, DC 20002
cholland@mathematica-mpr.com
(202) 484-5271

Submitter Comments

We conducted face validity testing through clinician interviews, and as noted in row 46 a total of 3 out of 6 clinicians interviewed indicated the measure could distinguish good from poor quality of care. As this measure is a CQM, we did not test the measure for data element validity because the gold standard we would normally use to test for data element validity, a manual review of the patient chart, is one of the data Sources: that clinicians may draw on to report the measure. We were unable to conduct empiric validity testing of the measure score because COVID-19 vaccination is new enough that there is limited published evidence on process or outcome measures obtainable from clinical sites that we could try to correlate with performance on the vaccination measure. We based our validity testing approach on the guidance from the CMS Measures Management System Blueprint, which notes that face validity testing may be sufficient for new measures and that the National Quality Forum also allows new measures to be tested through face validity (6.2.2.2.1).

For additional context on the face validity of the measure, in addition to the nine clinicians who provided input on measure usability, one non-clinical expert provided their feedback. This expert, who represents a federal entity and has experience in immunization policy and quality measurement, agreed that the measure would improve quality of care. Additionally, of the nine clinicians, two did not explicitly agree the measure would improve quality of care; however, they both provided conditional support for the measure given the state of the pandemic and agreed with submission of the measure to the 2022 Measures Under Consideration list.

MUC2022-060 First Year Standardized Waitlist Ratio (FYSWR)

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Number of patients in the practitioner group listed on the kidney or kidney-pancreas transplant waitlist or who received living donor transplants within the first year following initiation of dialysis.

Numerator Exclusions

N/A

Denominator

The denominator for the FYSWR is the expected number of waitlist or living donor transplant events in the practitioner group according to each patien's treatment history for patients within the first year following initiation of dialysis, adjusted for age, incident comorbidities, dual Medicare-Medicaid eligibility, Area Deprivation Index (from patient's residence zip code) and transplant center characteristics, among patients under 75 years of age who were not already waitlisted and did not have kidney transplantation prior to the initiation of ESRD dialysis.

Denominator Exclusions

Patients who were at age 75 or older on their initiation of dialysis date are excluded. Patients who were admitted to a skilled nursing home facility (SNF) or a hospice during the month of evaluation were excluded. These exclusions represent conditions for which transplant waitlist candidacy is highly unlikely, and which can be identified readily with available data. Patients were also excluded if waitlisted or transplanted prior to initiation of first dialysis. Patients who were attributed to dialysis practitioner groups with fewer than 11 patients or 2 expected events are not excluded from the measure. All patients who meet the denominator inclusion criteria are included and used to model a given dialysis practitioner group's expected waitlist rate. If a dialysis practitioner group has fewer than 11 patients or 2 expected events, then the dialysis practitioner group is excluded from reporting outcomes. The Nursing Home Minimum Dataset (MDS) and the Questions 17u and 22 on CMS Medical Evidence Form 2728 were used to identify patients in skilled nursing facilities. For hospice patients, a separate CMS file that contains final action claims submitted by Hospice providers was used to determine the hospice status. Patients are excluded if they are nursing home patients according to their Medical Evidence Form 2728

or according to the Minimum Dataset (MDS) data on their initiation of dialysis date. Patients with Medicare Hospice claims on their initiation of dialysis date are also excluded. Patients that were on the kidney or kidney-pancreas waitlist or had a transplant prior to initiation of dialysis were excluded.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Nephrology

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Claims Data;Registries

If applicable, specify the data source

N/A

Description of parts related to these sources

EQRS (formerly CROWNWeb), Medicare Claims, and the CMS Medical Evidence Form 2728 were used as the data Source: for establishing the denominator. CMS Medical Evidence Form 2728 was used for the age risk adjustment and exclusion of patients age 75 or older, and comorbidity condition adjustments. Organ Procurement and Transplant Network (OPTN) is the data source for the numerator (waitlisting or living donor kidney transplantation). Medicare claims were used for the hospice exclusion criteria. The Nursing Home Minimum Dataset and Questions 16u and 21 on the CMS Medical Evidence Form were used to identify SNF patients. Additionally, Medicare claims and a payment history file were used to determine dual eligibility status. The Medicare Provider Files from the CMS Integrated Data Repository (IDR) were used to identify practitioners group practice. Area Deprivation Index (ADI) was obtained from Census data (2011-2015) based on patient zip code. In order to assess the transplant center characteristics, Scientific Registry of Transplant Recipients (SRTR) data was used.

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Dialysis facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Chronic Conditions

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

Standardized First Kidney Transplant Waitlist Ratio for Incident Dialysis Patients (SWR), Centers for Medicare and Medicaid Services; Percentage of Prevalent Patients Waitlisted (PPPW), Centers for Medicare and Medicaid Services

How will this measure be distinguished from other similar and/or competing measures?

The measure provides individual practitioner and group practitioner accountability. See the Measure Information and Justification Forms for detailed information about the measure.

How will this measure add value to the CMS program?

The measure provides individual practitioner and group practitioner accountability. See the Measure Information and Justification Forms for detailed information about the measure.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

Data used in the measure are generated or collected by and used by healthcare personnel during provision of care (e.g., blood pressure, laboratory value, diagnosis, depression score) and coded by someone other than the person obtaining original information (e.g., Diagnosis-Related Group [DRG], International Classification of Diseases, 10th Revision, Clinical Modification/Procedure Coding System [ICD-10-CM/PCS] codes on claims). All data elements are in defined fields in a combination of electronic Sources: .

Method of Measure Calculation

Claims;Other digital method;Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Claims;Hybrid: Other digital method

Evidence of Performance Gap

The performance score information reported in this submission (mean, median, standard deviation) demonstrate opportunity for improvement.

Unintended Consequences

None anticipated.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

Other

Summarize the evidence

National or large regional studies provide strong empirical support for the association between processes under dialysis practitioner control and subsequent waitlisting. In one large regional study conducted on facilities in the state of Georgia, a standardized dialysis facility referral ratio was developed, adjusted for age, demographics and comorbidities (Paul S. et al, Clin J Am Soc Nephrol 2018;13:282-289). There was substantial variability across dialysis facilities in referral rates, and a Spearman correlation performed between ranking on the referral ratio and dialysis facility waitlist rates was highly significant ($r=0.35$, $p<0.001$). A national study using registry data (United States Renal Data System) from 2005-2007 examined the association between whether patients were informed about kidney transplantation (based on reporting on the Medical Evidence Form 2728) and subsequent access to kidney transplantation (waitlisting or receipt of a live donor transplant) (Kucirka LM et al. Am J Transplant 2012;12:351-357). Approximately 30% of patients were uninformed about kidney transplantation, and this was associated with half the rate of access to transplantation compared to patients who were informed. In a related survey study of 388 hemodialysis patients, whether provision of information about transplantation by nephrologists or dialysis staff occurred was directly confirmed with patients (Salter ML et al, J Am Soc Nephrol 2014;25:2871-2877). Patient report of provision of such information was associated with a three-fold increase in likelihood of waitlisting. Finally, a large survey study of 170 dialysis facilities in the Heartland Kidney Network (Iowa, Kansas, Missouri and Nebraska) was conducted to examine transplant education practices (Waterman AD et al, Clin J Am Soc Nephrol 2015;10:1617-1625). Facilities employing multiple (>3) transplant education strategies (e.g. provision of brochures, referral to formal transplant education program, distribution of transplant center contact information) had 36% higher waitlist rates compared to facilities employing fewer strategies.

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

281,479

Type of Evidence to Support the Measure

Other (enter here):: Two previous Technical Expert Panels (TEP) have been convened to discuss potential measures directed at improving access to kidney transplantation, in 2015 and most recently, in 2021 (2015 TEP Report: https://dialysisdata.org/sites/default/files/content/ESRD_Measures/Access_To_Kidney_Transplantation_TEP_Summary_Report.pdf ; 2021 TEP Report: https://dialysisdata.org/content/esrd-measures, please see Practitioner Level Measurement of Effective Access to Kidney Transplantation under Ongoing Technical Expert Panels section). Both were comprised of relevant stakeholders, including dialysis nephrologists, transplant nephrologists, transplant surgeons, social workers, researchers, and notably, patient representatives with a history of end-stage kidney disease. Discussions during both TEPs revealed broad support for the importance of waitlisting, and formal voting demonstrated a majority of TEP members were in favor of the development of quality measures targeting waitlisting (at the dialysis facility level for the 2015 TEP, and the practitioner level for the 2021 TEP).

In addition to the above, empirical support for the value of waitlisting to patients comes from a published study reporting on a large survey of 409 patients or family members who agreed to receiving

emails from the National Kidney Foundation (Husain S.A. et al, Am. J. Transplant 2018;18(11):2781-2790). Participants included both patients with advanced chronic kidney disease prior to transplant, and recipients of transplants, who were asked about their priorities in choice of a transplant center. Notably, participants were most likely (a plurality of participants) to rank waitlisting characteristics (such as ease of getting on the waitlist) as the most important feature, in contrast to other transplant center characteristics such as post-transplant outcomes and practical considerations (e.g. distance to center).

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Proxy social risk factors;Other (enter here):: Weighted SRTR mortality ratio; Weighted SRTR transplant ratio

Patient-level demographics: please select all that apply:

Age

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Comorbidities

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

Dual Eligibility for Medicare and Medicaid;Other (enter here):: Area Deprivation Index (ADI)

Patient community characteristic: please select all that apply:

N/A

Risk model performance

The C-statistic (also known as the Index of Concordance) was 0.75, meaning that the model correctly ordered 75% of the pairs of patient-months that were discordant with respect to the response variate.

Rationale for not using risk adjustment

N/A

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Inter-unit reliability (IUR)

Signal-to-Noise: Sample size

458,676

Signal-to-Noise: Statistical result

0.42

Signal-to-Noise: Interpretation of results

The value of IUR indicates that about 42% of the variation in the FYSWR measure can be attributed to the between-dialysis practitioner differences (signal) and about 58% of variation to within-dialysis practitioner variation (noise). The value of IUR implies a moderate degree of reliability.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Spearman correlation

Empiric Validity: Sample size

458,676

Empiric Validity: Statistical result

0.12

Empiric Validity: Methods and findings

Validity of the measure was tested by evaluating the association between the dialysis practitioner level measure performance, and subsequent mortality and transplant rates among all patients attributed to the practitioners. We hypothesized that practitioners with higher performance on the FYSWR measure would have subsequently higher transplant rates among their patients. This would be expected to follow from activities these practitioners conducted to improve the health and therefore suitability of their patients for transplant candidacy. Along similar lines, we hypothesized that practitioners with higher performance on the FYSWR measure would demonstrate lower subsequent mortality among their patients. However, we expected this to be a more modest association given the many other factors that can affect mortality within the dialysis population.

To evaluate the associations, we first divided dialysis practitioners into 3 tertiles (T1 to T3) based on their performance on the FYSWR (T1 to T3, from highest to lowest waitlisting). Tertiles were chosen in order to evaluate a gradient in effect, but still maintain sufficient numbers within each group for statistical precision. We then computed the corresponding second year mortality rate and transplant rate among patients assigned to each practitioner. We then applied the Cochran-Armitage trend test to evaluate the relationship between the tertile grouping and these practitioner-level outcomes. Finally, we examined the Spearman correlations between FYSWR and the second year mortality rate or second year transplant rate.

See the attached Measure Justification Form for the rest of the methods and findings.

Empiric Validity: Interpretation of results

Yes

Face Validity

No

Face Validity: Number of voting experts and patients/caregivers

N/A

Face Validity: Result

N/A

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Ratio

Measure Performance Score Interpretation

Higher score is better

Mean performance score

1.00

Median performance score

0.88

Minimum performance score

0.00

Maximum performance score

5.32

Standard deviation of performance scores

0.70

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare and Medicaid Services (CMS)

Measure Steward Contact Information

Golden Horten
7500 Security Blvd
Baltimore, MD 21244
golden.horten@cms.hhs.gov
(410) 786-4024

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Alexander Yaldo
1415 Washington Heights, Suite 33645 SPH I
Ann Arbor, MI 48109-2029
yaldo@med.umich.edu
(734) 936-5711

Secondary Submitter Contact Information

Jennifer Sardone
1415 Washington Heights, Suite 33645 SPH I
Ann Arbor, MI 48109-2029
jmsto@med.umich.edu
(734) 936-5711

Submitter Comments

For the estimated impact of the measure, 281,479 is the number of patients included in the measure calculations in the testing form. The MJF should be referenced for greater detail about validity testing and risk adjustment.

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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).

Numerator

PPPW: The numerator is the adjusted count of patient months in which the patient at the dialysis practitioner or practitioner group practice is on the kidney or kidney-pancreas transplant waitlist as of the last day of each month during the reporting year.

aPPPW: The numerator is the adjusted count of patient months in which the patient at the dialysis practitioner or practitioner group practice is on the kidney or kidney-pancreas transplant waitlist in an active status as of the last day of each month during the reporting year.

Numerator Details

PPPW: The adjusted count of patient months in which the patient at the dialysis practitioner or practitioner group practice is on the kidney or kidney-pancreas transplant waitlist, adjusted for patient-mix. To be included in the numerator for a particular month, the patient must be on the kidney or kidney-pancreas transplant waitlist as of the last day of the month during the reporting year.

aPPPW: The adjusted count of patient months in which the patient at the dialysis practitioner or practitioner group practice is active on the kidney or kidney-pancreas transplant waitlist, adjusted for patient-mix. To be included in the numerator for a particular month, the patient must be active on the kidney or kidney-pancreas transplant waitlist as of the last day of the month during the reporting year.

Numerator Exclusions

N/A

Denominator

All patient-months for patients who are under the age of 75 in the reporting month and who are assigned to a dialysis practitioner or practitioner group practice according to each patients treatment history during a given month during the reporting year.

Denominator Exclusions

Exclusion that are implicit in the denominator include:

Patients who were at age 75 or older in the reporting month

Patients who were admitted to a skilled nursing facility (SNF) during the month of evaluation were excluded from that month;

Patients who were admitted to a skilled nursing facility (SNF) within one year of dialysis initiation according to form CMS-2728

Patients determined to be in hospice were excluded from month of evaluation and the remainder of reporting period

Patients with dementia

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Fee for Service

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Nephrology

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Claims Data;Registries

If applicable, specify the data source

N/A

Description of parts related to these sources

EQRS (formerly CROWNWeb), Medicare Claims, and the CMS Medical Evidence Form 2728 were used as the data Sources: for establishing the denominator. EQRS was used for the age risk adjustment and exclusion of patients aged 75 or older. Organ Procurement and Transplant Network (OPTN) is the data source for the numerator (waitlisting in active status). Medicare claims from the year prior to the reporting period were used for comorbidity condition adjustments. Medicare claims during the reporting period were used for the hospice exclusion criteria. The Nursing Home Minimum Dataset and Questions 17u and 22 on the CMS Medical Evidence Form were used to identify SNF patients. Additionally, Medicare claims during the reporting period and a payment history file were used to determine dual eligibility status. The Medicare Provider Files from the CMS Integrated Data Repository (IDR) were used to identify dialysis practitioner's group practice. Area Deprivation Index (ADI) was obtained from Census data (2011-2015) based on patient zip code. In order to assess the transplant center characteristics, Scientific Registry of Transplant Recipients (SRTR) data was used.

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Dialysis facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Chronic Conditions

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

Standardized First Kidney Transplant Waitlist Ratio for Incident Dialysis Patients (SWR), Centers for Medicare and Medicaid Services

Percentage of Prevalent Patients Waitlisted (PPPW), Centers for Medicare and Medicaid Services

How will this measure be distinguished from other similar and/or competing measures?

The measure provides individual practitioner and group practitioner accountability. See the Measure Information and Justification Forms for detailed information about the measure.

How will this measure add value to the CMS program?

The measure provides individual practitioner and group practitioner accountability. See the Measure Information and Justification Forms for detailed information about the measure.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

Data used in the measure are generated or collected by and used by healthcare personnel during provision of care (e.g., blood pressure, laboratory value, diagnosis, depression score) and coded by someone other than the person obtaining original information (e.g., Diagnosis-Related Group [DRG], International Classification of Diseases, 10th Revision, Clinical Modification/Procedure Coding System [ICD-10-CM/PCS] codes on claims). All data elements are in defined fields in a combination of electronic Sources: .

Method of Measure Calculation

Claims;Other digital method;Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Claims;Hybrid: Other digital method

Evidence of Performance Gap

The performance score information reported in this submission (mean, median, standard deviation) demonstrate opportunity for improvement.

Unintended Consequences

None anticipated.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

Other

Summarize the evidence

National or large regional studies provide strong empirical support for the association between processes under dialysis practitioner control and subsequent waitlisting. In one large regional study conducted on facilities in the state of Georgia, a standardized dialysis facility referral ratio was developed, adjusted for age, demographics and comorbidities (Paul S. et al, Clin J Am Soc Nephrol 2018;13:282-289). There was substantial variability across dialysis facilities in referral rates, and a Spearman correlation performed between ranking on the referral ratio and dialysis facility waitlist rates was highly significant ($r=0.35$, $p<0.001$). A national study using registry data (United States Renal Data System) from 2005-2007 examined the association between whether patients were informed about kidney transplantation (based on reporting on the Medical Evidence Form 2728) and subsequent access to kidney transplantation (waitlisting or receipt of a live donor transplant) (Kucirka LM et al. Am J Transplant 2012;12:351-357). Approximately 30% of patients were uninformed about kidney transplantation, and this was associated with half the rate of access to transplantation compared to patients who were informed. In a related survey study of 388 hemodialysis patients, whether provision of information about transplantation by nephrologists or dialysis staff occurred was directly confirmed with patients (Salter ML et al, J Am Soc Nephrol 2014;25:2871-2877). Patient report of provision of such information was associated with a three-fold increase in likelihood of waitlisting. Finally, a large survey study of 170 dialysis facilities in the Heartland Kidney Network (Iowa, Kansas, Missouri and Nebraska) was conducted to examine transplant education practices (Waterman AD et al, Clin J Am Soc Nephrol 2015;10:1617-1625). Facilities employing multiple (>3) transplant education strategies (e.g. provision of brochures, referral to formal transplant education program, distribution of transplant center contact information) had 36% higher waitlist rates compared to facilities employing fewer strategies.

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

2,541,461

Type of Evidence to Support the Measure

Other (enter here):: Two previous Technical Expert Panels (TEP) have been convened to discuss potential measures directed at improving access to kidney transplantation, in 2015 and most recently, in 2021 (2015 TEP Report: https://dialysisdata.org/sites/default/files/content/ESRD_Measures/Access_To_Kidney_Transplantation_TEP_Summary_Report.pdf; 2021 TEP Report: https://dialysisdata.org/content/esrd-measures, please see Practitioner Level Measurement of Effective Access to Kidney Transplantation under Ongoing Technical Expert Panels section). Both were comprised of relevant stakeholders, including dialysis nephrologists, transplant nephrologists, transplant surgeons, social workers, researchers, and notably, patient representatives with a history of end-stage kidney disease. Discussions during both TEPs revealed broad support for the importance of waitlisting, and formal voting demonstrated a majority of TEP members were in favor of the development of quality measures targeting waitlisting (at the dialysis facility level for the 2015 TEP, and the practitioner level for the 2021 TEP).

In addition to the above, empirical support for the value of waitlisting to patients comes from a published study reporting on a large survey of 409 patients or family members who agreed to receiving emails from the National Kidney Foundation (Husain S.A. et al, Am. J. Transplant 2018;18(11):2781-2790). Participants include both patients with advanced chronic kidney disease prior to transplant, and recipients of transplants, and were asked about their priorities in choice of a transplant center. Notably, participants were most likely (a plurality of participants) to rank waitlisting characteristics (such as ease of getting on the waitlist) as the most important feature, in contrast to other transplant center characteristics such as post-transplant outcomes and practical considerations (e.g. distance to center).

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions;Proxy social risk factors;Other (enter here):: Weighted transplant center waitlist mortality ratio; Weighted transplant center transplant rate ratio

Patient-level demographics: please select all that apply:

Age;Sex

Patient-level health status & clinical conditions: please select all that apply:

Case-Mix Adjustment;Comorbidities

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

Dual Eligibility for Medicare and Medicaid;Other (enter here):: Area Deprivation Index (ADI)

Patient community characteristic: please select all that apply:

N/A

Risk model performance

PPPW Individual Practitioner: The C-statistic (also known as the Index of Concordance) was 0.769, meaning that the model correctly ordered 76.9% of the pairs of patient-months that were discordant with respect to the response variate. Month-specific C-statistics were computed in order to identify any trends by month in the model's discriminatory ability. PPPW Practitioner Group: The C-statistic (also known as the Index of Concordance) was 0.753, meaning that the model correctly ordered 75.3% of the pairs of patient-months that were discordant with respect to the response variate. Month-specific C-statistics were computed in order to identify any trends by month in the model's discriminatory ability. aPPPW Individual Practitioner: The C-statistic (also known as the Index of Concordance) was 0.783, meaning that the model correctly ordered 78.3% of the pairs of patient-months that were discordant with respect to the response variate. Month-specific C-statistics were computed in order to identify any trends by month in the model's discriminatory ability. aPPPW Practitioner Group: The C-statistic (also known as the Index of Concordance) was 0.763, meaning that the model correctly ordered 76.3% of the pairs of patient-months that were discordant with respect to the response variate. Month-specific C-statistics were computed in order to identify any trends by month in the model's discriminatory ability.

Rationale for not using risk adjustment

N/A

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Inter-unit reliability (IUR)

Signal-to-Noise: Sample size

8,331

Signal-to-Noise: Statistical result

0.89

Signal-to-Noise: Interpretation of results

PPPW Individual Practitioner: The IUR is 0.89. Dialysis practitioners with <11 eligible patients were excluded from this calculation. The value of IUR indicates that about 89% of the variation in the PPPW measure can be attributed to the between-dialysis practitioner differences (signal) and about 11% of variation to within-dialysis practitioner variation (noise). The value of IUR implies a high degree of reliability.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Spearman correlation

Empiric Validity: Sample size

8,331

Empiric Validity: Statistical result

0.248

Empiric Validity: Methods and findings

Validity of the measure was tested by evaluating the association between the dialysis practitioner level measure performance, and mortality and overall transplant rates among all patients attributed to the dialysis practitioners. We hypothesized that dialysis practitioners with higher performance on the PPPW measure would have higher transplant rates among their patients. This would be expected to follow from activities these dialysis practitioners conducted to improve the health and therefore suitability of their patients for transplant candidacy. Along similar lines, we hypothesized that dialysis practitioners with higher performance on the PPPW measure would demonstrate lower mortality among their patients. However, we expected this to be a more modest association given the many other factors that can affect mortality within the dialysis population.

To evaluate the associations, we first divided dialysis practitioners, into 3 tertiles (T1 to T3) based on their performance on the PPPW (T1 to T3, from highest to lowest waitlisting). Tertiles were chosen in order to evaluate a gradient in effect, but still maintain sufficient numbers within each group for statistical precision. We then computed the corresponding mortality rate and transplant rate among patients assigned to each dialysis practitioner in 2019. We then applied the Cochran-Armitage trend test to evaluate the relationship between the tertile grouping and these dialysis practitioner-level outcomes. Finally, we examined the Spearman correlation between the dialysis practitioner measure value and each of the outcomes respectively.

See attached Measure Justification Form for the rest of the information regarding the validity methods and findings.

Empiric Validity: Interpretation of results

Yes

Face Validity

No

Face Validity: Number of voting experts and patients/caregivers

N/A

Face Validity: Result

N/A

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between two manual reviewers

Sample Size

8,331

Statistic Name

Other (enter here):: Spearman correlation coefficient

Statistical Results

-0.074

Interpretation of results

PPPW Individual Practitioner:

As expected, higher PPPW performance correlated with higher transplant rate, with clear separation of transplant rates across dialysis practitioner tertiles of performance. The direction of the relationship with mortality was also as expected, and statistically significant, with numerically lower mortality with higher performance on the PPPW measure although the magnitude of the association was smaller than for transplant rate.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

19.1

Median performance score

18.5

Minimum performance score

0.0

Maximum performance score

71.4

Standard deviation of performance scores

8.8

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

The Centers for Medicare & Medicaid Services (CMS)

Measure Steward Contact Information

Golden Horten

7500 Security Blvd

Baltimore, MD 21244

golden.horten@cms.hhs.gov

4107864024

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Alexander Yaldo
1415 Washington Heights, Suite 33645 SPH I
Ann Arbor, MI 48109-2029
yaldo@med.umich.edu
(734) 936-5711

Secondary Submitter Contact Information

Jennifer Sardone
1415 Washington Heights, Suite 33645 SPH I
Ann Arbor, MI 48109-2029
jmsto@med.umich.edu
(734) 936-5711

Submitter Comments

For the estimated impact of the measure, 2,541,461, is the number of patient-months for the individual practitioner calculation.

MUC2022-065 Preventive Care and Wellness (composite)

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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).

Please refer to the 2022_MUC List Data_MIPS_PCW_Composite_CompositeCalculationAttachment_FINAL_05_09-22.docx attachment for more information on the exact composite calculation process.

Numerator

NOTE: CMS intends to align the Preventive Care and Wellness CQM components with the most recent versions of the MIPS component measures; any changes to the MIPS component measures will be carried over into this specification. Information in this form align with the MIPS performance period 2022 CQM specifications.

Numerator 1: Patients who received an influenza immunization OR who reported previous receipt of an influenza immunization (Previous Receipt - Receipt of the current season's influenza immunization from another provider OR from same provider prior to the visit to which the measure is applied [typically, prior vaccination would include influenza vaccine given since August 1st])

Numerator 2: Patients who received a pneumococcal vaccination on or after their 60th birthday and before the end of the measurement period; or had an adverse reaction to the vaccine before the end of the measurement period

Numerator 3: Women with one or more mammograms during the 27 months prior to the end of the measurement period

Numerator 4: Patients with one or more screenings for colorectal cancer. Appropriate screenings are defined by any one of the following criteria:

- Fecal occult blood test (FOBT) during the measurement period
- Flexible sigmoidoscopy during the measurement period or the four years prior to the measurement period
- Colonoscopy during the measurement period or the nine years prior to the measurement period

- Computed tomography (CT) colonography during the measurement period or the four years prior to the measurement period
- Fecal immunochemical DNA test (FIT-DNA) during the measurement period or the two years prior to the measurement period.

Numerator 5: Patients with a documented BMI during the encounter or during the previous twelve months, AND when the BMI is outside of normal parameters, a follow-up plan is documented during the encounter or during the previous twelve months of the current encounter

Numerator 6:

- Patients who were screened for tobacco use at least once within the measurement period
- Patients who received tobacco cessation intervention on the date of the encounter or within the previous 12 months
- Patients who were screened for tobacco use at least once within the measurement period AND who received tobacco cessation intervention if identified as a tobacco user on the date of the encounter or within the previous 12 months

Numerator 7: Patient visits where patients were screened for high blood pressure AND have a recommended follow-up plan documented, as indicated, if the blood pressure is elevated or hypertensive

Numerator Exclusions

N/A

Denominator

NOTE: CMS intends to align the Preventive Care and Wellness CQM components with the most recent versions of the MIPS component measures; any changes to the MIPS component measures will be carried over into this specification. Information in this form align with the MIPS performance period 2022 clinical quality measure specifications.

Denominator 1: All patients aged 6 months and older seen for a visit during the measurement period

Denominator 2: Patients 66 years of age and older with a visit during the measurement period

Denominator 3: Women 51 - 74 years of age with a visit during the measurement period

Denominator 4: Patients 50-75 years of age with a visit during the measurement period

Denominator 5: All patients aged 18 and older on the date of the encounter with at least one eligible encounter during the measurement period

Denominator 6:

- All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period

- All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period who were screened for tobacco use during the measurement period and identified as a tobacco user
- All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period

Denominator 7: All patient visits for patients aged 18 years and older at the beginning of the measurement period

Denominator Exclusions

NOTE: CMS intends to align the Preventive Care and Wellness CQM components with the most recent versions of the MIPS component measures; any changes to the MIPS component measures will be carried over into this specification. Information in this form align with the MIPS performance period 2022 clinical quality measure specifications.

Denominator Exclusion Population 1: N/A

Denominator Exclusion Population 2: Patient received hospice services any time during the measurement period

Denominator Exclusion Population 3:

- Women who had a bilateral mastectomy or who have a history of a bilateral mastectomy or for whom there is evidence of a right and a left unilateral mastectomy
- Hospice services used by patient any time during the measurement period
- Palliative care services used by patient any time during the measurement period
- Patients age 66 or older in Institutional Special Needs Plans (SNP) or residing in long term care for more than 90 consecutive days
- Patients 66 years of age and older with at least one claim/encounter for frailty during the measurement period AND a dispensed medication for dementia during the measurement period or the year prior to the measurement period
- Patients 66 years of age and older with at least one claim/encounter for frailty during the measurement period AND either one acute inpatient encounter with a diagnosis of advanced illness or two outpatient, observation, ED or nonacute inpatient encounters on different dates of service with an advanced illness diagnosis during the measurement period or the year prior to the measurement period

Denominator Exclusion Population 4:

- Patients with a diagnosis or past history of total colectomy or colorectal cancer
- Patient was provided hospice services any time during the measurement period
- Patient was provided palliative care services any time during the measurement period
- Patient age 66 or older in Institutional Special Needs Plans (SNP) or residing in long-term care for more than 90 consecutive days

- Patients 66 years of age and older with at least one claim/encounter for frailty during the measurement period AND a dispensed medication for dementia during the measurement period or the year prior to the measurement period
- Patients 66 years of age and older with at least one claim/encounter for frailty during the measurement period AND either one acute inpatient encounter with a diagnosis of advanced illness or two outpatient, observation, ED or nonacute inpatient encounters on different dates of service with an advanced illness diagnosis during the measurement period or the year prior to the measurement period

Denominator Exclusion Population 5:

- Documentation stating the patient has received or is currently receiving palliative or hospice care
- Documentation of patient pregnancy anytime during the measurement period prior to and including the current encounter

Denominator Exclusion Population 6: N/A

Denominator Exclusion Population 7: Patient not eligible due to active diagnosis of hypertension

Denominator Exceptions

NOTE: CMS intends to align the Preventive Care and Wellness CQM components with the most recent versions of the MIPS component measures; any changes to the MIPS component measures will be carried over into this specification. Information in this form alig

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

All Payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Family practice

Measure Type

Process

Is the measure a composite or component of a composite?

Composite measure

If Other, Please Specify

N/A

What data Sources: **are used for the measure?**

Registries

If applicable, specify the data Source:

N/A

Description of parts related to these Sources:

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Wellness and Prevention

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

This composite measure and the score from the Total Per Capita Cost (TPCC) cost measure, which measures the overall cost of care delivered to a beneficiary with a focus on the primary care they receive from their provider(s), could be used to assess impac

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

05726-C-MIPS

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

Yes

Previous Measure Information

In what prior year was this measure published?

2020

What was the MUC ID for the measure in this year?

MUC20-0043

List the CMS CBE MAP workgroup(s) in this year:

Clinician, 2020

What were the programs that MAP reviewed the measure for in this year?

Merit-based Incentive Payment System-Quality, 2020

What was the MAP recommendation in this year?

Merit-based Incentive Payment System-Quality, 2020, MAP did not recommend the measure for rulemaking with potential for mitigation.

Why was the measure not recommended by the MAP workgroups in this year?

Mitigation points include receipt of NQF endorsement and that CMS evaluate whether the components of the measure are appropriately weighted. MAP noted that the seven components of this composite measures are all currently used in MIPS and Part C and D program. CMS has noted their intention to remove the individual component measures if this composite measure is implemented in MIPS. MAP expressed divided concern for potential redundancy with the singular measures for the composite measure already in MIPS and concerns associated with the removal of the individual measures. MAP also expressed concerns related to some of the measure components being topped out. MAP expressed support for preventive measures in general. MAP noted that this measure may impact the 37 million Medicare beneficiaries who receive one or more preventive services, and the 1 in 6 Medicare beneficiaries who are younger than 65 years old who would seek preventive services.

MAP report page number being referenced for this year:

Measure Applications Partnership 2020-2021 Considerations for Implementing Measures in Federal Programs: Clinician, Hospital & PAC/LTC, 2020, page 13

What is the history or background for including this measure on the new measures under consideration list?

Measure previously submitted to MAP, refined and resubmitted per MAP recommendation

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

We could not identify any competing composite measures.

The composite uses existing measures in the MIPS program:

- Quality ID 110: Preventive Care and Screening: Influenza Immunization
- Quality ID 111: Pneumococcal Vaccination Status for Older Adults
- Quality ID 112: Breast Cancer Screening
- Quality ID 113: Colorectal Cancer Screening
- Quality ID 128: Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan
- Quality ID 226: Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention
- Quality ID 317: Preventive Care and Screening: Screening for High Blood Pressure and Follow-Up Documented

How will this measure be distinguished from other similar and/or competing measures?

CMS prioritized development of the Preventive Care and Wellness composite measure because, as a composite, it had several advantages for CMS and stakeholders when compared to using individual measures in a program. Composites can overcome statistical challenges such as small sample sizes while reducing data burden for interpretability (Peterson et al., 2010; Samuel, 2014; van Doorn-Klomberg et al., 2012). The summative nature of the composite's information permits tracking a broader, more comprehensive range of metrics than otherwise possible. This aggregated approach to calculating clinician performance makes composites well suited for pay-for-performance incentives or consumer decisions about clinicians (Peterson et al., 2010). Composite measures are an important strategy to maintain data fidelity as they are more likely to be stable over time, making incentives less sensitive to individual measure performance (Martsolf, 2012; Prentice et al., 2016). Potential implementation of this composite measure not only provides a more comprehensive assessment of a clinician's performance of preventive care than any single measure, but also provides CMS an opportunity to replace the individual measures in the program with a more robust measure, which aligns with the meaningful measure framework's goal. Citations: Martsolf, G. (2012). Creation and Evaluation of Composite Measures of Physician Practice Quality Using Aggregated Health Insurance Claims. The Pennsylvania State University, 194. Peterson, E. D., DeLong Elizabeth R., Masoudi Frederick A., Brien Sean M., Peterson Pamela N., Rumsfeld John S., Shahian David M., & Shaw Richard E. (2010). ACCF/AHA 2010 Position Statement on Composite Measures for Healthcare Performance Assessment. Circulation, 121(15), 1780-1791. https://doi.org/10.1161/CIR.0b013e3181d2ab98. Prentice, J. C., Frakt, A. B., & Pizer, S. D. (2016). Metrics That Matter. Journal of General Internal Medicine, 31(1), 70-73. https://doi.org/10.1007/s11606-015-3559-0. Samuel, C. A. (2014, March). Essays on Health Care Quality and Access: Cancer Care Disparities, Composite Measure Development, and Geographic Variations in Electronic Health Record Adoption. https://dash.harvard.edu/bitstream/handle/1/12274571/Samuel_gsas.harvard_0084L_11583.pdf?sequence=4&isAllowed=y. van Doorn-Klomberg, A.L., J.C. Braspenning, R.C. Feskens, M. Bouma, S.M. Campbell, and D. Reeves. "Precision of Individual and Composite Performance Scores: The Ideal Number of Indicators in an Indicator Set." Medical Care. doi: 10.1097/MLR.0b013e3182726bf1. Epub 2012.

How will this measure add value to the CMS program?

With rising rates of chronic conditions in the general population, wellness and preventive care have become increasingly important to improve outcomes and reduce costs. Research shows that performing the preventive services identified in the measure leads to identification of disease earlier in the care process (screenings) or prevention of disease (immunizations), which enables treatment to begin earlier, potentially improving patient outcomes. The composite measure provides an opportunity for providers and patients to identify and manage a patient's health risks for preventable conditions. This measure assigns a single performance score reflecting overall eligible clinician delivery of age- and sex-appropriate preventive screenings and wellness services to their patients. The seven services in this measure are (1) influenza vaccination, (2) pneumococcal vaccination, (3) breast cancer screening, (4) colorectal-cancer screening, (5) body mass index screening and follow-up, (6) tobacco use screening and intervention, and (7) screening for high blood pressure and follow-up, are recommended by USPSTF, ACIP, and AACE/ACE and apply to the general population (rather than a specific age group with specific risks, for example, older adults with cardiovascular risk). Although increased use of preventive care services may cause a short-term increase in health care costs, it also has the potential to improve patient quality of life and care. A study of preventive services covered under the Affordable Care Act

examined the extent to which lives could be saved if adults over 18 received them, including some addressed by this measure. The authors found preventive services ameliorate 9 of the 10 leading causes of death in America and could save at least 100,000 lives (Fox and Shaw 2015). Among the services referenced are screening for breast cancer, colon cancer, blood pressure, diabetes, and tobacco cessation, as well as influenza and pneumococcal vaccination. Higher rates of patient compliance with the appropriate and recommended preventive services could save additional lives and ensure better health outcomes.

Citations:

Fox, J. B., & Shaw, F. E. (2014). Clinical preventive services coverage and the affordable care act. American Journal of Public Health, 105(1), e7-e10. https://doi.org/10.2105/AJPH.2014.302289

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

data elements are in defined fields in electronic

Feasibility Assessment

All of the component measures are implemented and reported in MIPS as registry-based measures, which supports the feasibility of the composite measure. Furthermore, four clinical sites successfully provided data elements required for calculating and testing the Preventive Care and Wellness composite measure.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

We 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 made use of two different EHR systems across all sites. Most of the analyses in this report are based on data from these four sites for calendar year (CY) 2019. A total of 147 clinicians had patients who were eligible for inclusion in at least one Preventive Care and Wellness component measure in CY 2019; of these clinicians, 99 were eligible

to report at least two Preventive Care and Wellness component measures and saw at least 11 unique patients across the component measures.

We used two approaches to determine whether the Preventive Care and Wellness composite is topped out. The measure was considered topped out if (1) the truncated coefficient of variation (TCV), calculated by first removing the lower and upper 5th percentiles and then dividing the standard deviation by the mean of this truncated distribution, must be less than or equal to 0.10 or 10 percent; and (2) the 75th performance percentile must be statistically indistinguishable (within two standard errors) from the 90th percentile (CMS 2014).

We used a sample of 89 clinicians from 4 clinician networks who saw at least 11 unique patients and had eligible cases for at least two component measures during the 2019 measurement period. We found that the TCV in our sample was 18.4 percent, and the difference between the 75th and 90th percentiles (5.4) was greater than two times the standard deviation for the 90th percentile (2.2). These results indicated that the Preventive Care and Wellness composite was not topped out.

Citations:

CMS. "Analysis of Topped-Out Measures Finalized for the PY 2016 ESRD QIP." Updated June 19, 2014. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/AnalysisofTopped-OutMeasuresFinalizedforthePY2016ESRDQIP.pdf. Accessed August 20, 2020.

Unintended Consequences

This is a new measure and has not been implemented or submitted for NQF endorsement.

We received input on potential unintended consequences from out Technical Expert Panel (TEP) and workgroups. We also received public comment feedback on this topic.

Clinician members on the TEP expressed concern about potential gaming if the Preventive Care and Wellness composite measure is implemented. These TEP members suggested that clinicians could "boost" their composite score by focusing on seeing more patients that are eligible for services that are, in their view, easier to provide. TEP members raised QID 111: Pneumococcal Vaccination Status for Older Adults as an example because the numerator definition requires ever having a pneumococcal vaccination. Conversely, the numerator definition for QID 110: Preventive Care and Screening: Influenza Immunization requires an influenza immunization each flu season, which in their view makes it more difficult to meet. However, the denominator-weighting composite method is designed to decrease this risk. One TEP member stated that the composite weighting did not fully address their concern related to the potential for gaming.

One TEP member expressed concern that the composite would reflect the clinician's Electronic Health Record (EHR) documentation system rather than the quality of care provided by the clinician. This TEP member noted that some EHR systems are better able to capture numerator criteria in the composite, which enables clinicians to document these preventive care and wellness services more easily.

Patient representatives on the TEP noted that the Preventive Care and Wellness composite does not focus specifically on addressing disparities. These TEP members suggested developing and including

component measures that focus specifically on disparities. For example, the TEP member recommended targeting conditions-such as diabetes, kidney disease, and asthma-for future consideration.

Multiple public comments spoke of potential unintended consequences during implementation. Commenters noted concerns about the burden of reporting a composite measure compared to an individual measure. Other commenters suggested monitoring performance by social determinants of health, so that clinicians providing high-quality care based on their patients' demographic characteristics are not being unintentionally penalized. One commenter expressed concern that denominator weighting may unintentionally overweight less impactful component measures. Finally, one commenter noted that the benefits of the composite measure outweighs the potential unintended consequences.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

7

Outline the clinical guidelines supporting this measure

The measures in the Preventive Care and Wellness composite are based on seven preventive services recommended by the United States Preventive Services Task Force (USPSTF) and the Advisory Committee on Immunization Practices (ACIP).

Component 1: The Advisory Committee on Immunization Practices (ACIP, 2021) recommends routine annual influenza vaccination is recommended for all persons aged => 6 months who do not have contraindications. Optimally, vaccination should occur before onset of influenza activity in the community.

Component 2: In 2014, the Advisory Committee on Immunization Practices (ACIP) began recommending a dose of 13-valent pneumococcal conjugate vaccine (PCV13) be followed by a dose of 23-valent pneumococcal polysaccharide vaccine (PPSV23) 6-12 months later in adults aged 65 and older who have not previously received a pneumococcal vaccination, and in persons over the age of two years who are considered to be at higher risk for pneumococcal disease due to an underlying condition. The two vaccines should not be coadministered and intervals for administration of the two vaccines vary slightly depending on the age, risk group, and history of vaccination (Kobayashi, 2015). In 2015, ACIP updated its recommendation and changed the interval between PCV13 and PPSV23, from 6-12 months to at least one year for immunocompetent adults aged >=65 years who have not previously received pneumococcal vaccine. For immunocompromised vaccine-naÃ¯ve adults, the minimum acceptable interval between PCV13 and PPSV23 is 8 weeks. Both immunocompetent and immunocompromised adults aged >=65 years who have previously received a dose of PPSV23 when over the age of 65 should receive a dose of PCV13 at least one year after PPSV23 (>=1 year). Immunocompetent and immunocompromised adults aged >=65 who have previously received a dose of PPSV23 when under the age of 65, should also receive a dose of PCV13 at least one year after PPSV23 (>=1 year) and then another dose of PPSV23 at least one year after PCV13. It is recommended that for those that have this alternative three-dose schedule (2 PPSV23 and 1 PCV13), the three doses should be spread over a time period of five or more years (Kobayashi, 2015). In 2019, based on a review of accrued evidence ACIP changed the recommendation for PCV13 use in adults. ACIP recommends a routine single dose of PPSV23 for adults aged => 65 years. Shared clinical decision-making is recommended regarding administration of PCV13 to persons aged => 65 years who do not have an immunocompromising condition, cerebrospinal fluid leak, or cochlear implant and who have not previously received PCV13. If a

decision to administer PCV13 is made, PCV13 should be administered first, followed by PPSV23 at least 1 year later. In 2022, the Pneumococcal vaccination recommendations were simplified across age and risk group. Eligible adults may receive either PCV15 in series with PPSV23 or PCV20 alone.

Component 3: The United States Preventive Services Task Force (2016) recommends biennial screening mammography for women aged 50-74 years (B recommendation).

Component 4: The United States Preventive Services Task Force (2021) recommends screening for colorectal cancer starting at age 50 years and continuing until age 75 years. This is a Grade A recommendation. The USPSTF recommends screening for colorectal cancer in adults aged 45 to 49 years. This is a Grade B recommendation

Component 5: All adults should be screened annually using a BMI measurement. BMI measurements >25kg/m2 should be used to initiate further evaluation of overweight or obesity after taking into account age, gender, ethnicity, fluid status, and muscularity; therefore, clinical evaluation and judgment must be used when BMI is employed as the anthropometric indicator of excess adiposity, particularly in athletes and those with sarcopenia (Garvey, et al., 2016 AACE/ACE Guidelines, 2016. pp. 12-13) This is a grade A recommendation.

Component 6: The United States Preventive Services Task Force (2021) recommends that clinicians ask all adults about tobacco use, advise them to stop using tobacco, and provide behavioral interventions and U.S. Food and Drug Administration (FDA)-approved pharmacotherapy for cessation to adults who use tobacco. This is a grade A recommendation.

Component 7: The United States Preventive Services Task Force (2021) screening for hypertension in adults 18 years or older with office blood pressure measurement (OBPM). The USPSTF recommends obtaining blood pressure measurements outside of the clinical setting for diagnostic confirmation before starting treatment. This is a grade A recommendation.

Name the guideline developer/entity

Each component measure has a clinical guideline recommendation from either the United States Preventive Services Task Force, the Advisory Committee on Immunization Practices, or the American Association of Endocrinology.

Publication year

2022

Full citation +/- URL

Component 1: Grohskopf LA, Alyanak E, Ferdinands JM, et al. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices, United States, 2021-22 Influenza Season. MMWR Recomm Rep 2021;70(No. RR-5):1-28. DOI: http://dx.doi.org/10.15585/mmwr.rr7005a1

Component 2: Kobayashi M, Farrar JL, Gierke R, et al. Use of 15-Valent Pneumococcal Conjugate Vaccine and 20-Valent Pneumococcal Conjugate Vaccine Among U.S. Adults: Updated Recommendations of the Advisory Committee on Immunization Practices - United States, 2022. MMWR Morb Mortal Wkly Rep 2022;71:109-117. DOI: http://dx.doi.org/10.15585/mmwr.mm7104a1

Component 3: Siu A, on behalf of the U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force

Recommendation Statement. Annals of Internal Medicine 2016;164(4):279-297.
https://www.uspreventiveservicestaskforce.org/home/getfilebytoken/pth_4-Mnau_pZubaefDvUk

Component 4:

US Preventive Services Task Force. Screening for Colorectal Cancer US Preventive Services Task Force Recommendation Statement. JAMA. 2021;325(19):1965-1977. doi:10.1001/jama.2021.6238.
https://www.uspreventiveservicestaskforce.org/home/getfilebytoken/2Wc3FRHpVDPX2jT_WzjEXX

Component 5: Garvey W, Mechanick J, Brett M, et al. American Association of Clinical Endocrinologists and American College of Endocrinology Comprehensive Clinical Practice Guidelines For Medical Care of Patients with Obesity. Endocrine Practice. 2016;22(3):1-203.
https://www.sciencedirect.com/science/article/pii/S1530891X20446300?via%3Dihub

Component 6: US Preventive Services Task Force. Interventions for Tobacco Smoking Cessation in Adults, Including Pregnant Persons US Preventive Services Task Force Recommendation Statement. JAMA. 2021;325(3):265-279. doi:10.1001/jama.2020.25019
https://www.uspreventiveservicestaskforce.org/home/getfilebytoken/uM5P-6XpsqVfEC2axGnZaD

Component 7: US Preventive Services Task Force. Screening for Hypertension in Adults US Preventive Services Task Force Reaffirmation Recommendation Statement. JAMA. 2021;325(16):1650-1656. doi:10.1001/jama.2021.4987.
https://www.uspreventiveservicestaskforce.org/home/getfilebytoken/5P_7f2rbEquA-ZR-fJJfHK

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

Quality ID 110: Preventive Care and Screening: Influenza Immunization: The Advisory Committee on Immunization Practices (ACIP, 2021) recommends routine annual influenza vaccination is recommended for all persons aged => months who do not have contraindications. Optimally, vaccination should occur before onset of influenza activity in the community.

Additional guidelines are included in the 2022_MUC List Data_MIPS_PCW_Composite_EvidenceAttachment_FINAL_05-09-22.docx attachment.

What evidence grading system did the guideline use to describe strength of recommendation?

USPSTF

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

Grade A: The USPSTF recommends the service. There is high certainty that the net benefit is substantial.

Grade B: The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.

Grade C: The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient There is at least moderate certainty that the net benefit is small.

Grade D: The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.

Grade I: The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

USPSTF Grade A, Strong recommendation or similar

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

USPSTF

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

High: The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.

Moderate: The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as:

- The number, size, or quality of individual studies.
- Inconsistency of findings across individual studies.
- Limited generalizability of findings to routine primary care practice.
- Lack of coherence in the chain of evidence.

As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.

Low: The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of:

- The limited number or size of studies.
- Important flaws in study design or methods.
- Inconsistency of findings across individual studies.

- Gaps in the chain of evidence.
- Findings not generalizable to routine primary care practice.

Lack of information on important health outcomes.

More information may allow estimation of effects on health outcomes.

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

High or similar

List the guideline statement that most closely aligns with the measure concept.

Quality ID 110: Preventive Care and Screening: Influenza Immunization: The Advisory Committee on Immunization Practices (ACIP, 2021) recommends routine annual influenza vaccination is recommended for all persons aged => months who do not have contraindications. Optimally, vaccination should occur before onset of influenza activity in the community.

Additional guidelines are included in the 2022_MUC List Data_MIPS_PCW_Composite_EvidenceAttachment_FINAL_05-09-22.docx attachment.

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source: **of empirical data**

N/A

Summarize the empirical data

N/A

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

243,000,000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Other (enter here):: The Preventive Care and Wellness composite is a process measure that does not need risk adjustment because the measured processes are appropriate for all patients included in each component denominator, and the component measures excl

Cost estimate completed

Yes

Cost estimate methods and results

Preventive care is linked directly to the prevention of 9 of the 10 leading causes of death in the United States (Fox and Shaw, 2015). Depending on the condition avoided, the preventive care services in the Preventive Care and Wellness (composite) have the potential to avoid on average between $615 and $10,472 in direct medical costs per episode for acute conditions or per patient annually for chronic conditions. Estimated cost savings for each preventive care component measure are below. Years of cost literature varied according to Source: . All dollars have been adjusted for inflation to 2021 dollars using the U.S. Bureau of Labor Statistics CPI inflation Calculator.

Component measure 1, using condition of influenza: $615 on average per episode. This includes medical costs associated with being ill (not medically attended), outpatient care, hospitalization, and death (Molinari et al., 2007).

Component measure 2, using condition of pneumonia: $3,151 on average per episode. This includes medical costs associated with outpatient care, emergency/urgent care, hospitalization, and follow up visits within 28 days (Tong et al., 2018).

Component measure 3, using condition of breast cancer: $2,698 on average per patient annually. This includes inpatient and outpatient care during years that were not the initial or final year of care, both of which have higher costs on average. (Mariotto et al., 2011).

Component measure 4, using condition of colorectal cancer: $3,862 for female patient annually; $5,617 for male patient annually. This includes medical costs associated with inpatient and outpatient care during years that were not the initial or final year of care, both of which have higher costs on average. (Mariotto et al., 2011).

Component measure 5, using condition of diabetes: $10,472 on average per patient annually. This includes medical costs associated with institutional care (inpatient, nursing/residential facility, and hospice), outpatient care, and outpatient medications and supplies (American Diabetes Association, 2018).

Component measure 6, using condition of smoking-related disease: $906 on average per patient annually. This includes the percent of medical costs from the National Health Expenditure Accounts - which consists of historical and projected health care spending by good, service, Source: of funding, and sponsor (CMS, 2019) - that can be attributed to smoking-related disease for people with ages greater than or equal to 19 (Xu et al., 2014).

Component measure 7, using condition of hypertension: $2,905 on average per patient annually. This includes medical costs associated with inpatient, outpatient, medication, emergency room, and other for hypertension (Wang 2017).

Assumptions:

(1) Dollars have been adjusted for inflation to standardize for comparison across literature for each component.

(2) For pneumonia (the condition for component measure 2), the year of dollars before adjusting for inflation is assumed to be 2014, the final year of the data used for the cost analysis.

(3) The condition of diabetes is used for the BMI cost estimate (component measure 5), without consideration for other conditions that may stem from a high or low BMI.

(4) The condition of hypertension is used for the blood pressure estimate (component measure 7), without consideration for other conditions that may stem from a high or low blood pressure.

(5) The per-person estimate for smoking-related disease (the condition for component measure 6) is estimated among all people ages 19+ from the 2010 U.S. Census.

(6) The interaction between distinct component measures (e.g., a patient that has a high BMI might have both diabetes and hypertension) may impact the cost avoided that can be attributed to a single component measure.

Citations:

Component measure 1, using condition of influenza: Molinari, N.-A. M., Ortega-Sanchez, I. R., Thompson, W. W., Wortley, P. M., Weintraub, E., & Bridges, C. B. (2007). The annual impact of seasonal influenza in the US: Measuring disease burden and costs. Vaccine, 25(27), 5086-5096.

Component measure 2, using condition of pneumonia: Tong, S., Amand, C., Kieffer, A., & Kyaw, M. H. (2018). Trends in healthcare utilization and costs associated with pneumonia in the United States during 2008-2014. BMC Health Services Research, 18(1), 715. https://doi.org/10.1186/s12913-018-3529-4

Component measure 3, using condition of breast cancer: Mariotto, A. B., Robin Yabroff, K., Shao, Y., Feuer, E. J., & Brown, M. L. (2011). Projections of the cost of cancer care in the united states: 2010-2020. JNCI: Journal of the National Cancer Institute, 103(2), 117-128. https://doi.org/10.1093/jnci/djq495

Component measure 4, using condition of colorectal cancer: Mariotto, A. B., Robin Yabroff, K., Shao, Y., Feuer, E. J., & Brown, M. L. (2011). Projections of the cost of cancer care in the United States: 2010-2020. JNCI: Journal of the National Cancer Institute, 103(2), 117-128. https://doi.org/10.1093/jnci/djq495

Component measure 5, using condition of diabetes: American Diabetes Association. (2018). Economic costs of diabetes in the U. S. in 2017. Diabetes Care, 41(5), 917-928. https://doi.org/10.2337/dci18-0007

Component measure 6, using condition of smoking-related disease: National health expenditure data. (2019, December 17). [Government]. Centers for Medicare and Medicaid; CMS. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData

Howden, L. M., & Meyer, J. A. (2011). Age and Sex Composition: 2010 (2010 Census Briefs). United States Census Bureau. https://www.census.gov/library/publications/2011/dec/c2010br-03.html

Xu, X., Bishop, E. E., Kennedy, S. M., Simpson, S. A., & Pechacek, T. F. (2015). Annual healthcare spending attributable to cigarette smoking: an update. American Journal of Preventive Medicine, 48(3), 326-333. https://doi.org/10.1016/j.amepre.2014.10.012

Component measure 7, using condition of hypertension: Wang, G., hou, X., Zhuo, X., & Zhang, P. (2017). Annual total medical expenditures associated with hypertension by diabetes status in U. S. adults. American Journal of Preventive Medicine, 53(6), S182-S189. https://doi.org/10.1016/j.amepre.2017.07.018

Additional: Fox, J. B., & Shaw, F. E. (2014). Clinical preventive services coverage and the affordable care act. American Journal of Public Health, 105(1), e7-e10. https://doi.org/10.2105/AJPH.2014.302289

Cpi inflation calculator. (2021). [Government]. U.S. Bureau of Labor Statistics. Retrieved May 5, 2021, from https://www.bls.gov/data/inflation_calculator.htm

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

5

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

0

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

10

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

5

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise;Other (enter here): Test-Retest Reliability

Signal-to-Noise: Name of statistic

Signal-to-noise R statistic. The signal-to-noise statistic, R, summarizes the proportion of variation between clinician scores on a measure that is due to real differences in underlying characteristics (such as differences in medical care), rather than to background-level or random variation (for example, due to measurement or sampling error). We estimated signal-to-noise reliability for the composite measure using the iterative empirical Bayes method.

Signal-to-Noise: Sample size

99

Signal-to-Noise: Statistical result

0.977

Signal-to-Noise: Interpretation of results

The median reliability score of the composite measure was high (0.977). Reliability coefficients above 0.70 are considered sufficient to draw conclusions about groups, and values above 0.9 are considered sufficient to draw conclusions about individuals (Adams 2009). Reliability testing results remained high across the several methods used and for this component-level method. Citation: Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation, 2009. Available at https://www.rand.org/pubs/technical_reports/TR653.html. Accessed May 3, 2022.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

Test-retest reliability. We used bootstrap resampling to draw 2,000 random samples stratified by clinician, maintaining the same number of patients as in the original sample, calculated clinicians' composite scores and then computed the Spearman rank-order correlation (rho) and intraclass correlations (ICCs) between the scores.

Other: Sample size

99

Other: Statistical result

Mean Spearman rho = 0.971. Mean ICC = 0.967.

Other: Interpretation of results

Test-retest reliability: Our testing results suggest that the composite measure scores are very stable across multiple samples. The Spearman rho and ICC exceeded 0.9. For context, Spearman rho values above 0.8 indicate very strong agreement (Chan 2003); ICC values between 0.75 and 0.90 indicate good reliability, and ICC values greater than 0.90 indicate excellent reliability (Koo and Li 2015). Reliability testing results remained high across the several methods used and for this component-level method. Citations: Chan, Y.H. "Biostatistics 104: Correlational Analysis." Singapore Medical Journal, vol. 4, no. 12, 2003, pp. 614-619. Koo, T.K., and M.Y. Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." Journal of Chiropractic Medicine, vol. 15, 2016, pp. 155-163. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/pdf/main.pdf. Accessed October 31, 2021.

Empiric Validity

Yes

Empiric Validity: Statistic name

Calculation of correlations between pairs of the composite's component measures to assess construct validity. Please see the 2022_MUC List Data_MIPS_PCW_Composite_EmpiricalValidityAttachment_FINAL_05-09-22.docx attachment for more information about the conducted empirical validity testing.

Empiric Validity: Sample size

99

Empiric Validity: Statistical result

0.69

Empiric Validity: Methods and findings

We tested the construct validity of the composite measure by calculating the correlations between pairs of the seven component measures in the composite, examining every possible combination of component measures. We selected this approach in part because the individual component measures are already in use in MIPS and are presumed to be valid. With the validity of the component measures already established, we aimed to establish the construct of the composite as a whole. We based our approach on the rationale that all seven component measures assess the underlying concept of appropriate delivery of preventive care and wellness services (CMS 2021). We found positive

correlations among most of the measures which provides support for the construct validity of the composite measure. The Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention measure (QID 226) had a strong negative correlation with performance on the Preventive Care and Screening: Influenza Immunization (QID 110) and Breast Cancer Screening (QID 112) measures. Our TEP members noted that they would not expect all of the components to be positively correlated. In addition, these results suggest a quality gap that this composite measure will directly address. Implementing the composite measure will help draw attention to these activities and should improve the correlation of performance across the component measures.

In addition, we examined how the Preventive Care and Wellness composite scores differ among groups of patients by certain characteristics. Based on differences observed in the literature, we expected differences in composite scores across the following patient characteristics available in our data: sex (Applewhite et al., 2020; Hall et al., 2018; Borsky et al., 2018; Asch et al., 2006; Viera et al., 2006), race (Malhotra et al., 2017; Holden et al., 2015; Asch et al., 2006), insurance status (Malhotra et al., 2017; Lau et al., 2014; DeVoe et al., 2011), and Medicaid status (Goudie et al., 2020; Asch et al., 2006). This analysis was limited to one of the clinical sites, which included 46,728 patients.

We analyzed patient-level composite scores using a t-test with a stratified sample and multivariate logistic regression. We evaluated differences by sex, race (white vs. non-white), insurance status (insured vs. not insured), and Medicaid status using Cohen's D and absolute differences. Cohen's D is the difference in mean clinician scores divided by the pooled standard deviation. For Cohen's D, a value of 0.2 is considered a "small" effect size, 0.5 is a "medium" effect size, and 0.8 is a "large" effect size. We also analyzed absolute differences between patient subgroups. We found statistically significant differences among patient subpopulations using the composite measure, which supports the validity of the measure.

We conducted face validity to support empiric validity through clinician interviews, our Technical Expert Panel (TEP), patient and family workgroup, and public commenters for the Preventive Care and Wellness composite. Most feedback supported the preventive care and wellness as a concept. These Sources: noted that the composite score alone is useful for payors and accountability programs, and that granular component measure results are required for quality improvement and patient decision making.

We did not conduct Patient/Encounter Level (Data Element Level) Testing for the Preventive Care and Wellness composite measure. Each component measure is currently used in MIPS, which supports the reliability and validity of the data elements used to calculate the component measures included in the composite.

Additional details on the validity testing and results are included in the 2022_MUC List Data_MIPS_PCW_Composite_EmpiricalValidityAttachment_FINAL_05-09-22.docx attachment.

Citation:

Applewhite, A., Stancampiano, F., Harris, D., Manaois, A., Dimuna, J., Glenn, J., Heckman, M., Brushaber, D., Sher, T., Valery, J. (2020). A Retrospective Analysis of Gender-Based Difference in Adherence to Influenza Vaccination during the 2018-2019 Season. Journal of Primary Care & Community Health. 11, 1-6. https://journals.sagepub.com/doi/10.1177/2150132720958532. Accessed September 15, 2021.

Asch, S., Kerr, E., Keesey, J., Adams, J., Setodji, C., Malik, S., and McGlynn, E. (2006). Who Is at Greatest Risk for Receiving Poor-Quality Health Care? N Engl J Med, 354, 1147-1156. https://www.nejm.org/doi/full/10.1056/NEJMsa044464. Accessed September 15, 2021.

Borsky, A., Zhan, C., Miller, T., Ngo-Metzger, Q., Bierman, A., and Meyers, D. (2018). Few Americans Receive All High-Priority, Appropriate Clinical Preventive Services. Health Affairs, 37(6), 925-928. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2017.1248. Accessed September 15, 2021.

Centers for Medicare & Medicaid Services. "CMS Measures Management System Blueprint, Version 17.0." Washington, DC: Centers for Medicare & Medicaid Services, U.S. Department of Health and Human Services, September 2021. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Blueprint.pdf. Accessed November 3, 2021.

DeVoe, J., Fryer, G., Phillips, R., and Green, L. (2011). Receipt of Preventive Care Among Adults: Insurance Status and Usual Source: of Care. American Journal of Public Health, 93(5), 786-791. https://doi.org/10.2105/AJPH.93.5.786. Accessed September 15, 2021.

Goudie, A., Martin, B., Li, C., Lewis, K, Han, X., Kathe, N., Wilson, J., Thompson, J. (2020). Higher Rates of Preventive Health Care With Commercial Insurance Compared With Medicaid: Findings From the Arkansas Health Care Independence "Private Option" Program. Medical Care. 58(2), 120-127. https://journals.lww.com/lww-medicalcare/Fulltext/2020/02000/Higher_Rates_of_Preventive_Health_Care_With.5.aspx. Accessed September 15, 2021.

Hall, I., Tangka, F., Sabatino, S., Thompson, T., Graubard, B., Nancy Breen, N. (2018). Patterns and Trends in Cancer Screening in the United States. Prev Chronic Dis. 15(E97), 1-13. https://www.cdc.gov/pcd/issues/2018/17_0465.htm. Accessed September 15, 2021.

Holden, C., Chen, J., Dagher, R. (2015). Preventive Care Utilization Among the Uninsured by Race/Ethnicity and Income. American Journal of Preventive Medicine. 48(1),13-21. https://www.sciencedirect.com/science/article/pii/S0749379714004966?via%3Dihub. Accessed September 15, 2021.

Lau, J., Adams, S., Park, M., and Boscardin, W., Irwin, C. (2014). Improvement in Preventive Care of Young Adults After the Affordable Care Act. JAMA Pediatr. 168(12), 1101-1106. https://jamanetwork.com/journals/jamapediatrics/article-abstract/1913624. Accessed September 15, 2021.

Malhotra, J., Rotter, D., Tsui, J., Llanos, A., Balasubramanian, B., Demissie, K. (2017). Impact of Patient-Provider Race, Ethnicity, and Gender Concordance on Cancer Screening: Findings from Medical Expenditure Panel Survey. Cancer Epidemiol Biomarkers Prev 26(12), 1804-1811. https://aacrjournals.org/cebp/article/26/12/1804/71200/Impact-of-Patient-Provider-Race-Ethnicity-and. Accessed September 15, 2021.

Viera, A.J., Thorpe, J.M. & Garrett, J.M. (2006). Effects of sex, age, and visits on receipt of preventive healthcare services: a secondary analysis of national data. BMC Health Serv Res 6(15), https://doi.org/10.1186/1472-6963-6-15. Accessed September 15, 2021.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

10

Face Validity: Result

8

Patient/Encounter Level Testing

No

Type of Analysis

N/A

Sample Size

N/A

Statistic Name

N/A

Statistical Results

N/A

Interpretation of results

N/A

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

51.5

Median performance score

52.7

Minimum performance score

20.7

Maximum performance score

76.7

Standard deviation of performance scores

11.2

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

Centers for Medicare & Medicaid Services

Measure Steward Contact Information

Joel Andress

7500 Security Boulevard, Mailstop S3-10-26

Baltimore, MD 21244

Joel.Andress@cms.hhs.gov

4107865237

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Penoyer

955 Massachusetts Avenue, Suite 801

Cambridge, MA 02139

spenoyer@mathematica-mpr.com

6077156933

Secondary Submitter Contact Information

Christine Holland

1100 First Street, NE, 12th Floor

Washington, DC 20002

cholland@mathematica-mpr.com

2024845271

Submitter Comments

For additional context and rationale for data element testing, please refer to the section for Measure Score Level (Accountability Entity Level) Testing, Empiric Validity: Methods and findings.

MUC2022-098 Connection to Community Service Provider

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Number of pts 18 or older who had contact with a Community Service Provider (defined as any independent, for-profit, non-profit, state, territorial, or local agency capable of addressing core or supplemental health-related social needs) for at least 1 of their HRSNs within 60 days after screening (annually).

Numerator Exclusions

N/A

Denominator

Number of pts 18 or older who screened positive for at least 1 of the 5 HRSN domains (food insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety) during the period of performance (annually)

Denominator Exclusions

Patients who opt-out of connection with CSP

Denominator Exceptions

N/A

State of development

Field (Beta) Testing

State of Development Details

- Using a standard, validated screening tool, the CMS Accountable Healthcare Communities program has screened nearly 1 million patients for HRSN in 21 states, with 33% of beneficiaries screened having at least one HRSN.
- Of patients with at least one HRSN who were eligible for navigation, 74% of patients accepted navigation related to their HRSN.
- 18% of patients accepting navigation either reported at least one HRSN resolved (14%) or connection with a CSP without resolution (4%).

Sources:

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

- The HRSA Health Center Controlled Network (HCCN) grant cycle beginning in Q3 2022 will measure the number of participating FQHCs, RHCs, and lookalikes using social risk factor data to inform care plan development and facilitate closed-loop community referrals for at least 75% of patients identified as having a risk factor.
- According to BPHC, 70% of HRSA-funded Health Centers nationwide were enrolled in an HCCN for the 2019-2022 grant cycle.

Sources:

https://apply07.grants.gov/apply/opportunities/instructions/PKG00269508-instructions.pdf

https://bphc.hrsa.gov/program-opportunities/hccn/fy2022-faqs

- Medicare and CMMI are currently exploring quality measures related to identifying and addressing social needs to support several ACO-related initiatives.

Source:

https://www.nejm.org/doi/full/10.1056/NEJMp2202991

What is the target population of the measure?

All payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Primary care

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

Social and Economic Determinants

What data sources are used for the measure?

Administrative Data (non-claims);Electronic Clinical Data (non-EHR);Electronic Health Record;Standardized Patient Assessments;Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

Patient reported data and standardized assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSNs.

EHR-and non-EHR electronic clinical data, as well as patient reported data, will be used to determine whether contact was made with a CSP.

Administrative data will be used for measure stratification and ongoing performance monitoring.

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care;Behavioral health clinic;Community hospital;Emergency department;Federally qualified health center (FQHC);Hospital outpatient department (HOD);Hospital inpatient acute care facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Equity

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

The measure correlates to specific MIPS Quality Improvement Activities as follows:

- Use QDCR data for ongoing practice assessment and improvements (IA_PSPA_7)
- Use of toolsets or other reSources: to close healthcare disparities in communities (IA_PM_6)
- Practice Improvements that Engage Community ReSources: to Support Patient Health (IA_CC_14)
- Provide Clinical-Community Linkages (IA_PM_18)

Source:

https://qpp.cms.gov/mips/explore-measures?tab=improvementActivities&py=2022

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

Measure currently used in a CMS program being submitted as-is for a new or different program

Range of years this measure has been used by CMS Programs

Accountable Health Communities Pilot (2017-2022)

What other federal programs are currently using this measure?

Accountable Health Communities Pilot

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

Yes (enter here):: We strongly recommend stratifying the measures by race/ethnicity.

- Data from the AHC found racial/ethnic minorities were over-represented in the navigation-eligible groups.
- CMS has stated in its strategic plan that the imperative to stratify by race/ethnicity is a global issue for the Agency that applies to all measures.

Sources: :

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

https://www.cms.gov/sites/default/files/2022-04/Health%20Equity%20Pillar%20Fact%20Sheet_1.pdf

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

OCHIN's nationwide membership of Community Health Centers has documented over 1.2 million HRSN screenings as structured data in its centralized EHR record. - Further, CMS has the opportunity to leverage and apply CMMI's 5+ years of data and experience with AHC. Using a standard, validated screening tool, AHC has screened nearly 1 million beneficiaries for HRSN in 21 states, with 33% of beneficiaries screened having at least one HRSN. AHC used screening, referral, and navigation data files extracted by NewWave (Centers for Medicare & Medicaid Services [CMS] Enterprise Portal contractor)

and generated by Mathematica Policy Research (the AHC implementation contractor) using data submitted by bridge organizations.

Sources:

https://innovation.cms.gov/innovation-models/ahcm
https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf
https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

A number of CMMI models and participating entities have incorporated DOH screening and navigation data into their quality frameworks and care management plans for beneficiaries. CMMI's Comprehensive Primary Care Plus (CPC+) model reported in 2020 that 86% of ~1,500 Track 1 practices and 99% of ~1,500 Track 2 practices (together serving ~2.4M beneficiaries) are implementing DOH screening. CMMI required that by Program Year 3, Track 2 practices would use an electronic screening tool to assess patients' health-related social needs and store an inventory of reSources: to meet patients' needs; notably, by Program Year 2, Track 1 practices were as likely as Track 2 practices to report implementing these DOH functions, even absent a requirement that they do so.

Source:

https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2

Likewise, annual evaluations of other current CMMI models, including the State Innovation Model and Next Generation ACOs, report that participants are investing in staffing and infrastructure to conduct DOH screening and navigation. The 2021 Comprehensive End-Stage Renal Disease Care Model evaluation, for example, reported that "[m]any beneficiaries are protein malnourished and don't eat enough fresh produce. Some beneficiaries go to the hospital to get meals." ESRD Seamless Care Organizations have begun to monitor food insecurity and provide food gift cards to both low-income beneficiaries and those above the poverty level, to address beneficiaries' non-adherence to nutritional guidelines and reduce the risk of increased utilization and costs.

Sources:

https://downloads.cms.gov/files/cmmi/sim-rd2-test-ar3.pdf

https://innovation.cms.gov/data-and-reports/2020/nextgenaco-thirdevalrpt-fullreport

https://innovation.cms.gov/data-and-reports/2021/cec-annrpt-py4

Finally, a recent Physician's Foundation survey on provider well-being and attitudes towards social drivers of health found that 90% of providers want greater means and reSources: to address DOH in their practices.

Source:

https://physiciansfoundation.org/wp-content/uploads/2022/03/SDOH-Survey-Report.pdf

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Other digital method;Hybrid: Manual abstraction

Evidence of Performance Gap

CMS has already identified social and economic determinants as both a measurement priority and gap in Meaningful Measures 2.0, and as a central part of its Health Equity strategic plan pillar moving forward. Other public and private organizations such as ASPE, NQF and NCQA have identified this as a critical gap.

Sources:

https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization

https://aspe.hhs.gov/pdf-report/second-impact-report-to-congress

https://www.cms.gov/sites/default/files/2022-04/Health%20Equity%20Pillar%20Fact%20Sheet_1.pdf

https://www.qualityforum.org/News_And_Resources/Press_Releases/2019/National_Quality_Forum_Leads_National_Call_to_Address_Social_Determinants_of_Health__through_Quality_and_Payment_Innovation.aspx

https://blog.ncqa.org/ncqa-releases-its-social-determinants-of-health-reSource: -guide/

Unintended Consequences

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: . This challenge was noted as a primary barrier to connecting beneficiaries to resources in the AHC Year 1 evaluation. There is a well-documented and well-tested catalog of additional tools, infrastructure, and investments that can be implemented to support practices in acting on this measure.

Sources: :

https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/Unintended%20consequences%20of%20screening%20for%20social%20determinants.pdf

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

https://nhchc.org/wp-content/uploads/2020/04/NHCHC_Community-Information-Exchange2.pdf

https://governor.nc.gov/news/north-carolina-creates-nation%E2%80%99s-first-statewide-infrastructure-connecting-healthcare-and-human

https://blueshieldcafoundation.org/sites/default/files/publications/downloadable/Investing%20in%20Health%20-%20A%20Federal%20Action%20Plan%20-January%202021_Final.pdf

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

1

Outline the clinical guidelines supporting this measure

The USPSTF provides a "B" recommendation that recommends that clinicians screen for Intimate Partner Violence (one of the HRSNs included in the denominator of the proposed measure) in women of

reproductive age and provide or refer women who screen positive to ongoing support services. They cite adequate evidence that available screening instruments can identify IPV in women, and that screening for IPV in women of reproductive age and providing or referring women who screen positive to ongoing support services has a moderate net benefit.

However, in addition to this individual measure, USPSTF also released recently released a technical brief on screening and interventions for social risk factors which notes 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. The report also highlights the lack of unintended consequences encountered during implementation of social risk screening and intervention in studies reporting these outcomes, despite any perceived barriers.

Sources:

https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/intimate-partner-violence-and-abuse-of-elderly-and-vulnerable-adults-screening

https://uspreventiveservicestaskforce.org/uspstf/sites/default/files/inline-files/Social%20Risk%20Factors%20Tech%20Brief_Assembled%20for%20Web_Sep%202021_1.pdf

Name the guideline developer/entity

U.S. Preventive Services Task Force (USPSTF)

Publication year

2018

Full citation +/- URL

JAMA. 2018;320(16):1678-1687. doi:10.1001/jama.2018.14741
https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/intimate-partner-violence-and-abuse-of-elderly-and-vulnerable-adults-screening

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

The USPSTF recommends that clinicians screen for intimate partner violence (IPV) in women of reproductive age and provide or refer women who screen positive to ongoing support services.

What evidence grading system did the guideline use to describe strength of recommendation?

USPSTF

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

A: The USPSTF recommends the service. There is high certainty that the net benefit is substantial. Offer or provide this service.

B: The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. Offer or provide this service.

C: The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient pReferemces:. There is at least moderate certainty that the net benefit is small. Offer or provide this service for selected patients depending on individual circumstances.

D: The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. Discourage the use of this service.

I: The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

USPSTF Grade B or D, Moderate recommendation or similar

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

USPSTF

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

A. The USPSTF recommends the service. There is high certainty that the net benefit is substantial. Offer or provide this service.

B. The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. Offer or provide this service.

C. The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small. Offer or provide this service for selected patients depending on individual circumstances.

D. The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. Discourage the use of this service.

I. The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

USPSTF Grade B or D, Moderate recommendation or similar

List the guideline statement that most closely aligns with the measure concept.

The USPSTF recommends that clinicians screen for intimate partner violence (IPV) in women of reproductive age and provide or refer women who screen positive to ongoing support services.

Number of systematic reviews that inform this measure concept

3

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

See attached document with supporting evidence and research.

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies)

Summarize the empirical data

- Using a standard, validated screening tool, the CMS Accountable Healthcare Communities program has screened nearly 1 million patients for HRSN in 21 states, with 33% of beneficiaries screened having at least one HRSN.
- Of patients with at least one HRSN who were eligible for navigation, 74% of patients accepted navigation related to their HRSN.
- 18% of patients accepting navigation either reported at least one HRSN resolved (14%) or connection with a CSP without resolution (4%).

Sources: :

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

CMMI's Comprehensive Primary Care Plus (CPC+) model reported in 2020 that 86% of ~1,500 Track 1 practices and 99% of ~1,500 Track 2 practices (together serving ~2.4M beneficiaries) are implementing DOH screening.

Source:

https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines;Peer-Reviewed Systematic Review;Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through stratification of results

Cost estimate completed

Yes

Cost estimate methods and results

The cost avoided to payers and providers annually is likely to be significant given the research demonstrating increased utilization, readmissions, cost and increased financial liability for providers caring for patients with increased social risk.

Sources: :

https://www.gsfb.org/wp-content/uploads/2018/08/Berkowitz-S.-A.-Basu-S.-Meigs-J.-B.-Seligman-H.-K.-Food-Insecurity-and-Health-Care.pdf

https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2020.01742

https://pubmed.ncbi.nlm.nih.gov/32897345/

For example, extensive research exists demonstrating increased healthcare expenditures to patients including Medicare beneficiaries associated with DOH. The example below provides the annualized increase in annual healthcare expenditures (PMPY in 2015 dollars) associated with food insecurity across different disease categories across all payor types in the peer-reviewed literature:

- Diabetes Mellitus: $4,413.61
- Hypertension: $2,175.20
- Heart Disease: $5,144.05
- Overall: $1,863

Source:

https://www.gsfb.org/wp-content/uploads/2018/08/Berkowitz-S.-A.-Basu-S.-Meigs-J.-B.-Seligman-H.-K.-Food-Insecurity-and-Health-Care.pdf

The AHC Year 1 evaluation found that Medicare FFS beneficiaries in the Assistance Track intervention group had 9% fewer ED visits than those in the control group in the first year after screening. (No Medicaid utilization/cost data reported yet.)

Source:

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

3,162

Meaningful to Patients: Number indicating survey/tool is meaningful

2,441

Meaningful to Clinicians: Numbers consulted

10,078

Meaningful to Clinicians: Number indicating survey/tool is meaningful

8,800

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

4

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

4

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Other (enter here):: Predictive and Empirical Validity

Testing methodology and results

Empirical validity (through AHC and CPC+ practice implementation across 3+ million beneficiaries over last ~ 5-year time frame) and Psychometric and Pragmatic Property Analysis (see https://pubmed.ncbi.nlm.nih.gov/31753276/)

Burden for Provider: Was a provider workflow analysis conducted?

Yes

If yes, how many sites were evaluated in the provider workflow analysis?

3,224

Did the provider workflow have to be modified to accommodate the new measure?

Yes

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Other (enter here): Inter-rater reliability

Signal-to-Noise: Name of statistic

N/A

Signal-to-Noise: Sample size

N/A

Signal-to-Noise: Statistical result

N/A

Signal-to-Noise: Interpretation of results

N/A

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

Cohen's Kappa

Other: Sample size

1,008

Other: Statistical result

Within social domains, percentages reporting a social risk tended to be higher by the AHC than the YCLS. Using unadjusted kappas, the AHC and YCLS items had substantial agreement for measures of food insecurity only. When examining the adjusted kappas that account for bias and prevalence, agreement between the AHC and YCLS items was substantial or higher (kappas > 0.60) for all social risks except housing quality (kappa = 0.52). The YCLS and CHW had substantial agreement (kappa 0.75) on housing.

Other: Interpretation of results

These results are the first to suggest that both the AHC and YCLS have high reliability and concurrent and predictive validity, supporting their use in healthcare settings, including by primary care physicians to engage in social risk-informed care.

Source:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652127/

Empiric Validity

Yes

Empiric Validity: Statistic name

Adjusted odds ratio

Empiric Validity: Sample size

30,098

Empiric Validity: Statistical result

1.60

Empiric Validity: Methods and findings

Study 1: A reported social risk on the AHC and YCLS measures was strongly associated with having fair or poor self-rated health

Source:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652127

Study 2: HFSS questions 1 and 2 were most frequently endorsed among food-insecure families (92.5% and 81.9%, respectively). An affirmative response to either question 1 or 2 had a sensitivity of 97% and specificity of 83% and was associated with increased risk of reported poor/fair child health (adjusted odds ratio [aOR]: 1.56; P < .001), hospitalizations in their lifetime (aOR: 1.17; P < .001), and developmental risk (aOR: 1.60; P < .001).

Source:

https://pubmed.ncbi.nlm.nih.gov/20595453/

Study 3: Sensitivity of each two-item combination was high for the US population and high-risk demographic groups compared with the eighteen-item CFSM (Table 2). Sensitivity ranged from 96.4% for items 2 and 3 for households with children and incomes <200 % of the federal poverty line, to 99.8% for items 1 and 3 for Spanish-speaking households. (results for all combinations are available from the corresponding author upon request). Specificity was lower, ranging from 73.7% for items 1 and 2 for households with children and incomes <100 % of the federal poverty line, to 94.5% for items 2 and 3 for households with a respondent aged >60 years. Accuracy was high for all two-item combinations.

Source:

https://www.cambridge.org/core/journals/public-health-nutrition/article/brief-assessment-of-food-insecurity-accurately-identifies-highrisk-us-adults/81A4F5E162241E289A5181A10C056125

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

7

Face Validity: Result

7

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Agreement between two survey instruments in a randomized sample of patients receiving health plan premium subsidies.

Sample Size

1,008

Statistic Name

Kappa

Statistical Results

0.6

Interpretation of results

These results are the first to suggest that both the AHC and YCLS have high reliability and concurrent and predictive validity, supporting their use in healthcare settings, including by primary care physicians to engage in social risk-informed care.

Source:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652127/

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

0.18

Median performance score

0.18

Minimum performance score

0.18

Maximum performance score

0.18

Standard deviation of performance scores

0

Does the performance measure use survey or patient-reported data?

Yes

Surveys or patient-reported outcome tools

The measure by design is not prescriptive about the use of specific tools to establish patients who have been screened for the HRSNs listed in the denominator. However, several tools, including the AHC

screening tool, are available nationally and have been designed and tested across several modes of administration, including phone, electronic, and paper.

Section 5: Measure Contact Information

Measure Steward

OCHIN

Measure Steward Contact Information

Ned Mossman

PO Box 5426

Portland, Oregon 97228

mossmann@ochin.org

(503) 943-5946

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

OCHIN was one of the first entities to enable EHR-based documentation and review of patient DOH screening in 2016. As of May 2022, OCHIN members have collected over 1.2 million DOH screenings, with than 40-50,000 new screenings added each month. These members, largely FQHCs, Public Health Departments, and Rural Health Centers, have engaged in screening and action on their patients' DOH because they understand acutely the consequences these factors have on their patients' ability to achieve optimal health. For example, OCHIN-led research has demonstrated that even with equivalent care, individuals who experience social risks still fare worse in key cardiovascular disease outcomes.

In COVID-19's wake, food insecurity, housing instability, IPV, and other basic DOH have reached unprecedented levels and revealed searing racial disparities. In 2021, 21% of Black individuals were projected to experience food insecurity, compared to 11% of white individuals. Likewise, 22% of Asian, 22% of Black, and 20% of Latino renters are not caught up on rent, compared to 9% of white renters.

In its 2022 Strategic plan, CMS placed screening for and acting on health-related social needs as a key goal underpinning its strategic health equity pillar. To that end, Administrator Brooks-LaSure has

charged the whole of CMS with building health equity into its core work across all programs, and HHS Secretary Becerra has pledged "to take a department-wide approach to the advancement of equity, consistent with President Biden's charge to federal departments and agencies, and this would include examination of ways to address the social determinants of health." In particular, Secretary Berra has noted the importance of collecting more robust DOH data to address the disparities exposed by COVID-19 and leveraging the data and experience from the CMMI Accountable Health Community (AHC) model, which has screened nearly one million beneficiaries.

CMS has recognized the importance of making DOH measures standard across programs, identifying the development and implementation of "measures that reflect social and economic determinants" as a key priority and measurement gap to be addressed through Meaningful Measures 2.0.

A growing set of constituencies have called on CMS to provide leadership in measuring and addressing DOH, citing various rationales for doing so. Healthcare experts have increasingly recognized that equity is unachievable without addressing DOH, calling for CMS to require program "participants to uniformly screen for and document drivers of health" and "build DOH measures into MIPS and all APMs." The Health Care Payment Learning & Action Network (LAN) -- a group of public and private health care leaders providing thought leadership, strategic direction, and ongoing support to accelerate adoption of APMs -- has identified promoting equity and addressing DOH as key facets of APM resiliency.

Likewise, physicians and other providers have called on CMS to create standard patient-level DOH measures -- beyond socioeconomic status (SES), hierarchical condition category (HCC) score, or duals status -- recognizing that these risk factors transcend specific subpopulations; drive demand for healthcare services; escalate physician burnout; and penalize physicians caring for those patients via worse Merit-based Incentive Payment System (MIPS) scores.

Together with the DOH screening measures submitted to the MUC process in 2021 (MUC 2021-134 and 2021-136), measures on connection to CSPs and resolution of HRSNs represent a crucial and necessary step to create a collective construct of measures built on experience from existing CMS programs and leveraging existing evaluation and measurement work to demonstrate the feasibility and validity of the approach.

Sources:

https://www.sciencedirect.com/science/article/pii/S0749379722001672

https://www.feedingamerica.org/sites/default/files/2021-03/National%20Projections%20Brief_3.9.2021_0.pdf

https://www.cbpp.org/research/poverty-and-inequality/tracking-the-covid-19-recessions-effects-on-food-housing-and

https://www.cms.gov/sites/default/files/2022-04/Health%20Equity%20Pillar%20Fact%20Sheet_1.pdf

https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization

https://blueshieldcafoundation.org/sites/default/files/publications/downloadable/Investing%20in%20Health%20-%20A%20Federal%20Action%20Plan%20-January%202021_Final.pdf

https://www.healthaffairs.org/do/10.1377/hblog20201216.672904/full/

https://hcp-lan.org/apm-measurement-effort/2020-2021-apm/

https://physiciansfoundation.org/wp-content/uploads/2020/11/PF-QPP-Open-Comment-Submission-v.f_-.pdf

https://pubmed.ncbi.nlm.nih.gov/27942709/

https://physiciansfoundation.org/wp-content/uploads/2020/10/2020-Physicians-Foundation-Survey-Part3.pdf

https://pubmed.ncbi.nlm.nih.gov/30610144/

https://pubmed.ncbi.nlm.nih.gov/32897345/

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Number of pts 18 or older who report that at least 1 of their HRSNs was resolved within 12 months after screening (annually)

Numerator Exclusions

N/A

Denominator

Number of pts 18 or older who screened positive for at least 1 of the 5 HRSN domains (food insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety) in the 12 months prior to the period of performance (annually)

Denominator Exclusions

Patients who opt-out of connection with Community Service Provider

Denominator Exceptions

N/A

State of development

Field (Beta) Testing

State of Development Details

- Using a standard, validated screening tool, the CMS Accountable Healthcare Communities program has screened nearly 1 million patients for HRSN in 21 states, with 33% of beneficiaries screened having at least one HRSN.
- Of patients with at least one HRSN who were eligible for navigation, 74% of patients accepted navigation related to their HRSN.
- 18% of patients accepting navigation either reported at least one HRSN resolved (14%) or connection with a CSP without resolution (4%).

Sources: :

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

- The HRSA Health Center Controlled Network (HCCN) grant cycle beginning in Q3 2022 will measure the number of participating FQHCs, RHCs, and lookalikes using social risk factor data to inform care plan development and facilitate closed-loop community referrals for at least 75% of patients identified as having a risk factor.
- According to BPHC, 70% of HRSA-funded Health Centers nationwide were enrolled in an HCCN for the 2019-2022 grant cycle.

Sources:

https://apply07.grants.gov/apply/opportunities/instructions/PKG00269508-instructions.pdf

https://bphc.hrsa.gov/program-opportunities/hccn/fy2022-faqs

Medicare and CMMI are currently exploring quality measures related to identifying and addressing social needs to support several ACO-related initiatives.

Source:

https://www.nejm.org/doi/full/10.1056/NEJMp2202991

What is the target population of the measure?

All payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Primary care

Measure Type

Intermediate Outcome

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

Social and Economic Determinants

What data sources are used for the measure?

Administrative Data (non-claims);Electronic Clinical Data (non-EHR);Electronic Health Record;Standardized Patient Assessments;Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

Patient reported data and standardized assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSNs.

EHR-and non-EHR electronic clinical data, as well as patient reported data, will be used to determine whether contact was made with a CSP.

Administrative data will be used for measure stratification and ongoing performance monitoring.

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care;Behavioral health clinic;Community hospital;Emergency department;Federally qualified health center (FQHC);Hospital outpatient department (HOD);Hospital inpatient acute care facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Equity

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

The measure correlates to specific MIPS Quality Improvement Activities as follows:

- Use QDCR data for ongoing practice assessment and improvements (IA_PSPA_7)
- Use of toolsets or other reSources: to close healthcare disparities in communities (IA_PM_6)
- Practice Improvements that Engage Community ReSources: to Support Patient Health (IA_CC_14)
- Provide Clinical-Community Linkages (IA_PM_18)

Source:

https://qpp.cms.gov/mips/explore-measures?tab=improvementActivities&py=2022

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

Measure currently used in a CMS program being submitted as-is for a new or different program

Range of years this measure has been used by CMS Programs

Accountable Health Communities Pilot (2017-2022)

What other federal programs are currently using this measure?

Accountable Health Communities Pilot

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

Yes (enter here):: We strongly recommend stratifying the measures by race/ethnicity.

- Data from the AHC found racial/ethnic minorities were over-represented in the navigation-eligible groups.

- CMS has stated in its strategic plan that the imperative to stratify by race/ethnicity is a global issue for the Agency that applies to all measures.

Sources: :

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

https://www.cms.gov/sites/default/files/2022-04/Health%20Equity%20Pillar%20Fact%20Sheet_1.pdf

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

- OCHIN's nationwide membership of Community Health Centers has documented over 1.2 million HRSN screenings as structured data in its centralized EHR record.
- Further, CMS has the opportunity to leverage and apply CMMI's 5+ years of data and experience with AHC. Using a standard, validated screening tool, AHC has screened nearly 1 million beneficiaries for HRSN in 21 states, with 33% of beneficiaries screened having at least one HRSN. AHC used screening, referral, and navigation data files extracted by NewWave (Centers for Medicare & Medicaid Services [CMS] Enterprise Portal contractor) and generated by

Mathematica Policy Research (the AHC implementation contractor) using data submitted by bridge organizations.

Sources:

https://innovation.cms.gov/innovation-models/ahcm

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

A number of CMMI models and participating entities have incorporated DOH screening and navigation data into their quality frameworks and care management plans for beneficiaries. CMMI's Comprehensive Primary Care Plus (CPC+) model reported in 2020 that 86% of ~1,500 Track 1 practices and 99% of ~1,500 Track 2 practices (together serving ~2.4M beneficiaries) are implementing DOH screening. CMMI required that by Program Year 3, Track 2 practices would use an electronic screening tool to assess patients' health-related social needs and store an inventory of reSources: to meet patients' needs; notably, by Program Year 2, Track 1 practices were as likely as Track 2 practices to report implementing these DOH functions, even absent a requirement that they do so.

Source:

https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2

Likewise, annual evaluations of other current CMMI models, including the State Innovation Model and Next Generation ACOs, report that participants are investing in staffing and infrastructure to conduct DOH screening and navigation. The 2021 Comprehensive End-Stage Renal Disease Care Model evaluation, for example, reported that "[m]any beneficiaries are protein malnourished and don't eat enough fresh produce. Some beneficiaries go to the hospital to get meals." ESRD Seamless Care Organizations have begun to monitor food insecurity and provide food gift cards to both low-income beneficiaries and those above the poverty level, to address beneficiaries' non-adherence to nutritional guidelines and reduce the risk of increased utilization and costs.

Sources:

https://downloads.cms.gov/files/cmmi/sim-rd2-test-ar3.pdf

https://innovation.cms.gov/data-and-reports/2020/nextgenaco-thirdevalrpt-fullreport

https://innovation.cms.gov/data-and-reports/2021/cec-annrpt-py4

Finally, a recent Physician's Foundation survey on provider well-being and attitudes towards social drivers of health found that 90% of providers want greater means and reSources: to address DOH in their practices.

Source:

https://physiciansfoundation.org/wp-content/uploads/2022/03/SDOH-Survey-Report.pdf

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Other digital method;Hybrid: Manual abstraction

Evidence of Performance Gap

CMS has already identified social and economic determinants as both a measurement priority and gap in Meaningful Measures 2.0, and as a central part of its Health Equity strategic plan pillar moving forward. Other public and private organizations such as ASPE, NQF and NCQA have identified this as a critical gap.

Sources:

https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization

https://aspe.hhs.gov/pdf-report/second-impact-report-to-congress

https://www.cms.gov/sites/default/files/2022-04/Health%20Equity%20Pillar%20Fact%20Sheet_1.pdf

https://www.qualityforum.org/News_And_Resources/Press_Releases/2019/National_Quality_Forum_Leads_National_Call_to_Address_Social_Determinants_of_Health__through_Quality_and_Payment_Innovation.aspx

https://blog.ncqa.org/ncqa-releases-its-social-determinants-of-health-reSource: -guide/

Unintended Consequences

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: . This challenge was noted as a primary barrier to connecting beneficiaries to resources in the AHC Year 1 evaluation. There is a well-documented and well-tested catalog of additional tools, infrastructure, and investments that can be implemented to support practices in acting on this measure.

Sources: :

https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/Unintended%20consequences%20of%20screening%20for%20social%20determinants.pdf

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

https://nhchc.org/wp-content/uploads/2020/04/NHCHC_Community-Information-Exchange2.pdf

https://governor.nc.gov/news/north-carolina-creates-nation%E2%80%99s-first-statewide-infrastructure-connecting-healthcare-and-human

https://blueshieldcafoundation.org/sites/default/files/publications/downloadable/Investing%20in%20Health%20-%20A%20Federal%20Action%20Plan%20-January%202021_Final.pdf

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

1

Outline the clinical guidelines supporting this measure

The USPSTF provides a "B" recommendation that recommends that clinicians screen for Intimate Partner Violence (one of the HRSNs included in the denominator of the proposed measure) in women of

reproductive age and provide or refer women who screen positive to ongoing support services. They cite adequate evidence that available screening instruments can identify IPV in women, and that screening for IPV in women of reproductive age and providing or referring women who screen positive to ongoing support services has a moderate net benefit.

However, in addition to this individual measure, USPSTF also released recently released a technical brief on screening and interventions for social risk factors which notes 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. The report also highlights the lack of unintended consequences encountered during implementation of social risk screening and intervention in studies reporting these outcomes, despite any perceived barriers.

Sources:

https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/intimate-partner-violence-and-abuse-of-elderly-and-vulnerable-adults-screening

https://uspreventiveservicestaskforce.org/uspstf/sites/default/files/inline-files/Social%20Risk%20Factors%20Tech%20Brief_Assembled%20for%20Web_Sep%202021_1.pdf

Name the guideline developer/entity

U.S. Preventive Services Task Force (USPSTF)

Publication year

2018

Full citation +/- URL

JAMA. 2018;320(16):1678-1687. doi:10.1001/jama.2018.14741
https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/intimate-partner-violence-and-abuse-of-elderly-and-vulnerable-adults-screening

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

The USPSTF recommends that clinicians screen for intimate partner violence (IPV) in women of reproductive age and provide or refer women who screen positive to ongoing support services.

What evidence grading system did the guideline use to describe strength of recommendation?

USPSTF

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

A. The USPSTF recommends the service. There is high certainty that the net benefit is substantial. Offer or provide this service.

B. The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. Offer or provide this service.

C. The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferemces. There is at least moderate certainty that the net benefit is small. Offer or provide this service for selected patients depending on individual circumstances.

D. The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. Discourage the use of this service.

I. The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

USPSTF Grade B or D, Moderate recommendation or similar

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

USPSTF

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

A. The USPSTF recommends the service. There is high certainty that the net benefit is substantial. Offer or provide this service.

B. The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. Offer or provide this service.

C. The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small. Offer or provide this service for selected patients depending on individual circumstances.

D. The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. Discourage the use of this service.

I. The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

USPSTF Grade B or D, Moderate recommendation or similar

List the guideline statement that most closely aligns with the measure concept.

The USPSTF recommends that clinicians screen for intimate partner violence (IPV) in women of reproductive age and provide or refer women who screen positive to ongoing support services.

Number of systematic reviews that inform this measure concept

3

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

See attached document with supporting evidence and research.

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies)

Summarize the empirical data

- Using a standard, validated screening tool, the CMS Accountable Healthcare Communities program has screened nearly 1 million patients for HRSN in 21 states, with 33% of beneficiaries screened having at least one HRSN.
- Of patients with at least one HRSN who were eligible for navigation, 74% of patients accepted navigation related to their HRSN.
- 18% of patients accepting navigation either reported at least one HRSN resolved (14%) or connection with a CSP without resolution (4%).

Sources: :

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

CMMI's Comprehensive Primary Care Plus (CPC+) model reported in 2020 that 86% of ~1,500 Track 1 practices and 99% of ~1,500 Track 2 practices (together serving ~2.4M beneficiaries) are implementing DOH screening.

Source:

https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines;Peer-Reviewed Systematic Review;Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through stratification of results

Cost estimate completed

Yes

Cost estimate methods and results

The cost avoided to payers and providers annually is likely to be significant given the research demonstrating increased utilization, readmissions, cost and increased financial liability for providers caring for patients with increased social risk.

Sources: :

https://www.gsfb.org/wp-content/uploads/2018/08/Berkowitz-S.-A.-Basu-S.-Meigs-J.-B.-Seligman-H.-K.-Food-Insecurity-and-Health-Care.pdf

https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2020.01742

https://pubmed.ncbi.nlm.nih.gov/32897345/

For example, extensive research exists demonstrating increased healthcare expenditures to patients including Medicare beneficiaries associated with DOH. The example below provides the annualized increase in annual healthcare expenditures (PMPY in 2015 dollars) associated with food insecurity across different disease categories across all payor types in the peer-reviewed literature:

- Diabetes Mellitus: $4,413.61
- Hypertension: $2,175.20
- Heart Disease: $5,144.05
- Overall: $1,863

Source:

https://www.gsfb.org/wp-content/uploads/2018/08/Berkowitz-S.-A.-Basu-S.-Meigs-J.-B.-Seligman-H.-K.-Food-Insecurity-and-Health-Care.pdf

The AHC Year 1 evaluation found that Medicare FFS beneficiaries in the Assistance Track intervention group had 9% fewer ED visits than those in the control group in the first year after screening. (No Medicaid utilization/cost data reported yet.)

Source:

https://innovation.cms.gov/data-and-reports/2020/ahc-first-eval-rpt

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

3,162

Meaningful to Patients: Number indicating survey/tool is meaningful

2,441

Meaningful to Clinicians: Numbers consulted

10,078

Meaningful to Clinicians: Number indicating survey/tool is meaningful

8,800

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

4

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

4

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Other (enter here):: Predictive and Empirical Validity

Testing methodology and results

Empirical validity (through AHC and CPC+ practice implementation across 3+ million beneficiaries over last ~ 5-year time frame) and Psychometric and Pragmatic Property Analysis (see https://pubmed.ncbi.nlm.nih.gov/31753276/)

Burden for Provider: Was a provider workflow analysis conducted?

Yes

If yes, how many sites were evaluated in the provider workflow analysis?

3,224

Did the provider workflow have to be modified to accommodate the new measure?

Yes

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Other (enter here): Inter-rater reliability

Signal-to-Noise: Name of statistic

N/A

Signal-to-Noise: Sample size

N/A

Signal-to-Noise: Statistical result

N/A

Signal-to-Noise: Interpretation of results

N/A

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

Cohen's Kappa

Other: Sample size

1,008

Other: Statistical result

Within social domains, percentages reporting a social risk tended to be higher by the AHC than the YCLS. Using unadjusted kappas, the AHC and YCLS items had substantial agreement for measures of food insecurity only. When examining the adjusted kappas that account for bias and prevalence, agreement between the AHC and YCLS items was substantial or higher (kappas > 0.60) for all social risks except housing quality (kappa = 0.52). The YCLS and CHW had substantial agreement (kappa 0.75) on housing.

Other: Interpretation of results

These results are the first to suggest that both the AHC and YCLS have high reliability and concurrent and predictive validity, supporting their use in healthcare settings, including by primary care physicians to engage in social risk-informed care.

Source:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652127/

Empiric Validity

Yes

Empiric Validity: Statistic name

Adjusted odds ratio

Empiric Validity: Sample size

30,098

Empiric Validity: Statistical result

1.60

Empiric Validity: Methods and findings

Study 1: A reported social risk on the AHC and YCLS measures was strongly associated with having fair or poor self-rated health

Source:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652127

Study 2: HFSS questions 1 and 2 were most frequently endorsed among food-insecure families (92.5% and 81.9%, respectively). An affirmative response to either question 1 or 2 had a sensitivity of 97% and specificity of 83% and was associated with increased risk of reported poor/fair child health (adjusted odds ratio [aOR]: 1.56; P < .001), hospitalizations in their lifetime (aOR: 1.17; P < .001), and developmental risk (aOR: 1.60; P < .001).

Source:

https://pubmed.ncbi.nlm.nih.gov/20595453/

Study 3: Sensitivity of each two-item combination was high for the US population and high-risk demographic groups compared with the eighteen-item CFSM (Table 2). Sensitivity ranged from 96.4% for items 2 and 3 for households with children and incomes <200 % of the federal poverty line, to 99.8% for items 1 and 3 for Spanish-speaking households. (results for all combinations are available from the corresponding author upon request). Specificity was lower, ranging from 73.7% for items 1 and 2 for households with children and incomes <100 % of the federal poverty line, to 94.5% for items 2 and 3 for households with a respondent aged >60 years. Accuracy was high for all two-item combinations.

Source:

https://www.cambridge.org/core/journals/public-health-nutrition/article/brief-assessment-of-food-insecurity-accurately-identifies-highrisk-us-adults/81A4F5E162241E289A5181A10C056125

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

7

Face Validity: Result

7

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Agreement between two survey instruments in a randomized sample of patients receiving health plan premium subsidies.

Sample Size

1,008

Statistic Name

Kappa

Statistical Results

0.6

Interpretation of results

These results are the first to suggest that both the AHC and YCLS have high reliability and concurrent and predictive validity, supporting their use in healthcare settings, including by primary care physicians to engage in social risk-informed care.

Source:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652127/

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

0.18

Median performance score

0.18

Minimum performance score

0.18

Maximum performance score

0.18

Standard deviation of performance scores

0

Does the performance measure use survey or patient-reported data?

Yes

Surveys or patient-reported outcome tools

The measure by design is not prescriptive about the use of specific tools to establish patients who have been screened for the HRSNs listed in the denominator. However, several tools, including the AHC

screening tool, are available nationally and have been designed and tested across several modes of administration, including phone, electronic, and paper.

Section 5: Measure Contact Information

Measure Steward

OCHIN

Measure Steward Contact Information

Ned Mossman

PO Box 5426

Portland, Oregon 97228

mossmann@ochin.org

(503) 943-5946

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

OCHIN was one of the first entities to enable EHR-based documentation and review of patient DOH screening in 2016. As of May 2022, OCHIN members have collected over 1.2 million DOH screenings, with than 40-50,000 new screenings added each month. These members, largely FQHCs, Public Health Departments, and Rural Health Centers, have engaged in screening and action on their patients' DOH because they understand acutely the consequences these factors have on their patients' ability to achieve optimal health. For example, OCHIN-led research has demonstrated that even with equivalent care, individuals who experience social risks still fare worse in key cardiovascular disease outcomes.

In COVID-19's wake, food insecurity, housing instability, IPV, and other basic DOH have reached unprecedented levels and revealed searing racial disparities. In 2021, 21% of Black individuals were projected to experience food insecurity, compared to 11% of white individuals. Likewise, 22% of Asian, 22% of Black, and 20% of Latino renters are not caught up on rent, compared to 9% of white renters.

In its 2022 Strategic plan, CMS placed screening for and acting on health-related social needs as a key goal underpinning its strategic health equity pillar. To that end, Administrator Brooks-LaSure has

charged the whole of CMS with building health equity into its core work across all programs, and HHS Secretary Becerra has pledged "to take a department-wide approach to the advancement of equity, consistent with President Biden's charge to federal departments and agencies, and this would include examination of ways to address the social determinants of health." In particular, Secretary Berra has noted the importance of collecting more robust DOH data to address the disparities exposed by COVID-19 and leveraging the data and experience from the CMMI Accountable Health Community (AHC) model, which has screened nearly one million beneficiaries.

CMS has recognized the importance of making DOH measures standard across programs, identifying the development and implementation of "measures that reflect social and economic determinants" as a key priority and measurement gap to be addressed through Meaningful Measures 2.0.

A growing set of constituencies have called on CMS to provide leadership in measuring and addressing DOH, citing various rationales for doing so. Healthcare experts have increasingly recognized that equity is unachievable without addressing DOH, calling for CMS to require program "participants to uniformly screen for and document drivers of health" and "build DOH measures into MIPS and all APMs." The Health Care Payment Learning & Action Network (LAN) -- a group of public and private health care leaders providing thought leadership, strategic direction, and ongoing support to accelerate adoption of APMs -- has identified promoting equity and addressing DOH as key facets of APM resiliency.

Likewise, physicians and other providers have called on CMS to create standard patient-level DOH measures -- beyond socioeconomic status (SES), hierarchical condition category (HCC) score, or duals status -- recognizing that these risk factors transcend specific subpopulations; drive demand for healthcare services; escalate physician burnout; and penalize physicians caring for those patients via worse Merit-based Incentive Payment System (MIPS) scores.

Together with the DOH screening measures submitted to the MUC process in 2021 (MUC 2021-134 and 2021-136), measures on connection to CSPs and resolution of HRSNs represent a crucial and necessary step to create a collective construct of measures built on experience from existing CMS programs and leveraging existing evaluation and measurement work to demonstrate the feasibility and validity of the approach.

Sources:

https://www.sciencedirect.com/science/article/pii/S0749379722001672

https://www.feedingamerica.org/sites/default/files/2021-03/National%20Projections%20Brief_3.9.2021_0.pdf

https://www.cbpp.org/research/poverty-and-inequality/tracking-the-covid-19-recessions-effects-on-food-housing-and

https://www.cms.gov/sites/default/files/2022-04/Health%20Equity%20Pillar%20Fact%20Sheet_1.pdf

https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization

https://blueshieldcafoundation.org/sites/default/files/publications/downloadable/Investing%20in%20Health%20-%20A%20Federal%20Action%20Plan%20-January%202021_Final.pdf

https://www.healthaffairs.org/do/10.1377/hblog20201216.672904/full/

https://hcp-lan.org/apm-measurement-effort/2020-2021-apm/

https://physiciansfoundation.org/wp-content/uploads/2020/11/PF-QPP-Open-Comment-Submission-v.f_-.pdf

https://pubmed.ncbi.nlm.nih.gov/27942709/

https://physiciansfoundation.org/wp-content/uploads/2020/10/2020-Physicians-Foundation-Survey-Part3.pdf

https://pubmed.ncbi.nlm.nih.gov/30610144/

https://pubmed.ncbi.nlm.nih.gov/32897345/

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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 fluociniolone 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.

Numerator

Number of patients 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 listed in chart OR if the IOP was >25 mm Hg, a plan of care was documented.

Plan of care includes one of the following: placement on IOP lowering medication (i.e., placement on a new medication, change in frequency or dose of an existing medication, or re-prescribing/renewing an existing medication), order for or performance of a IOP lowering procedure, referral to eye care provider for management of elevated IOP, or return within 4 weeks for IOP re-check.

Notes:

For patients who receive more than one injection during the measurement period (12 months), screening only needs to occur once to meet the numerator.

Tonometry with documented IOP should occur for the same eye that was injected.

Numerator Exclusions

none

Denominator

Patients who had an intravitreal or periocular corticosteroid injection (e.g., triamcinolone, preservative-free triamcinolone, dexamethasone, dexamethasone intravitreal implant, or fluociniolone intravitreal implant) with a patient encounter during the measurement period.

Denominator Exclusions

Patients with a diagnosis of hypotony.

Denominator Exceptions

Lost to follow-up

Patients who received a prior intravitreal or periocular steroid injection within the last six (6) months and had a subsequent IOP evaluation with IOP<25mm Hg within seven (7) weeks of treatment

State of development

Field (Beta) Testing

State of Development Details

ASRS believes these measures meet the minimum requirements outlined by CMS to be considered for inclusion in the Merit-based Incentive Payment System (MIPS). Specifically, we completed accountable entity reliability testing at the individual clinician level and a systematic face validity assessment by 15 physicians. Because these measures rely on the same data elements used for the previous version submitted in 2020, we also provided the data element validity results from the first round of testing as they further demonstrate that the measures will produce valid results. ASRS was unable to complete empiric validity testing at the accountable entity level since these are newly developed measures and we were unable to identify any existing and related measures against which valid comparisons could be made. We will likely continue to struggle to identify measures that could be used to complete this additional step in testing given the limited number of ophthalmology-specific measures in MIPS. Unfortunately, the MERIT tool does not provide any opportunity for us to justify why this empiric validity testing could not be completed and required that we select Field (beta) testing for the state of development. We hope that this additional context ensures that these measures are not rejected due to the inability to select Fully developed as the state of development.

What is the target population of the measure?

All Payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Ophthalmology

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Paper Medical Records

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

ASRS was unable to identify any applicable cost measures. ASRS believes that there are linkages to the following IAs: IA_ PSPA_7: Use of QCDR data for ongoing practice assessment and improvements, IA_PSPA_16: Use of decision support and standardized treatment protocols and IA_ CC_13: Practice Improvements for Bilateral Exchange of Patient Information These IAs could be used in conjunction with this measure to ensure that tools, pathways, reminders and outreach are used in the management of this condition, that bilateral exchange of information occurs, and to enable feedback and timely tracking of patient care through the use of QCDR data.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

8072

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

ASRS previously assessed the feasibility of collecting the required data elements of a similar measure across three practices with two different EHRs. The majority of the required data elements for this measure were found to be feasible (see feasibility scorecard results for specifics). The additional testing of this updated measure further demonstrated that two practices were able to collect and report the required data elements.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Testing using data from 1/1/2021-12/31/2021 across 19 physicians in two practices demonstrated that performance varied from 25.00-100.00%. While the mean performance score was 71.38%, the results were derived from retina specialists -whose care is more likely aligned with evidence-based recommendations as opposed to non-fellowship general ophthalmologists who still routinely administer intra-/periocular steroid treatments as part of their care. Because of this, we believe that wider sampling will demonstrate a clearer gap in care and resulting sufficient performance variation to enable benchmarking of this measure within the MIPS program.

Unintended Consequences

No unintended consequences were identified during testing of this measure. ASRS will continue to monitor whether any are identified during implementation of the measure.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research

Summarize the empirical data

While current clinical guidelines do not address the need to assess for elevated IOP following corticosteroid injection, a systematic review completed by Kiddee and colleagues (2013) identified that 10.9% to 79.0% of these patients will develop clinically significant IOP elevations with the large variation in incidence dependent largely on the specific steroid utilized and dose administered. The timing of IOP elevation also varies based on the type and dose; although, the available literature consistently shows IOP peaking in the 4-8 week range following injection with higher and earlier elevations following intravitreal triamcinolone injections as compared to intravitreal dexamethasone implants. This review recommended that IOP be assessed every two weeks in the first month and monthly for an additional six months at a minimum. Well-designed randomized controlled trials also support initial follow-up of no later than seven weeks. The Standard of Care vs. Corticosteroid for Retinal Vein Occlusion (SCORE) study where pressures peaked within 52.5 days following 4 mg intravitreal triamcinolone acetonide injection and the GENEVA study examining the effectiveness of dexamethasone intravitreal injections saw IOP peak within 60 days (Haller, 2010; Aref, 2015). For patients with a diagnosis of glaucoma, these symptoms can occur earlier and we would expect the follow up timeframe would occur sooner such as within the first four weeks following the injection (Vie, 2017).

Reference:

Aref AA, Scott IU, Oden NL, Ip MS, Blodi BA, VanVeldhuisen PC for the SCORE Study Investigator Group. Incidence, risk factors, and timing of elevated intraocular pressure after intravitreal triamcinolone acetonide injection for macular edema secondary to retinal vein occlusion. SCORE study report 15. JAMA Ophthalmol. 2015;133:1022-1029.

Haller JA, Bandello F, Belfort R, et al. for the OZURDEX GENEVA Study Group. Randomized, sham-controlled trial of dexamethasone intravitreal implant in patients with macular edema due to retinal vein occlusion. Opthalmol. 2010;117:1134-1146.

Kiddee W, Trope GE, Sheng L, et al. Intraocular pressure monitoring post intravitreal steroids: a systematic review. Survey Ophthalmol. 2013;58:291-310.

Vie AL, Kodjikian L, Malcles A, et al. Tolerance of intravitreal dexamethasone implants in patients with ocular hypertension or open-angle glaucoma. Retina 2017 Jan;37(1):173-178. doi: 10.1097/IAE.0000000000001114.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through exclusions (e.g., process measures)

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Beta-binomial model at the clinician level

Signal-to-Noise: Sample size

17

Signal-to-Noise: Statistical result

0.828

Signal-to-Noise: Interpretation of results

Physicians with at least five eligible cases were included in performance score reliability testing and a median reliability of 0.828 suggests good reliability; a reliability > 0.70 is generally considered adequate reliability.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

No

Empiric Validity: Statistic name

N/A

Empiric Validity: Sample size

N/A

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

N/A

Empiric Validity: Interpretation of results

N/A

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

15

Face Validity: Result

14

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between eCQM and manual reviewer

Sample Size

80

Statistic Name

Kappa

Statistical Results

0.78

Interpretation of results

This measure was updated to include a plan of care for those individuals whose intraocular pressure was not controlled (>25 mm Hg). While ASRS has not yet validated the data elements for a plan of care, previous testing of a similar eCQM provides information on the other critical data elements used in this measure. The overall reliability of the EHR extract vs. manual abstraction resulted in a prevalence adjusted kappa of 0.81 (95% CI: 0.66 to 0.96) for the denominator and a prevalence adjusted kappa of 0.77 (95% CI: 0.44-0.79) for the numerator. Eligible encounter: n = 80; % agreement = 100%; prevalence adjusted kappa = n/a Intravitreal steroid injection: n = 80; % agreement = 97.5%; prevalence adjusted kappa = 0.95 (95% CI: 0.88-1.0) Tonometry w/in 7 weeks: n = 80; % agreement = 82.5%; prevalence adjusted kappa = 0.95 (95% CI: 0.64-0.92) Lost to follow-up (exception): n = 80; % agreement = 96.3%; prevalence adjusted kappa = 0.95 (95% CI: 0.84-1.0) There was at least "substantial" agreement between the EHR retrospective data report and an independent review of a sample of patient medical records when comparing denominator criteria and numerator criteria based on calculation of prevalence adjusted kappa.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

71.38

Median performance score

80.77

Minimum performance score

25

Maximum performance score

100

Standard deviation of performance scores

23.07

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

American Society of Retina Specialists

Measure Steward Contact Information

Allison Madson
20 N Wacker Dr., Ste 2030
Chicago, Illinois 60606
allison.madson@asrs.org
(312) 578-8760

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

N/A

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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

Numerator

Number of patients who: had 1) a vitreous examination AND 2) peripheral dilated examination with documentation of scleral depression of the affected eye or contact lens (e.g., 3-mirror Goldmann) that provides visualization to the ora for 360 degrees OR if the retina cannot be adequately visualized, then ultrasound was performed OR was referred to another provider for additional examination (e.g., if retina cannot be visualized and ultrasound is not available) AND were re-evaluated no later than 8 weeks from the initial examination with: 1) a vitreous examination AND 2) adequate dilated examination to evaluate the peripheral retina for tears or detachment OR if the retina cannot be adequately visualized, then ultrasound was performed OR was referred to another provider for additional examination (e.g., if retina cannot be visualized and ultrasound is not available)

Numerator Exclusions

none

Denominator

Patients with a diagnosis of acute PVD in either eye and eligible encounter during measurement period

Note:

Acute can be captured as new onset vitreous separation or vitreous detachment. For the purposes of this measure acute PVD is PVD with recent onset of 30 days or less.

Denominator Exclusions

Patients with a post-operative encounter of the eye with the acute PVD within 2 weeks before the initial encounter or 8 weeks after initial acute PVD encounter Patients with a diagnosis of acute vitreous hemorrhage Note: The diagnosis of vitreous hemorrhage includes any degree of vitreous hemorrhage rather than "meaningful" vitreous hemorrhage since it is difficult to quantify and no criteria exist. If a patient is diagnosed with vitreous hemorrhage, it is assumed that it is meaningful. Vitreous hemorrhage should occur at the same time as PVD and/or have an onset of 30 days or less to ensure vitreous hemorrhage is acute and not chronic.

Denominator Exceptions

Lost to follow-up

State of development

Field (Beta) Testing

State of Development Details

ASRS believes these measures meet the minimum requirements outlined by CMS to be considered for inclusion in the Merit-based Incentive Payment System (MIPS). Specifically, we completed accountable entity reliability testing at the individual clinician level and a systematic face validity assessment by 15 physicians. Because these measures rely on the same data elements used for the previous version submitted in 2020, we also provided the data element validity results from the first round of testing as they further demonstrate that the measures will produce valid results. ASRS was unable to complete empiric validity testing at the accountable entity level since these are newly developed measures and we were unable to identify any existing and related measures against which valid comparisons could be made. We will likely continue to struggle to identify measures that could be used to complete this additional step in testing given the limited number of ophthalmology-specific measures in MIPS. Unfortunately, the MERIT tool does not provide any opportunity for us to justify why this empiric validity testing could not be completed and required that we select Field (beta) testing for the state of development. We hope that this additional context ensures that these measures are not rejected due to the inability to select Fully developed as the state of development.

What is the target population of the measure?

all payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Ophthalmology

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Paper Medical Records

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

ASRS was unable to identify any applicable cost measures. ASRS believes that there are linkages to the following IAs: IA_BE_15: Engagement of Patients, Family, and Caregivers in Developing a Plan of Care IA_PSPA_7: Use of QCDR data for ongoing practice assessment and improvements These IAs could be used in conjunction with this measure to ensure that patients are engaged in the management of acute episodes of this condition, particularly by leveraging EHR technology and to enable feedback and timely tracking of patient care through the use of QCDR data.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

ASRS previously assessed the feasibility of collecting the required data elements for this measure across three practices with two different EHRs. The majority of the required data elements for this measure were found to be feasible (see feasibility scorecard results for specifics). The additional testing of this updated measure further demonstrated that two practices were able to collect and report the required data elements.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Testing using data from 1/1/2021-12/31/2021 across 19 physicians in two practices demonstrated that performance varied from 0.00 to 5.31%. Performance scores are generally low for this measure.

Unintended Consequences

No unintended consequences were identified during testing of this measure. ASRS will continue to monitor whether any are identified during implementation of the measure.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

1

Outline the clinical guidelines supporting this measure

This measure is based on a guideline developed by the American Academy of Ophthalmology (AAO). This guideline was used given that it directly discusses the appropriate initial evaluation and follow-up of patients with an acute posterior vitreous detachment by clinicians.

AAO's Preferred Practice Pattern guidelines are based on the best available scientific data as interpreted by panels of knowledgeable health professionals. In some instances, such as when results of carefully conducted clinical trials are available, the data are particularly persuasive and provide clear guidance. In other instances, the panels have to rely on their collective judgment and evaluation of available evidence. Preferred Practice Pattern guidelines should be clinically relevant and specific enough to provide useful information to practitioners. Where evidence exists to support a recommendation for care, the recommendation should be given an explicit rating that shows the strength of evidence. To accomplish these aims, methods from the Scottish Intercollegiate Guideline Network (SIGN) and the

Grading of Recommendations Assessment, Development and Evaluation (GRADE) group are used. GRADE is a systematic approach to grading the strength of the total body of evidence that is available to support recommendations on a specific clinical management issue. Organizations that have adopted GRADE include SIGN, the World Health Organization, the Agency for Healthcare Research and Policy, and the American College of Physicians.

All studies used to form a recommendation for care are graded for strength of evidence individually, and that grade is listed with the study citation.

To rate individual studies, a scale based on SIGN is used. The definitions and levels of evidence to rate individual studies are outlined within the PPP publication.

Name the guideline developer/entity

American Academy of Ophthalmology

Publication year

2019

Full citation +/- URL

American Academy of Ophthalmology Retina/Vitreous Preferred Practice Pattern Panel. Preferred Practice Pattern Guidelines. Posterior Vitreous Detachment, Retinal Breaks, and Lattice Degeneration PPP 2019. San Francisco, CA: American Academy of Ophthalmology; 2019. Available at: https://www.aao.org/ppp

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

The eye examination should include the following elements:

Examination of the vitreous for hemorrhage, detachment, and pigmented cells.

Careful examination of the peripheral fundus using scleral depression.

There are no symptoms that can reliably distinguish between a PVD with or without an associated retinal break; therefore, a peripheral retinal examination is required. The preferred method of evaluating patients for peripheral vitreoretinal pathology is to use an indirect ophthalmoscope combined with scleral depression. Many patients with retinal tears have blood and pigmented cells in the anterior vitreous. In fully dilated eyes, slit-lamp biomicroscopy with a mirrored contact lens or a condensing lens is an alternative method in fully dilated eyes instead of a scleral depressed indirect examination of the peripheral retina.

Follow-up Evaluation: The guidelines in Table 3 are recommendations for the timing of re-evaluation in the absence of additional symptoms. Patients with new symptoms or a change in symptoms may require

more frequent evaluation. Patients with no positive findings at the initial examination should be seen at the intervals recommended in the Comprehensive Adult Medical Eye Evaluation PPP. All patients with risk factors should be advised to contact their ophthalmologist promptly if new symptoms such as flashes, floaters, peripheral visual field loss, or decreased visual acuity develop.

Type of Lesion - Symptomatic PVD with no retinal break.

Follow-up Interval - Depending on symptoms, risk factors, and clinical findings, patients may be followed within 2 months, then 6 to12 months.

What evidence grading system did the guideline use to describe strength of recommendation?

GRADE method

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

Strong recommendation: Used when the desirable effects of an intervention clearly outweigh the undesirable effects or clearly do not Discretionary recommendation: Used when the trade-offs are less certain - either because of low-quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced.

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

Other (enter here):: The grading of evidence was strong for the initial exam recommendations and discretionary for the follow-up examination recommendations in the 2014 AAO Preferred Practice Pattern Guidelines. Posterior Vitreous Detachment, Retinal Brea

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

GRADE method

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

Strong recommendation: Used when the desirable effects of an intervention clearly outweigh the undesirable effects or clearly do not Discretionary recommendation: Used when the trade-offs are less certain - either because of low-quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced.

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

Other (enter here):

The grading of evidence was strong for the initial exam recommendations and discretionary for the follow-up examination recommendations in the 2014 AAO Preferred Practice Pattern Guidelines. Posterior Vitreous Detachment, Retinal Breaks, and Lattice Degeneration. The evidence Citations: remain the same in the 2019 release.

List the guideline statement that most closely aligns with the measure concept.

The eye examination should include the following elements:

Examination of the vitreous for hemorrhage, detachment, and pigmented cells.

Careful examination of the peripheral fundus using scleral depression.

There are no symptoms that can reliably distinguish between a PVD with or without an associated retinal break; therefore, a peripheral retinal examination is required. The preferred method of evaluating patients for peripheral vitreoretinal pathology is to use an indirect ophthalmoscope combined with scleral depression. Many patients with retinal tears have blood and pigmented cells in the anterior vitreous. In fully dilated eyes, slit-lamp biomicroscopy with a mirrored contact lens or a condensing lens is an alternative method in fully dilated eyes instead of a scleral depressed indirect examination of the peripheral retina.

Follow-up Evaluation: The guidelines in Table 3 are recommendations for the timing of re-evaluation in the absence of additional symptoms. Patients with new symptoms or a change in symptoms may require more frequent evaluation. Patients with no positive findings at the initial examination should be seen at the intervals recommended in the Comprehensive Adult Medical Eye Evaluation PPP. All patients with risk factors should be advised to contact their ophthalmologist promptly if new symptoms such as flashes, floaters, peripheral visual field loss, or decreased visual acuity develop.

Type of Lesion - Symptomatic PVD with no retinal break.

Follow-up Interval - Depending on symptoms, risk factors, and clinical findings, patients may be followed within 2 months, then 6 to12 months.

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through exclusions (e.g., process measures)

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Beta-binomial model at the clinician level

Signal-to-Noise: Sample size

19

Signal-to-Noise: Statistical result

0.978

Signal-to-Noise: Interpretation of results

Physicians with at least five eligible cases were included in performance score reliability testing and a median reliability of 0.978 suggests excellent reliability; a reliability > 0.70 is generally considered adequate reliability.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

No

Empiric Validity: Statistic name

N/A

Empiric Validity: Sample size

N/A

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

N/A

Empiric Validity: Interpretation of results

N/A

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

15

Face Validity: Result

13

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between eCQM and manual reviewer

Sample Size

74

Statistic Name

Kappa

Statistical Results

0.73

Interpretation of results

Previous testing of a similar eCQM provides information on the critical data elements used in this measure. The overall reliability of the EHR extract vs. manual abstraction 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. Denominator: Eligible encounter: n = 74; % agreement = 100%; prevalence adjusted kappa = n/a Acute PVD: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Numerator 1: Vitreous exam - initial: n = 74; % agreement = 98.6%; prevalence adjusted kappa = 0.97 (95% CI: 0.92-1.0) Peripheral exam initial: n = 74; % agreement = 94.6%; prevalence adjusted kappa = 0.89 (95% CI: 0.79-0.99)

Inadequate retina visualization - initial*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Ultrasound - initial: n = 74; % agreement = 98.6%; prevalence adjusted kappa = 0.97 (95% CI: 0.92-1.0) Referral - initial*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a *no cases documented Numerator 2: Vitreous exam - f/u: n = 74; % agreement = 86.5%; prevalence adjusted kappa = 0.73 (95% CI: 0.57-0.89) Peripheral exam - f/u: n = 74; % agreement = 91.9%; prevalence adjusted kappa = 0.84 (95% CI: 0.72-0.96) Inadequate retina visualization - f/u*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Ultrasound - f/u: n = 74; % agreement = 98.6%; prevalence adjusted kappa = 0.97 (95% CI: 0.92-1.0) Referral - f/u*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Patient lost to follow-up*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a *no cases documented There was "almost perfect" agreement between the EHR retrospective data report and an independent review of a sample of patient medical records when comparing denominator criteria and numerator criteria based on calculation of prevalence adjusted kappa.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

1

Median performance score

0.24

Minimum performance score

0

Maximum performance score

5.31

Standard deviation of performance scores

1.55

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

American Society of Retina Specialists

Measure Steward Contact Information

Allison Madson

20 N Wacker Dr., Ste 2030

Chicago, Illinois 60606

allison.madson@asrs.org

312-578-8760

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

N/A

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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

Numerator

Number of patients who: had 1) a vitreous examination AND 2) peripheral dilated examination with documentation of scleral depression of the affected eye or contact lens (e.g., 3-mirror Goldmann) that provides visualization to the ora for 360 degrees OR if the retina cannot be adequately visualized, then ultrasound was performed OR was referred to another provider for additional examination (e.g., if retina cannot be visualized and ultrasound is not available) AND were re-evaluated no later than 2 weeks from the initial examination with: 1) a vitreous examination AND 2) adequate dilated examination to evaluate the peripheral retina for tears or detachment OR if the retina cannot be adequately visualized, then ultrasound was performed OR was referred to another provider for additional examination (e.g., if retina cannot be visualized and ultrasound is not available)

Numerator Exclusions

none

Denominator

Patients with a diagnosis of acute PVD and acute vitreous hemorrhage in either eye and eligible encounter during measurement period Notes: Acute can be captured as new onset vitreous separation or vitreous detachment. For the purposes of this measure acute PVD is PVD with recent onset of 30 days or less. The measure includes any degree of vitreous hemorrhage rather than "meaningful" vitreous hemorrhage since it is difficult to quantify and no criteria exist. If a patient is diagnosed with vitreous hemorrhage, it is assumed that it is meaningful. Vitreous hemorrhage should occur at the same time as PVD and/or have an onset of 30 days or less to ensure vitreous hemorrhage is acute and not chronic.

Denominator Exclusions

Patients with a post-operative encounter of the eye with the acute PVD within 2 weeks before the initial encounter or 8 weeks after initial acute PVD encounter.

Denominator Exceptions

Lost to follow-up

State of development

Field (Beta) Testing

State of Development Details

ASRS believes these measures meet the minimum requirements outlined by CMS to be considered for inclusion in the Merit-based Incentive Payment System (MIPS). Specifically, we completed accountable entity reliability testing at the individual clinician level and a systematic face validity assessment by 15 physicians. Because these measures rely on the same data elements used for the previous version submitted in 2020, we also provided the data element validity results from the first round of testing as they further demonstrate that the measures will produce valid results. ASRS was unable to complete empiric validity testing at the accountable entity level since these are newly developed measures and we were unable to identify any existing and related measures against which valid comparisons could be made. We will likely continue to struggle to identify measures that could be used to complete this additional step in testing given the limited number of ophthalmology-specific measures in MIPS. Unfortunately, the MERIT tool does not provide any opportunity for us to justify why this empiric validity testing could not be completed and required that we select Field (beta) testing for the state of development. We hope that this additional context ensures that these measures are not rejected due to the inability to select Fully developed as the state of development.

What is the target population of the measure?

All payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Ophthalmology

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Paper Medical Records

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

ASRS was unable to identify any applicable cost measures. ASRS believes that there are linkages to the following IAs: IA_BE_15: Engagement of Patients, Family, and Caregivers in Developing a Plan of Care IA_PSPA_7: Use of QCDR data for ongoing practice assessment and improvements These IAs could be used in conjunction with this measure to ensure that patients are engaged in the management of acute episodes of this condition, particularly by leveraging EHR technology and to enable feedback and timely tracking of patient care through the use of QCDR data.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources.

Feasibility Assessment

ASRS previously assessed the feasibility of collecting the required data elements for this measure across three practices with two different EHRs. The majority of the required data elements for this measure were found to be feasible (see feasibility scorecard results for specifics). The additional testing of this updated measure further demonstrated that two practices were able to collect and report the required data elements.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Testing using data from 1/1/2021-12/31/2021 across 19 physicians in two practices demonstrated that performance varied from 0.00 to 38.10%. Performance scores are generally low for this measure.

Unintended Consequences

No unintended consequences were identified during testing of this measure. ASRS will continue to monitor whether any are identified during implementation of the measure.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

1

Outline the clinical guidelines supporting this measure

This measure is based on a guideline developed by the American Academy of Ophthalmology (AAO). This guideline was used given that it directly discusses the appropriate initial evaluation and follow-up of patients with an acute posterior vitreous detachment by clinicians. AAO's Preferred Practice Pattern guidelines are based on the best available scientific data as interpreted by panels of knowledgeable health professionals. In some instances, such as when results of carefully conducted clinical trials are

available, the data are particularly persuasive and provide clear guidance. In other instances, the panels have to rely on their collective judgment and evaluation of available evidence. Preferred Practice Pattern guidelines should be clinically relevant and specific enough to provide useful information to practitioners. Where evidence exists to support a recommendation for care, the recommendation should be given an explicit rating that shows the strength of evidence. To accomplish these aims, methods from the Scottish Intercollegiate Guideline Network (SIGN) and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) group are used. GRADE is a systematic approach to grading the strength of the total body of evidence that is available to support recommendations on a specific clinical management issue. Organizations that have adopted GRADE include SIGN, the World Health Organization, the Agency for Healthcare Research and Policy, and the American College of Physicians. All studies used to form a recommendation for care are graded for strength of evidence individually, and that grade is listed with the study citation. To rate individual studies, a scale based on SIGN is used. The definitions and levels of evidence to rate individual studies are outlined within the PPP publication.

Name the guideline developer/entity

American Academy of Ophthalmology

Publication year

2019

Full citation +/- URL

American Academy of Ophthalmology Retina/Vitreous Preferred Practice Pattern Panel. Preferred Practice Pattern Guidelines. Posterior Vitreous Detachment, Retinal Breaks, and Lattice Degeneration PPP 2019. San Francisco, CA: American Academy of Ophthalmology; 2019. Available at: https://www.aao.org/ppp.

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

The eye examination should include the following elements:

Examination of the vitreous for hemorrhage, detachment, and pigmented cells

Careful examination of the peripheral fundus using scleral depression.

There are no symptoms that can reliably distinguish between a PVD with or without an associated retinal break; therefore, a peripheral retinal examination is required. The preferred method of evaluating patients for peripheral vitreoretinal pathology is to use an indirect ophthalmoscope combined with scleral depression. Many patients with retinal tears have blood and pigmented cells in the anterior vitreous. In fully dilated eyes, slit-lamp biomicroscopy with a mirrored contact lens or a

condensing lens is an alternative method in fully dilated eyes instead of a scleral depressed indirect examination of the peripheral retina.

A spontaneous vitreous hemorrhage can be the presenting sign of PVD or may occur during the evolution of the PVD. Two-thirds of patients who present with associated vitreous hemorrhage were found to have at least one break. In this subgroup, one-third had more than one break and approximately 88% of the breaks occurred in the superior quadrants. If media opacity or patient cooperation precludes an adequate examination of the peripheral retina, B-scan ultrasonography should be performed to search for retinal tears, RRD, mass lesions, or other causes of vitreous hemorrhage. Bilateral patching and/or elevation of the head while sleeping may be used when attempting to clear the vitreous hemorrhage. If no abnormalities are found, frequent follow-up examinations are recommended (i.e., every 1 to 2 weeks initially). Wide-field color photography can detect some peripheral retinal breaks but does not replace careful ophthalmoscopy and may be useful in patients not able to tolerate the exam. Even if the vitreous hemorrhage is sufficiently dense to obscure the posterior pole, the peripheral retina frequently can be examined using indirect ophthalmoscopy and scleral depression. Patients who present with vitreous hemorrhage sufficient to obscure all retinal details and have a negative B-scan ultrasonographic evaluation should be followed closely. When a retinal tear is suspected, repeat ultrasonographic examination should be performed within 1 to 2 weeks of the initial evaluation.

What evidence grading system did the guideline use to describe strength of recommendation?

GRADE method

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

Strong recommendation: Used when the desirable effects of an intervention clearly outweigh the undesirable effects or clearly do not Discretionary recommendation: Used when the trade-offs are less certain - either because of low-quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

Other (enter here):

The grading of evidence was strong for the initial exam recommendations in the 2014 AAO Preferred Practice Pattern Guidelines. Posterior Vitreous Detachment, Retinal Breaks, and Lattice Degeneration. The evidence Citations: remain the same in the 2019 release.

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

GRADE method

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

Strong recommendation: Used when the desirable effects of an intervention clearly outweigh the undesirable effects or clearly do not Discretionary recommendation: Used when the trade-offs are less

certain - either because of low-quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

Other (enter here):

The grading of evidence was strong for the initial exam recommendations in the 2014 AAO Preferred Practice Pattern Guidelines. Posterior Vitreous Detachment, Retinal Breaks, and Lattice Degeneration. The evidence Citations: remain the same in the 2019 release.

List the guideline statement that most closely aligns with the measure concept.

The eye examination should include the following elements:

Examination of the vitreous for hemorrhage, detachment, and pigmented cells

Careful examination of the peripheral fundus using scleral depression.

There are no symptoms that can reliably distinguish between a PVD with or without an associated retinal break; therefore, a peripheral retinal examination is required. The preferred method of evaluating patients for peripheral vitreoretinal pathology is to use an indirect ophthalmoscope combined with scleral depression. Many patients with retinal tears have blood and pigmented cells in the anterior vitreous. In fully dilated eyes, slit-lamp biomicroscopy with a mirrored contact lens or a condensing lens is an alternative method in fully dilated eyes instead of a scleral depressed indirect examination of the peripheral retina.

A spontaneous vitreous hemorrhage can be the presenting sign of PVD or may occur during the evolution of the PVD. Two-thirds of patients who present with associated vitreous hemorrhage were found to have at least one break. In this subgroup, one-third had more than one break and approximately 88% of the breaks occurred in the superior quadrants. If media opacity or patient cooperation precludes an adequate examination of the peripheral retina, B-scan ultrasonography should be performed to search for retinal tears, RRD, mass lesions, or other causes of vitreous hemorrhage. Bilateral patching and/or elevation of the head while sleeping may be used when attempting to clear the vitreous hemorrhage. If no abnormalities are found, frequent follow-up examinations are recommended (i.e., every 1 to 2 weeks initially). Wide-field color photography can detect some peripheral retinal breaks but does not replace careful ophthalmoscopy and may be useful in patients not able to tolerate the exam. Even if the vitreous hemorrhage is sufficiently dense to obscure the posterior pole, the peripheral retina frequently can be examined using indirect ophthalmoscopy and scleral depression. Patients who present with vitreous hemorrhage sufficient to obscure all retinal details and have a negative B-scan ultrasonographic evaluation should be followed closely. When a retinal tear is suspected, repeat ultrasonographic examination should be performed within 1 to 2 weeks of the initial evaluation.

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through exclusions (e.g., process measures)

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Beta-binomial model at the clinician level

Signal-to-Noise: Sample size

18

Signal-to-Noise: Statistical result

0.973

Signal-to-Noise: Interpretation of results

Physicians with at least five eligible cases were included in performance score reliability testing and a median reliability of 0.973 suggests excellent reliability; a reliability > 0.70 is generally considered adequate reliability.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

No

Empiric Validity: Statistic name

N/A

Empiric Validity: Sample size

N/A

Empiric Validity: Statistical result

N/A

Empiric Validity: Methods and findings

N/A

Empiric Validity: Interpretation of results

N/A

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

15

Face Validity: Result

13

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between eCQM and manual reviewer

Sample Size

74

Statistic Name

Kappa

Statistical Results

0.73

Interpretation of results

Previous testing of a similar eCQM provides information on the critical data elements used in this measure. The overall reliability of the EHR extract vs. manual abstraction 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. Denominator: Eligible encounter: n = 74; % agreement = 100%; prevalence adjusted kappa = n/a Acute PVD: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Vitreous hemorrhage: n = 40; % agreement = 97.5%; prevalence adjusted kappa = 0.95 Numerator 1: Vitreous exam - initial: n = 74; % agreement = 98.6%; prevalence adjusted kappa = 0.97 (95% CI: 0.92-1.0) Peripheral exam - initial: n = 74; % agreement = 94.6%; prevalence adjusted kappa = 0.89 (95% CI: 0.79-0.99) Inadequate retina visualization - initial*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Ultrasound initial: n = 74; % agreement = 98.6%; prevalence adjusted kappa = 0.97 (95% CI: 0.92-1.0) Referral - initial*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a *no cases documented Numerator 2: Vitreous exam f/u: n = 74; % agreement = 86.5%; prevalence adjusted kappa = 0.73 (95% CI: 0.57-0.89) Peripheral exam- f/u: n = 74; % agreement = 91.9%; prevalence adjusted kappa = 0.84 (95% CI: 0.72-0.96) Inadequate retina visualization - f/u*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Ultrasound - f/u: n = 74; % agreement = 98.6%; prevalence adjusted kappa = 0.97 (95% CI: 0.92-1.0) Referral - f/u*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a Patient lost to follow-up*: n = 74; % agreement = 100.0%; prevalence adjusted kappa = n/a *no cases documented There was "almost perfect" agreement between the EHR retrospective data report and an independent review of a sample of patient medical records when comparing denominator criteria and numerator criteria based on calculation of prevalence adjusted kappa.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

7.9

Median performance score

1.22

Minimum performance score

0

Maximum performance score

38.10

Standard deviation of performance scores

10.96

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

American Society of Retina Specialists

Measure Steward Contact Information

Allison Madson

20 N Wacker Dr., Ste 2030

Chicago, Illinois 60606

allison.madson@asrs.org

312-578-8760

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

N/A

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

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Individuals who demonstrated improvement or maintenance of functioning, as demonstrated by results of a follow-up assessment using the 12-item WHODAS 2.0 or Sheehan Disability Index 30 to 180 days after the index assessment during, the measurement period.

Numerator Details:

Improvement or maintenance: Improvement is defined as any positive improvement in score at the follow-up assessment compared to the index assessment. Maintenance is defined as no change in score at the follow-up assessment compared to the index assessment.

Follow-up Assessment: The follow-up assessment is the 12-item WHODAS 2.0 or Sheehan Disability Index (SDS) assessment completed at an encounter 30 to 180 days after the encounter with the index assessment, within the measurement period. If there are multiple assessments completed within the follow-up time window (i.e. a 150-day window, beginning 30 days after the index assessment), the assessment that will be counted as the follow-up is the last assessment completed during the window.

12-item WHODAS 2.0: WHODAS 2.0 assesses change-over-time in functioning for all individuals with mental health and/or substance use disorders. The domains covered in the tool are communication and understanding, mobility, self-care, social functioning, life activities (work and home), and participation in society. Response options include: (0) None, (1) Mild, (2) Moderate, (3) Severe, and (4) Extreme or Cannot Do. A 12-item and 36-item version of the WHODAS 2.0 are available. Summed scores on the 12-item and 36-item WHODAS 2.0 are converted to a summary scale from 0 to 100 (where 0 = no disability; 100 = full disability). There is no recommended cutoff score. A higher score on the WHODAS 2.0 equates to a lower level of functioning.

Sheehan Disability Index (SDS): SDS assesses change-over-time in functioning for individuals with mental health and/or substance use disorders. The domains covered in the tool are work/school, social life/leisure activities, and family life/home responsibilities. Response options include: (0) Not at all, (1-3) Mildly, (4-6) Moderately, (7-9) Markedly, and (10) Extremely, regarding how current symptoms have

disrupted activities in each of the domains covered by the assessment. The 3 items are summed into a single dimensional measure of global functioning from 0 to 30 (where 0 = unimpaired and 30 = highly impaired). There is no recommended cutoff score. A higher score on the SDS equates to a lower level of functioning.

Index Assessment: The outpatient encounter where the individual first completed the WHODAS 2.0 or SDS was counted as the baseline assessment. If there were multiple assessments during the measurement period, the first assessment completed during the denominator identification period was counted as the baseline assessment.

Measurement Period: A 15-month period starting 3 months prior to the measurement year through the end of the measurement year.

Numerator Exclusions

N/A

Denominator

Individuals aged 18 and older with a mental and/or substance use disorder and an encounter with an index assessment completed using the 12-item WHODAS 2.0 or Sheehan Disability Index (SDS) during the denominator identification period.

Denominator Details:

Age Range: Individuals aged 18 and older as of the date of the index encounter.

Codes Used to Identify Diagnoses: 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), any Mental, Behavioral, and Neurodevelopmental disorder diagnosis, F01-99

Codes Used to Identify Encounter: Current Procedural Terminology (CPT) codes: 99201-99205, 99211-99215, 99241-99245, 90791-90792, 90785, 90832-90840, 90845, 90847, 90849, 90853, 90880, 90875, 90876, 90901, 90911-90913, 99441-99444, 90865, 96112-96113, 96116, 96118, 96120, 96121, 96125, 96127, 96130-96133, 96136-96137, 99401-99409

Index assessment: The outpatient encounter where the individual first completed the WHODAS 2.0 or SDS was counted as the baseline assessment. If there were multiple assessments during the measurement period, the first assessment completed during the denominator identification period was counted as the baseline assessment.

WHODAS 2.0: WHODAS 2.0 assesses change-over-time in functioning for all individuals with mental health and/or substance use disorders. The domains covered in the tool are communication and understanding, mobility, self-care, social functioning, life activities (work and home), and participation in society. Response options include: (0) None, (1) Mild, (2) Moderate, (3) Severe, and (4) Extreme or Cannot Do. A 12-item and 36-item version of the WHODAS 2.0 are available. Summed scores on the 12-item and 36-item WHODAS 2.0 are converted to a summary scale from 0 to 100 (where 0 = no disability;

100 = full disability). There is no recommended cutoff score. A higher score on the WHODAS 2.0 equates to a lower level of functioning.

SDS: SDS assesses change-over-time in functioning for individuals with mental health and/or substance use disorders. The domains covered in the tool are work/school, social life/leisure activities, and family life/home responsibilities. Response options include: (0) Not at all, (1-3) Mildly, (4-6) Moderately, (7-9) Markedly, and (10) Extremely, regarding how current symptoms have disrupted activities in each of the domains covered by the assessment. The 3 items are summed into a single dimensional measure of global functioning from 0 to 30 (where 0 = unimpaired and 30 = highly impaired). There is no recommended cutoff score. A higher score on the SDS equates to a lower level of functioning.

Denominator identification period: Period in which individuals had an encounter with a baseline assessment using the WHODAS 2.0 or SDS. The denominator identification period was defined by a 12-month window starting 4 months prior to the measurement year through the first 8 months of the measurement year.

Denominator Exclusions

1. Patients whose functional capacity or motivation (or lack thereof) to improve may impact the accuracy of results of validated tools.

2. Patients deceased during the measurement period.

Denominator exclusions included situations where the patient's functional capacity or motivation (or lack thereof) to improve may impact the accuracy of results of validated tools, such as acute medical conditions, delirium, dementia, intellectual, and development disorders. A patient met criteria for exclusion if there were documentation of an exclusion diagnosis at any point during the denominator intake period or the patient's date of death occurred within the measurement year.

ICD-10-CM codes used to identify denominator exclusions:

F00-09: Mental disorders due to known physiological conditions

F70-79: Intellectual disabilities

F80-89: Developmental Disorders

Patient death: Value of TRUE in the vital status indicator = patient death during the measurement year.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Adults 18 and over, all payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Psychiatry

Measure Type

Outcome - (PRO-PM)

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Patient Reported Data and Surveys;Registries

If applicable, specify the data source

Other:

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Behavioral Health

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

This measurement can be linked to a number of Improvement Activities (IA) including those covering screening and follow-up for certain conditions - e.g., MDD prevention and treatment intervention (IA-BMH-5), Depression Screening (IA-BMH-4) and Unhealthy Alcohol Use (IA-BMH-9). Measurement-based care (MBC) processes encompass screening and assessment of patients and include the use of the information for patient engagement and follow-up care. Other links to IA include Collaborative Care Management Training Program (IA-BMH-10), which has as a focus the implementation of MBC in the treatment and care of those with mental health and substance use disorders. Further, Electronic Health Record Enhancements for BH data capture (IA-BMH-8) can be linked to this MBC process measure primarily because electronic data capture of patient reported assessments provides for an efficient and effective means of reducing the burden with implementation of MBC. Links to cost measures include

certain impacts on Medicare Beneficiary Spending (MSPB-1) that may be expected to occur primarily post inpatient stay with outpatient follow-up care to prevent subsequent preventable hospitalizations.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

06126

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

Submitted previously but not included in MUC List

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

NQF 0422: Functional Status Change for Patients with Knee Impairments

NQF 0423: Functional Status Change for Patients with Hip Impairments

NQF 0424: Functional Status Change for Patients with Lower Leg, Foot or Ankle Impairments

NQF 0426: Functional Status Change for Patients with Shoulder Impairments

NQF 0427: Functional Status Change for Patients with Elbow, Wrist or Hand Impairments

NQF 0428: Functional status change for patients with General orthopaedic impairments

NQF 0429: Change in Basic Mobility as Measured by the AM-PAC:

NQF 0430: Change in Daily Activity Function as Measured by the AM-PAC:

NQF 0688: Percent of Long-Stay Residents Whose Need for Help with Daily Activities Has Increased

NQF 0700: Functional Status Change for Patients with Low Back Impairments

2286: Functional Change: Change in Self Care Score

2287: Functional Change: Change in Motor Score

2321: Functional Change: Change in Mobility Score

2612: CARE: Improvement in Mobility

2613: CARE: Improvement in Self Care

NQF 2632: Functional Outcome Measure: Change in Mobility Among Long-Term Care Hospital (LTCH) Patients Requiring Ventilator Support

NQF 2633: Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients

NQF 2634: IRF Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients

NQF 2635: IRF Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients

NQF 2636: IRF Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients

2769: Functional Change: Change in Self Care Score for Skilled Nursing Facilities

2774: Functional Change: Change in Mobility Score for Skilled Nursing Facilities

2775: Functional Change: Change in Motor Score for Skilled Nursing Facilities

2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities

2777: Functional Change: Change in Self Care Score for Long Term Acute Care Facilities

2778: Functional Change: Change in Mobility Score for Long Term Acute Care Facilities

2958: Average change in functional status following lumbar spine fusion surgery

2962: Average change in functional status following total knee replacement surgery

How will this measure be distinguished from other similar and/or competing measures?

There are a range of NQF-endorsed measures assessing functional status, but most are focused on post-operative functional outcomes or on specific domains of functioning (e.g., Mobility, Self-Care) for patients in particular settings of care (e.g., Skilled Nursing Facilities, Long-Term Acute Care Facilities).

How will this measure add value to the CMS program?

There are significant gaps in measurement of mental/behavioral health care, including measurement of patient outcomes. This measure is intended to lay the groundwork for expanded measurement of patient outcomes, using patient-reported data on functional status and recovery, among other domains. Mental health clinicians have few measures in the MIPS program that are relevant to their practice; this measure presents an opportunity for these clinicians to report data to CMS that is directly relevant to their clinical practice.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

Yes (enter here):: Measure results are stratified by age category, sex, and majority mental health comorbidity (e.g., substance use, mood, and anxiety disorders).

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

Measure specification and logic were examined to determine the extent to which data are readily available or could be readily captured for performance measurement. Based on feedback from the alpha testing, 71% of the respondents answered the question "How easy was it for you to access PROMS to assess level of functioning (i.e., access the WHODAS, Sheehan Disability Scale, or other level of functioning scale)?" with moderately to extremely easy. For the question "On average, how easy was it for your patients to complete the WHODAS or other function scale?", 64% of the respondents thought that it is moderately to extremely easy as well. However, for those providers not employing an EHR, entry of relevant clinical data requires manual entry, which appears to be a significant barrier. EHR data requires a low burden for data collection at scale. However, accuracy and missingness in data elements is always a potential limitation. Data derived from psychiatric EHR data will often only encompass the patient's mental health diagnoses. Scores of medical illness burden (e.g., Charlson Comorbidity Index) are dependent on medical diagnosis information. As such, any estimation of medical illness burden is limited and may require reliance on proxy measures (e.g., medications). Given the known impact of chronic medical illness of mental health, a limited ability to control for medical illness burden is a limitation. Additionally, many EHRs separate billing and clinical information. As a result, the diagnostic information available is limited to 'problem lists' that are dependent on providers to reconcile. Encounter-level diagnosis information from billing would permit more refined assessment of acute illness states (e.g., delirium, intoxication). The challenges we encountered were primarily related to data collection in PsychPRO. The portal portion of the registry was set to collect WHODAS 2.0 data responses, however, it was observed that either the clinicians had difficulty implementing MBC in their practice or they prescribed the assessment measures, but the patients failed to complete them. The use of measurement-based care in clinical practice may require both considerable reSources: and changes to clinician workflow. During the testing of the MBC Process measure, participating practices using the PsychPRO registry required time and training to be oriented to both the PsychPRO system and the use of standardized assessment tools in practice. The measure development team provided technical assistance and education in the form of Learning Collaborative webinars, newsletters, and one-on-one consultation to practices. Clinicians were also provided with a clear explanation and rationale of the

recommended process of care being measured by the MBC Process measure. Additionally, the data collection period included the Coronavirus Disease 2019 (COVID-19) pandemic which impacted both aspects of the data collection. To adjust to these challenges, we used DSM-5 Field Trials data to supply additional data as detailed above.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

There are many available standardized assessments to monitor functioning in health care. Still, functioning assessment tools are used less frequently and less consistently than symptom severity scales (Evans & Lam, 2014) even though functional impairment is a key component of the diagnosis of a mental or substance use disorder per Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) (APA, 2013) and a better indicator of service needs, treatment outcomes, and quality care (Reed, Spaudling & Bufka, 2009; Kilbourne et al., 2018). Notably, a systematic search of over 90 depression treatment outcome meta-analyses revealed that less than 5% of clinical trials measure and report functioning outcomes despite the high economic burden in the US associated with direct costs specific to diagnosis and treatment ($2.1 billion) and indirect costs related to disability and premature mortality ($4.2 billion) (McKnight & Kashdan, 2009). From measure testing, performance (i.e., pass rate) was accessed only at the clinic/practice level. for the measure at the clinic level varied from 0.51 to 0.58. Due Performance on varies from 0.51 to 0.58 based on age, gender, and psychiatric comorbidity groupings of the test population.

Referemces:

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: Evans VC, Lam RW.

Assessments of functional improvement: self-versus clinician-ratings. Medicographia. 2014;36(4):512-520.

McKnight PE, Kashdan TB. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. Clin Psychol Rev. 2009;29(3):243-59.

Kilbourne AM, Beck K, Spaeth-Rublee B, Ramanuj P, O'Brien RW, Tamoyasu N, Pincus HA. Measuring and improving quality of mental health care: a global perspective. World Psychiatry, 2018; 17:30-38.

Unintended Consequences

No unintended consequences have been identified.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies)

Summarize the empirical data

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.

Guideline Recommendations:

APA-2021: 'Statement 2: Use of Quantitative Measures APA recommends (1C) that the initial psychiatric evaluation of a patient with a possible psychotic disorder include a quantitative measure to identify and determine the severity of symptoms and impairments of functioning that may be a focus of treatment'. Research support: low

APA-2016: Guideline VII. Quantitative Assessment Guideline Statements: APA suggests (2C) that the initial psychiatric evaluation of a patient include quantitative measures of symptoms, level of functioning, and quality of life. Research support: low VA/DOD-2017: 'Recommendation 4. For patients with suspected PTSD, we recommend an appropriate diagnostic evaluation that includes determination of DSM criteria, acute risk of harm to self or others, functional status, medical history, past treatment history, and relevant family history. A structured diagnostic interview may be considered.' Strong For

VA/DOD-2016: 'Recommendation 3. For patients with suspected depression, we recommend an appropriate diagnostic evaluation that includes a determination of functional status, medical history, past treatment history, and relevant family history'. Strong For

NICE-2018: '1.1.3 When assessing for PTSD, ask people specific questions about re experiencing, avoidance, hyperarousal, dissociation, negative alterations in mood and thinking, and associated functional impairment.'

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

Yes

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through stratification of results

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

12

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

0000

Meaningful to Patients: Numbers consulted

12

Meaningful to Patients: Number indicating survey/tool is meaningful

12

Meaningful to Clinicians: Numbers consulted

22

Meaningful to Clinicians: Number indicating survey/tool is meaningful

22

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency

Testing methodology and results

WHODAS 2.0 and SDS have established excellent internal consistency reliability (alpha coefficients 0.98, 0.89, respectively) in prior testing across a spectrum of mental illness diagnoses. References:

Bastiaens, L., Galus, J. & Goodlin, M. The 12 item WHODAS as primary self report outcome measure in a correctional community treatment center for dually diagnosed patients. Psychiatric Quarterly 86, 219-224 (2015).

Federici, S., Bracalenti, M., Meloni, F. & Luciano, J. V. World Health Organization disability assessment schedule 2.0: An international systematic review. Disability and rehabilitation 39, 2347-2380 (2017).

Saltychev, M., Katajapuu, N., Brlund, E. & Laimi, K. Psychometric properties of 12-item self-administered World Health Organization disability assessment schedule 2.0 (WHODAS 2.0) among general population and people with non-acute physical causes of disability - systematic review. Disability and rehabilitation 43, 789-794 (2021).

Leon, A. C., Olfson, M., Portera, L., Farber, L. & Sheehan, D. V. Assessing Psychiatric Impairment in Primary Care with the Sheehan Disability Scale. The International Journal of Psychiatry in Medicine 27, 93-105, doi:10.2190/T8EM-C8YH-373N-1UWD (1997).

Sheehan, K. H. & Sheehan, D. V. Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale. International clinical psychopharmacology 23, 70-83 (2008).

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

mean signal-to-noise reliability

Signal-to-Noise: Sample size

48

Signal-to-Noise: Statistical result

.82

Signal-to-Noise: Interpretation of results

Signal-to-noise reliability of 0.70 or greater is adequate for distinguishing the relative performance of one entity from another. The reliability of the performance measure at the clinic/practice level is above this a priori threshold.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Spearman's Rho

Empiric Validity: Sample size

15

Empiric Validity: Statistical result

0.44

Empiric Validity: Methods and findings

Correlation of measure performance to other valid indicators or conceptually related measures. We compared provider and site performance on the outcome measure to the provider and site-level performance on a conceptually related measure. The measure with provider and site-level data available for testing determined to be conceptually related was the Depression Remission at Six Months (NQF 0711) outcome measure. Improved/maintained functioning rates are expected to be positively

correlated with depressive remission rates. To assess whether this relationship held true, we tested the measure distributions for normality and assessed the significance of the Spearman's rank correlation coefficient. A Spearman's rho (0.10-0.19 = negligible correlation, 0.20-0.29 = weak, 0.30-0.39 = moderate, 0.40-0.69 = strong, >0.70 = very strong. The outcome measure is strongly positively correlated with the NQF 0711 Depressive Remission outcome measure at the provider level and strongly positively correlated at the site level, indicating strong convergent validity. For this analysis, only PsychPRO registry data were used, as the community psychiatry program dataset lacked necessary data elements to calculate performance on NQF 0711 measure. No minimum denominator threshold was applied at the provider or site level. Remission of depression and improvement/maintenance of functioning both indicate quality of care. A conceptually-related outcome measure for which we possessed data elements was not available to facilitate the examination of the rate's discriminant validity. Caution is noted in interpreting these results given the preponderance of zero performance rates among providers and sites for both measures.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

11

Face Validity: Result

9

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Other (enter here)

Sample Size

All records

Statistic Name

Other (enter here):: Other (enter here):

Statistical Results

Coverage % = 7.2 for Patient - Diagnosis - Encounter - Assessment (Numerator) linkage

Interpretation of results

Critical data elements demonstrated overall low missingness and high validity by individual element. However, when the required linkages for measure testing were examined, certain relationships demonstrate poor coverage.

See attached Testing Report (Functioning_TSR) for additional details.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

34.4

Median performance score

32

Minimum performance score

0

Maximum performance score

77.8

Standard deviation of performance scores

3.3

Does the performance measure use survey or patient-reported data?

Yes

Surveys or patient-reported outcome tools

12-item WHODAS 2.0

Section 5: Measure Contact Information

Measure Steward

American Psychiatric Association

Measure Steward Contact Information

Andrew Lyzenga
800 Maine Ave SW, Suite 900
DC, District of Columbia 20024
alyzenga@psych.org
(202) 744-9776

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

Please see Testing Report (Functioning_TSR) for full results of measure testing and analysis.

MUC2022-127 Initiation, Review, And/Or Update To Suicide Safety Plan For Individuals With Suicidal Thoughts, Behavior, Or Suicide Risk

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

NUMERATOR 1:

Individuals in the target population for whom a complete suicide safety plan is initiated, reviewed, or updated in collaboration between the individual and their clinician at the time the suicidal ideation behavior or risk is identified (concurrent or within 24 hours of index clinical encounter), during the measurement period.

NUMERATOR 2:

Individuals in the target population for whom a suicide safety plan is initiated, reviewed, or updated in collaboration between the individual and their clinician at the time the suicidal ideation, behavior or risk is identified (concurrent or within 24 hours of clinical encounter) (i.e., individuals who satisfy Numerator 1) AND reviewed and updated within 120 days after the index clinical encounter, during the measurement period.

Numerator Details:

Suicide safety plan: A brief intervention that involves the patient with suicidal ideation, behavior or risk and their clinician working in collaboration to identify and document: a written list of warning signs, internal coping strategies the patient can use to stay safe without involving others, Sources: of support (including access to professional services), and ways to make their environment safe.

Measurement Period: A 16-month period, starting 4 months prior to the measurement year through the 12 months of the measurement year.

Numerator Exclusions

N/A

Denominator

Individuals aged 18 and older with a mental and/or substance use disorder with suicidal ideation and/or behavior symptoms OR deemed a suicide risk based on their clinician's evaluation using the CRPSR or similar tool and have an encounter with an index assessment completed using the CSSRS during the denominator identification period.

Denominator Details:

Age Range: Individuals aged 18 and older as of the date of the baseline encounter.

Suicidal Ideation and/or Behavior Symptoms: Any non-zero score on the CSSRS or clinician determination of increased suicide risk.

Codes Used to Identify Mental and/or Substance Use Disorder Diagnoses: 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), any Mental, Behavioral, and Neurodevelopmental disorder diagnosis, F01-99.

Codes used to identify outpatient encounters:

99201-99205, 99211-99215, 99241-99245, 90791-90792, 90785, 90832-90840, 90845, 90847, 90849, 90853, 90880, 90875, 90876, 90901, 90911-90913, 99441-99444, 90865, 96112-96113, 96116, 96118, 96120, 96121, 96125, 96127, 96130-96133, 96136-96137, 99401-99409

Baseline Assessment: The encounter when the individual first completes the CSSRS was counted as the baseline assessment. If there are multiple assessments during the measurement period, the first assessment completed during the denominator identification period was counted as the baseline.

Denominator Identification Period: The period in which individuals can have an encounter with a baseline assessment using the CSSRS. The denominator encounter period is the 12-month window starting 4 months prior to the measurement year and ending 8 months into the measurement year.

Denominator Exclusions

Denominator exclusions included situations where the patient's functional capacity or motivation (or lack thereof) to improve may impact the accuracy of results of validated tools, such as acute medical conditions, delirium, dementia, intellectual, and development disorders. A patient met criteria for exclusion if there were documentation of an exclusion diagnosis at any point during the denominator intake period or the patient's date of death occurred within the measurement year. ICD-10-CM codes used to identify denominator exclusions (e.g., patients with reduced functional capacity): F00-09: Mental disorders due to known physiological conditions; F70-79: Intellectual disabilities; F80-89: Developmental Disorders Patient death: Value of TRUE in the vital status indicator = patient death during the measurement year.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Adults 18 and over, all payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Psychiatry

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Administrative Data (non-claims);Electronic Health Record;Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

The measure is calculated using administrative codes (ICD-10-CM, CPT) and patient assessment data derived from the EHR.

Providers collect information from patients via survey instrument; the CSSRS instrument may be included in EHRs or patient portals. Paper surveys can be uploaded to the EHR as PDFs. The registry then extracts data from EHRs and calculates the measure.

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

Yes

What one healthcare domain applies to this measure?

Behavioral Health

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

This measurement can be linked to a number of Improvement Activities (IA) including those covering screening and follow-up for certain conditions - e.g., MDD prevention and treatment intervention (IA-BMH-5), Depression Screening (IA-BMH-4) and Unhealthy Alcohol Use (IA-BMH-9). Measurement-based care (MBC) processes encompass screening and assessment of patients and include the use of the information for patient engagement and follow-up care. Other links to IA include Collaborative Care Management Training Program (IA-BMH-10), which has as a focus the implementation of MBC in the treatment and care of those with mental health and substance use disorders. Further, Electronic Health Record Enhancements for BH data capture (IA-BMH-8) can be linked to this measure primarily because electronic data capture of patient reported assessments provides for an efficient and effective means of reducing the burden with implementation of measurement-based care. Links to cost measures include certain impacts on Medicare Beneficiary Spending (MSPB-1) that may be expected to occur primarily post inpatient stay with outpatient follow-up care to prevent subsequent preventable hospitalizations.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

06122

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

NQF # 104e: Adult Major Depressive Disorder (MDD): Suicide Risk Assessment

NQF #1365e: Child and Adolescent Major Depressive Disorder (MDD): Suicide Risk Assessment

How will this measure be distinguished from other similar and/or competing measures?

The existing quality measures related to suicide screening focus only on mood disorders (e.g., Major Depressive Disorder and Bipolar Disorder), despite strong evidence that suicide risk is increased across all mental and substance use disorders as well as subthreshold mental and substance use conditions (APA, 2013). Furthermore, of the available suicide related quality measures, none address suicide safety

planning or outcomes. Therefore, development of process and outcome quality measures related to suicide prevention across a wider range of subthreshold and fully diagnosable mental and substance use disorders is imperative. Suicide screening alone is a good first step but should be used in conjunction with interventions that are evidence-based, such as SSP, which is reviewed and followed-up until the suicide risk is diminished and a reduction in suicidal ideation and behaviors attained (AFSP, 2018; National Action Alliance for Suicide Prevention, 2018). The proposed measures address the gap in quality measures related to the continuum of care and improvement in outcomes for individuals with suicidal ideation, behavior, or risk.

How will this measure add value to the CMS program?

There are significant gaps in measurement of mental/behavioral health care, including measurement of patient outcomes. This measure is intended to lay the groundwork for expanded measurement of patient outcomes, using patient-reported data on suicidal behavior and ideation. Mental health clinicians have few measures in the MIPS program that are relevant to their practice; this measure presents an opportunity for these clinicians to report data to CMS that is directly relevant to their clinical practice.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

Yes (enter here):: age, sex, and mental health comorbidities

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

Measure specification and logic were examined to determine the extent to which data are readily available or could be readily captured for performance measurement. Based on feedback from the alpha testing, the clinicians reported that the results of the assessment of suicidal thoughts and behaviors were slightly to moderately useful in helping them decide to complete the Suicide Safety Plan. They also reported that it was slightly to moderately easy for them to work with their patients on the SSP. The clinicians reported that they found it slightly to moderately useful to develop Suicide Safety Planning overall. EHR data requires a low burden for data collection at scale. However, accuracy and missingness in data elements is always a potential limitation. Data derived from psychiatric EHR data will often only encompass the patient's mental health diagnoses. Scores of medical illness burden (e.g., Charlson Comorbidity Index) are dependent on medical diagnosis information. As such, any estimation of medical illness burden is limited and may require reliance on proxy measures (e.g., medications). Given the known impact of chronic medical illness of mental health, a limited ability to control for medical illness

burden is a limitation. Additionally, many EHRs separate billing and clinical information. As a result, the diagnostic information available is limited to 'problem lists' that are dependent on providers to reconcile. Encounter-level diagnosis information from billing would permit more refined assessment of acute illness states (e.g., delirium, intoxication). The challenges we encountered were related to data collection in PsychPRO and getting the data from one of the three external data Sources: in a timely manner. The registry was set to collect suicide process measure data responses; however, it was observed that either the clinicians had difficulty implementing MBC in their practice or they prescribed the assessment measures, but the patients failed to complete them. Based on feedback from some of the clinicians, the use of measurement-based care in clinical practice requires both considerable resources and changes to clinic workflow, especially for small solo practices. During the alpha testing of the suicide process measure as well as testing of all measures in this initiative, participating practices using the PsychPRO registry required time and training to be oriented to both the PsychPRO system and the use of standardized assessment tools in practice. The measure development team provided technical assistance and education in the form of Learning Collaborative webinars, newsletters, and one-on-one consultation to practices. Clinicians were also provided with a clear explanation and rationale of the recommended process of care being measured by the Process measures. Additionally, the data collection period included the Coronavirus Disease 2019 (COVID-19) pandemic, which impacted both aspects of the data collection in PsychPRO as well as in one of the external data sources. To adjust to these challenges, we used external data sources to provide data and replace the registry data as detailed above.

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Other digital method

Evidence of Performance Gap

The uptake and frequency of use of standardized PROMs associated with suicide ideation and behavior and clinician-rated assessments of suicide risk, including the use of safety plans, varies within and across behavioral health specialties as well as primary and emergency care, where suicidal persons most often present for care(Waldrop & McGuiness, 2017; Harding et al, 2011; Kilbourne et al, 2010). Currently, only hard-copy versions of safety planning documents have been used in most settings, with slow uptake of electronic versions. Hard-copy safety plans provided to patients are prone to misplacement, creating a barrier to their use and follow-up(Little et al, 2018). Even with use of suicide safety plans at an index visit, Gamarra et al. found that less than 50% of suicidal persons had explicit evidence of ongoing review or utilization of the safety plan in ongoing treatments (Gamarra et al, 2015). The development and implementation of the proposed quality measures related to safety planning and the review, update (if necessary), and utilization of those plans over the course of treatment may incentivize quality care that addresses the low rate of (re)assessment and poor outcomes. These quality measures will help to advance the Zero Suicide initiative set forth in the 2012 National Strategy for Suicide Prevention (National Alliance for Suicide Prevention, 2018) and ultimately improve the quality of care for patients with suicide ideation, behaviors or suicide risk. From measure testing, performance (i.e., pass rate) on both indicators for the measure at the provider and clinic levels. Performance on INDICATOR 1 varies from 0 to 1.00 at the provider level and 0 to 0.91 at the clinic/practice level. These variabilities were

observed across age, gender, and psychiatric comorbidity groups in the patient population. For INDICATOR 2, performance varied from 0 to 0.60 at the provider and 0 to 0.43 at the clinic/practice levels and across age, gender, and psychiatric comorbidity in the patient population. Referemces: Waldrop & McGuinness. Measurement-Based Care in Psychiatry. J Psychosoc Nurs Ment Health Serv. 2017 Nov 1;55(11):30-35. Harding et al. Measurement-based care in psychiatric practice: a policy framework for implementation. J Clin Psychiatry. 2011 Aug;72(8):1136-43. Kilbourne et al. Implementing composite quality metrics for bipolar disorder: towards a more comprehensive approach to quality measurement. Gen Hosp Psychiatry. 2010 Nov-Dec;32(6):636-43. Little et al. Integrating Safety Plans for Suicidal Patients Into Patient Portals: Challenges and Opportunities. Psychiatric Services 2018; 69:618-619. Gamarra, et al. Assessing variability and implementation fidelity of suicide prevention safety planning in a regional VA healthcare system. 2015. Crisis, 36: 433-439. National Action Alliance for Suicide Prevention: Transforming Health Systems Initiative Work Group. Recommended Standard Care for People with Suicide Risk: Making Health Care Suicide Safe. Washington, DC: Education Development Center, Inc.; 2018.

Unintended Consequences

In public comments submitted on this measure, several respondents expressed concerns with the burden of administering the assessment tools specified in the measures, including the associated data entry and data collection requirements. Of note, one respondent representing a managed-care organization, noted the need for additional administrative staff to successfully implement the measures. Another respondent, identifying as a provider, echoed concerns about how measures may strain current clinical workflow and indicated the measures would be very challenging to use and report in practices or clinical settings with little or no administrative support. The implementation of measurement-based care (MBC) can require significant changes in practice for clinicians. MBC entails routine use of assessment instruments, which may not be part of providers' usual workflow, and can require a different mode of interaction with patients. Patients also need to adjust to the need for timely completion of patient-reported outcome measures (PROMs), and clinicians need to work closely with their patients to explain the purpose of assessment tools and how they will be used to inform and adjust treatment approaches. For these reasons, adoption of MBC may take several months and require multiple QI initiatives (e.g., PDSA cycles). APA and NCQA have conducted regular learning collaborative sessions during the development and testing of this measure set, providing technical assistance, answering questions, and working through challenges faced by participants. As MBC is more widely adopted as part of routine clinical practice, data collection difficulties are expected to become less of a barrier to implementation.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies)

Summarize the empirical data

Suicide safety planning (SSP), which involves counseling the individual with suicidal thoughts or behaviors around reducing access to lethal means, teaching brief problem-solving and coping skills, and helping the individual increase social support and identify emergency contacts (Boudreaux et al., 2017; Goodman et al., 2020; Hegerl, 2016; Stanley & Brown, 2012; Stanley et al., 2016), is effective and critical in suicide prevention, as echoed in recent clinical practice guidelines and recommendations from the Joint Commission and the National Action Alliance for Suicide (JC, 2018; Stanley et al., 2015). It has been identified as the best practice for suicide prevention by the ASFP and the Suicide Prevention ReSource: (NAASP, 2018). In fact, SSP has been found to be clinically useful and feasible by both suicidal individuals and clinicians and is associated with reduction in suicidal behaviors. Individuals with suicidal ideation and behaviors also report that the SSP helps them maintain their safety and increases the likelihood of remaining in care (Brodsky et al., 2018; Chesin et al., 2017). A 2017 randomized control trial (RCT) by Bryan and colleagues evaluated the effectiveness of crisis response planning for the prevention of suicide attempts among active-duty Army Soldiers (N=97) presenting for an emergency behavioral health appointment. Participants were randomly assigned to receive a contract for safety, a standard crisis response plan, or an enhanced crisis response plan. Crisis response planning was associated with a reduction in suicide attempts, more rapid decline in suicidal ideation, and fewer inpatient hospital days (Bryan et al, 2017). A 2018 study by Stanley, Brown, and colleagues used a cohort comparison approach to determine whether a Safety Planning Intervention (SPI) administered in EDs with follow-up contact for suicidal patients was associated with reduced suicidal behavior and improved outpatient treatment engagement in the 6 months following discharge. Patients receiving a SPI were less likely to engage in suicidal behavior than those in usual care, had approximately half the odds of suicidal behavior over 6 months, and more than double the odds of attending at least one outpatient mental health visit (Stanley et al, 2018). While the evidence base for suicide safety planning is still evolving, the compelling evidence of these interventions at helping to reduce suicide risk, suicide ideation and behavior and increasing patients' engagement in treatment outweighs the limited completed studies in the US in this area. However, the Zero Suicide Effort with its focus on suicide safety planning has stimulated a significant uptick of research efforts. Also, given the significant mortality, morbidity, and costs (e.g., ED visits, hospitalizations, reduced human capital, etc.) associated with suicide and suicide attempts, the results that have emerged from early studies of this intervention are extremely compelling, and show it to be one of the most promising approaches to suicide reduction that currently exist. Moreover, as opposed to other suicide prevention initiatives, which tend to be focused at the public health level (e.g., reduction in access to firearms and other lethal means, changes in packaging of medications, etc.), suicide safety planning is an intervention that can be implemented and measured at the individual clinician or clinical practice level.

References: Report: Evidence [Process] 1a.16)

Bryan CJ, Mintz J, Clemans TA, et al. Effect of crisis response planning vs. Contracts for safety on suicide risk in U.S. Army soldiers: A randomized clinical trial. J Affect Disord. Apr 1 2017;212:64-72.

Brodsky BS, Spruch-Feiner A, Stanley B. The zero suicide model: Applying evidence-based suicide prevention practices to clinical care. Front Psychiatry. 2018;9:33. doi:10.3389/fpsyt.2018.00033

Chesin MS, Stanley B, Haigh EA, et al. Staff views of an emergency department intervention using safety planning and structured follow-up with suicidal veterans. Arch Suicide Res. 2017;21(1):127-137.

Goodman M, et al. An Open Trial of a Suicide Safety Planning Group Treatment: "Project Life Force", Arch Suicide Res. 2020 Apr 14;1-14.

Little V, et al. Integrating Safety Plans for Suicidal Patients Into Patient Portals: Challenges and Opportunities. Psychiatric Services 2018; 69:618-619; doi: 10.1176/appi.ps.201700458

Skovgaard Larsen JL. MYPLAN - A Mobile Phone Application for Supporting People at Risk of Suicide. Crisis. 2016 May;37(3):236-40. doi: 10.1027/0227-5910/a000371. Epub 2016 Feb 2.

Stanley B, Chaudhury SR, Chesin M, et al. An emergency department intervention and follow-up to reduce suicide risk in the VA: Acceptability and effectiveness. Psychiatr Serv. Jun 1 2016;67(6):680-683.

Stanley B, Brown GK, Brenner LA, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. JAMA Psychiatry. 2018;75(9):894-900. doi:10.1001/jamapsychiatry.2018.1776

Zonana J, et al. The Impact of Safety Plans in an Outpatient Clinic. Crisis. 2018 Jul;39(4):304-309.

Boudreaux ED, Brown GK, Stanley B, Sadasivam RS, Camargo CA Jr, Miller IW. Computer Administered Safety Planning for Individuals at Risk for Suicide: Development and Usability Testing. J Med Internet Res. 2017 May 15;19(5):e149. doi: 10.2196/jmir.6816. PMID: 28506957; PMCID: PMC5447822.

Goodman M, Sullivan SR, Spears AP, Dixon L, Sokol Y, Kapil-Pair KN, Galfalvy HC, Hazlett EA, Stanley B. An Open Trial of a Suicide Safety Planning Group Treatment: "Project Life Force". Arch Suicide Res. 2021 Jul-Sep;25(3):690-703. doi: 10.1080/13811118.2020.1746940. Epub 2020 Apr 14. PMID: 32290789.

Hegerl U. Prevention of suicidal behavior. Dialogues Clin Neurosci. 2016 Jun;18(2):183-90. doi: 10.31887/DCNS.2016.18.2/uhegerl. PMID: 27489458; PMCID: PMC4969705.

Stanley, B., & Brown, G. K. (2012). Safety planning intervention: A brief intervention to mitigate suicide risk. Cognitive and Behavioral Practice, 19(2), 256-264.

Stanley B, Chaudhury SR, Chesin M, et al. An emergency department intervention and follow-up to reduce suicide risk in the VA: Acceptability and effectiveness. Psychiatr Serv. Jun 1 2016;67(6):680-683.

JC. (2018). Sentinel Event Alert 56: Detecting and treating suicide ideation in all settings. Retrieved November 1, 2021 from https://www.jointcommission.org/assets/1/18/SEA_56_Suicide.pdf

Stanley B, Brown GK, Currier GW, Lyons C, Chesin M, Knox KL. Brief Intervention and Follow-Up for Suicidal Patients With Repeat Emergency Department Visits Enhances Treatment Engagement. Am J

Public Health. 2015 Aug;105(8):1570-2. doi: 10.2105/AJPH.2015.302656. Epub 2015 Jun 11. PMID: 26066951; PMCID: PMC4504270.

NAASP. (2018). Recommended standard care for people with suicide risk: Making health care suicide safe. Retrieved November 1, 2021 from https://theactionalliance.org/sites/default/files/action_alliance_recommended_standard_care_final.pdf

Brodsky BS, Spruch-Feiner A, Stanley B. The zero suicide model: Applying evidence-based suicide prevention practices to clinical care. Front Psychiatry. 2018;9:33. doi:10.3389/fpsyt.2018.00033

Chesin MS, Stanley B, Haigh EA, et al. Staff views of an emergency department intervention using safety planning and structured follow-up with suicidal veterans. Arch Suicide Res. 2017;21(1):127-137.

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through stratification of results

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

12

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

12

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

mean signal-to-noise reliability (beta binomial method)

Signal-to-Noise: Sample size

189

Signal-to-Noise: Statistical result

Rate 1: 0.85; Rate 2: 0.82

Signal-to-Noise: Interpretation of results

Signal-to-noise reliability of 0.70 or greater is adequate for distinguishing the relative performance of one entity from another. The reliability of the measure at both the provider and clinic/practice levels is above this threshold value for both Rate 1 & Rate 2.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Correlation - Spearman's Rho

Empiric Validity: Sample size

189

Empiric Validity: Statistical result

.22

Empiric Validity: Methods and findings

We tested for correlation between each process measure performance rate with performance on NQF #0104: Adult Major Depressive Disorder (MDD): Suicide Risk Assessment at the provider and site levels. Rate 1 demonstrated moderate correlation with NQF 0104 at the provider and site level, while rate 2 showed weak and moderate correlation at the provider and site level, respectively. A conceptually unrelated outcome measure for which we possessed data elements was not available.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

11

Face Validity: Result

10

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: The current measure employs diagnostic (ICD-10-CM) and procedural (CPT) coding and structured data elements commonly captured as part of outpatient encounters in electronic health record and claims data with established reliability and validity. These data elements are routinely used for quality measure testing in the same population and care setting as evaluated in the current measure, i.e., adults with mental and/or substance use disorders seeking outpatient care. The CSSRS as an instrument as established reliability and validity as well.

To determine if suicide safety plans were completed with meaningful information, we conducted a manual review of 547 randomly selected patients (10 per site if available) from one data Source: . The first two items of the suicide safety plan were reviewed.

Additional data element validity testing was performed. See Summary Report for details.

Sample Size

547

Statistic Name

Other (enter here):: Percent Absent

Statistical Results

.01

Interpretation of results

Suicide safety plan data element is valid.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

Rate 1: 20.9; Rate 2: 2.3

Median performance score

Rate 1: 17.4; Rate 2: 0.0

Minimum performance score

0

Maximum performance score

Rate 1: 91.5; Rate 2: 34.6

Standard deviation of performance scores

Rate 1: 1.1; Rate 2: 0.4

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

American Psychiatric Association

Measure Steward Contact Information

Andrew Lyzenga

800 Maine Ave SW, Suite 900

DC, District of Columbia 20024

alyzenga@psych.org

202-744-9776

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

Please see the attached testing report (SuicideProcess_TSR) for full measure testing results and analyses.

MUC2022-131 Reduction in Suicidal Ideation or Behavior Symptoms

Program

Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Individuals who demonstrated a reduction in suicidal ideation and/or behavior symptoms as demonstrated by results of a follow-up assessment using the C-SSRS+ within 120 days after the index assessment during the measurement period. Reduction: Any decrease in score. Follow-up Assessment: Follow-up assessment using the CSSRS at a separate encounter from the baseline assessment. This assessment was administered within 90 days (+30 days) after the baseline assessment within the 16-month measurement period. If there are multiple assessments during the measurement period, the last assessment completed within 90 days (+30 days) after the baseline assessment was counted as the follow-up assessment. Columbia-Suicide Severity Rating Scale 'Screen Version': Suicidal ideation and behavior should be assessed using the Columbia-Suicide Severity Rating Scale 'Screen Version' or the 'Since Last Visit' version of the CSSRS. The CSSRS includes a 6-item patient self-reported tool that asked about wish for death, thoughts of suicide, suicidal thoughts with method without specific thoughts or intent, suicidal intent without and with specific plan, and suicide behavior along with the intensity of suicidal ideation subscale. The subscale is rated on a 5-point scale (1=least severe to 5=most severe). Baseline Assessment: Defined in denominator details. Measurement Period: A 16-month period, starting 4 months prior to the measurement year through the 12 months of the measurement year.

Numerator Exclusions

N/A

Denominator

Individuals aged 18 and older with a mental and/or substance us disorder with suicidal ideation and/or behavior symptoms OR deemed a suicide risk based on their clinician's evaluation using the CRPSR or similar tool and have an encounter with an index assessment completed using the CSSRS during the denominator identification period.

Denominator Details:

Age Range: Individuals aged 18 and older as of the date of the baseline encounter.

Suicidal Ideation and/or Behavior Symptoms: Any non-zero score on the CSSRS or clinician determination of increased suicide risk.

Codes Used to Identify Mental and/or Substance Use Disorder Diagnoses: 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), any Mental, Behavioral, and Neurodevelopmental disorder diagnosis, F01-99.

Codes used to identify outpatient encounters:

99201-99205, 99211-99215, 99241-99245, 90791-90792, 90785, 90832-90840, 90845, 90847, 90849, 90853, 90880, 90875, 90876, 90901, 90911-90913, 99441-99444, 90865, 96112-96113, 96116, 96118, 96120, 96121, 96125, 96127, 96130-96133, 96136-96137, 99401-99409

Baseline Assessment: The encounter when the individual first completes the CSSRS was counted as the baseline assessment. If there are multiple assessments during the measurement period, the first assessment completed during the denominator identification period was counted as the baseline is.

Denominator Identification Period: The period in which individuals can have an encounter with a baseline assessment using the CSSRS. The denominator encounter period is the 12-month window starting 4 months prior to the measurement year and ending 8 months into the measurement year.

Denominator Exclusions

Denominator exclusions included situations where the patient's functional capacity or motivation (or lack thereof) to improve may impact the accuracy of results of validated tools, such as acute medical conditions, delirium, dementia, intellectual, and development disorders. A patient met criteria for exclusion if there were documentation of an exclusion diagnosis at any point during the denominator intake period or the patient's date of death occurred within the measurement year. ICD-10-CM codes used to identify denominator exclusions: F00-09: Mental disorders due to known physiological conditions; F70-79: Intellectual disabilities; F80-89: Developmental Disorders Patient death: Value of TRUE in the vital status indicator = patient death during the measurement year.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Adults 18 and over, all payer

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Psychiatry

Measure Type

Outcome - (PRO-PM)

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

The measure is calculated using administrative codes (ICD-10-CM, CPT) and patient assessment data derived from the EHR.

Providers collect information from patients via survey instrument; the CSSRS instrument may be included in EHRs or patient portals. Paper surveys can be uploaded to the EHR as PDFs. The registry then extracts data from EHRs and calculates the measure.

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Behavioral Health

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

This measurement can be linked to a number of Improvement Activities (IA) including those covering screening and follow-up for certain conditions - e.g., MDD prevention and treatment intervention (IA-BMH-5), Depression Screening (IA-BMH-4) and Unhealthy Alcohol Use (IA-BMH-9). Measurement-based care (MBC) processes encompass screening and assessment of patients and include the use of the information for patient engagement and follow-up care. Other links to IA include Collaborative Care Management Training Program (IA-BMH-10), which has as a focus the implementation of MBC in the treatment and care of those with mental health and substance use disorders. Further, Electronic Health Record Enhancements for BH data capture (IA-BMH-8) can be linked to this measure primarily because electronic data capture of patient reported assessments provides for an efficient and effective means of reducing the burden with implementation of measurement-based care. Links to cost measures include

certain impacts on Medicare Beneficiary Spending (MSPB-1) that may be expected to occur primarily post inpatient stay with outpatient follow-up care to prevent subsequent preventable hospitalizations.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

Yes

CMIT ID

06118

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

9999

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

No

Which measure(s) already in a program is your measure similar to and/or competing with?

N/A

How will this measure be distinguished from other similar and/or competing measures?

N/A

How will this measure add value to the CMS program?

N/A

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

Yes (enter here):: social and clinical risk factors

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

Measure specification and logic were examined to determine the extent to which data are readily available or could be readily captured for performance measurement. Primary challenges encountered were related to data collection in PsychPRO. The registry was set to collect suicide outcomes measures data responses; however, it was observed that either the clinicians had difficulty implementing MBC in their practice or they prescribed the assessment measure, but the patients refused to complete them. Additionally, the data collection period included the Coronavirus Disease 2019 (COVID-19) pandemic which impacted both aspects of the data collection. Finally, clinicians using PsychPRO were obligated to enter diagnosis information for patients if not additionally employing an EHR, which presented an additional barrier impacting data completeness and measure estimation. COVID-19 impacted the collection of data in one of the three external data Sources: as well. Based on feedback from the alpha testing, the clinicians reported that their patients found it slightly to moderately easy to complete the CRRSS using the portal. The clinicians also reported that the visualization of how the patient scored was slightly to moderately clinically useful to them. EHR data requires a low burden for data collection at scale. However, accuracy and missingness in data elements is always a potential limitation. Data derived from psychiatric EHR data will often only encompass the patient's mental health diagnoses. Scores of medical illness burden (e.g., Charlson Comorbidity Index) are dependent on medical diagnosis information. As such, any estimation of medical illness burden is limited and may require reliance on proxy measures (e.g., medications). Given the known impact of chronic medical illness of mental health, a limited ability to control for medical illness burden is a limitation. Additionally, many EHRs separate billing and clinical information. As a result, the diagnostic information available is limited to 'problem lists' that are dependent on providers to reconcile. Encounter-level diagnosis information from billing would permit more refined assessment of acute illness states (e.g., delirium, intoxication). The challenges we encountered were related to data collection in PsychPRO. The registry was set to collect suicide outcome measure data responses within the MBC framework; however, it was observed that either the clinicians had difficulty implementing MBC in their practice or they prescribed the assessment measures, but the patients failed to complete them. Based on feedback from some of the clinicians, the use of MBC in clinical practice requires both considerable resources and changes to clinic workflow, especially for small solo practices. During the alpha testing of the suicide PRO-PM as well as testing of all measures in this initiative, participating practices using the PsychPRO registry required time and training to be oriented to both the PsychPRO system and the use of standardized assessment tools in practice. The measure development team provided technical assistance and education in the form of Learning Collaborative webinars, newsletters, and one-on-one consultation to practices. Clinicians were also provided with a clear explanation and rationale of the recommended process of care being measured by the outcome measures. Additionally, the data collection period included the Coronavirus Disease 2019 (COVID-19) pandemic, which impacted both aspects of the data collection in PsychPRO as well as in one of the external data sources. To adjust to these challenges, we used external data sources to provide data and replace the registry data as detailed above.

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Other digital method

Evidence of Performance Gap

Suicide is a preventable cause of lost lives, yet each year over 40,000 Americans die by suicide (Hedegaard, 2018). Americans at risk for suicide present to multiple settings within a month to a year of their deaths. Safety planning, means reduction, and connecting suicidal persons to treatment are effective and critical elements in suicide prevention (Suicide Prevention ReSource: Center, 2015; Betz et al, 2016; National Alliance for Suicide Prevention, 2018), as echoed in the most updated clinical practice guidelines for assessment and treatment of suicidal persons(DoD/VA, 2019).

At the provider level, 2.7% of providers performed better than the mean provider rate, 15.3% performed the same, 4.8% performed worse, and 77.3% failed to reach the 10-patient threshold for testing.

Referemces:

Hedegaard H, et al. Suicide Mortality in the United States, 19992017. NCHS Data Brief No. 330, November 2018. https://www.cdc.gov/nchs/products/databriefs/db330.htm

Suicide Prevention Resource Center, Department of Defense Strategy for Suicide Prevention. 2015. Retrieved from: https://sprc.org/resources-programs/defense-strategy-suicide-prevention-dssp

Betz et al. Lethal means access and assessment among suicidal emergency department patients. Depress Anxiety. 2016 Jun;33(6):502-11.

National Alliance for Suicide Prevention, Recommended Standard Care for People with Suicide Risk: Making Health Care Suicide Safe. 2018. Retrieved from: https://theactionalliance.org/sites/default/files/action_alliance_recommended_standard_care_final.pdf

U.S. Department of Veterans Affairs and U.S. Department of Defense Clinical Practice Guideline for the Assessment and Management of Patients at Risk for Suicide (2019). Retrieved from: https://www.healthquality.va.gov/guidelines/MH/srb/VADoDSuicideRiskFullCPGFinal5088212019.pdf.

Unintended Consequences

No unintended consequences identified.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies)

Summarize the empirical data

Suicide safety planning (SSP), which involves counseling the suicidal individual around reducing access to lethal means, teaching brief problem-solving and coping skills, and helping the individual increase social support and identify emergency contacts (Boudreaux et al., 2017; Stanley & Brown, 2012; Stanley et al., 2015; Stanley et al., 2016), is effective and critical in suicide prevention as echoed in recent clinical

practice guidelines (Department of Veterans Affairs Department of Defense, 2013; 2019) and recommendations from the Joint Commission (The Joint Commission, 2016) and the National Action Alliance for Suicide Prevention ([Action Alliance]; National Action Alliance for Suicide Prevention, 2018). It has been identified as the best practice for suicide prevention by the American Foundation for Suicide Prevention ([AFSP]; ASFP, 2018) and the Suicide Prevention Resource Center ([SPRC]; Suicide Prevention ReSources: Center, 2009). In fact, this effective suicide prevention initiative has been found to be clinically useful and feasible by both suicidal individuals and clinicians, associated with reduction in suicidal behaviors. Individuals with suicidal ideation and behaviors also report that the SSP helps them maintain their safety and increases the likelihood of them remaining in care (Brodsky et al., 2018; Chesin et al., 2017; Stanley et al., 2016; Stanley et al., 2018). Several studies have provided compelling support for suicide safety planning interventions, suggesting that such interventions are associated with reductions in suicidal behavior and increased treatment engagement. These include: A 2017 randomized control trial (RCT) by Bryan and colleagues evaluating the effectiveness of crisis response planning for the prevention of suicide attempts among active-duty Army Soldiers (N=97) presenting for an emergency behavioral health appointment. Participants were randomly assigned to receive a contract for safety, a standard crisis response plan, or an enhanced crisis response plan. Crisis response planning was associated with a reduction in suicide attempts, more rapid decline in suicidal ideation, and fewer inpatient hospital days (Bryan et al, 2017). A 2018 study by Stanley, Brown, and colleagues using a cohort comparison approach to determine whether a Safety Planning Intervention (SPI) administered in EDs with follow-up contact for suicidal patients was associated with reduced suicidal behavior and improved outpatient treatment engagement in the 6 months following discharge. Patients receiving a SPI were less likely to engage in suicidal behavior than those in usual care, had approximately half the odds of suicidal behavior over 6 months, and more than double the odds of attending at least 1 outpatient mental health visit (Stanley et al, 2018). A number of organizations have developed tools and resources for suicide prevention, including:

CDC: https://www.cdc.gov/violenceprevention/pdf/suicideTechnicalPackage.pdf

National Action Alliance for Suicide Prevention:
https://theactionalliance.org/sites/default/files/clinicalcareinterventionreport.pdf

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

Yes

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Addressed through stratification of results

Cost estimate completed

No

Cost estimate methods and results

N/A

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

12

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

12

Meaningful to Patients: Numbers consulted

12

Meaningful to Patients: Number indicating survey/tool is meaningful

12

Meaningful to Clinicians: Numbers consulted

11

Meaningful to Clinicians: Number indicating survey/tool is meaningful

9

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

Yes

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

11

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

11

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency

Testing methodology and results

Instrument reliability was not tested. Instead, we used an already established and standardized assessment tool, CSSRS, with established inter-rater reliability coefficient and a sensitivity and specificity above 95% as well as a high validity 35,36. In another study by Posner and colleagues (2011), the convergent and divergent validity of the CSSRS were good and the instrument was shown to have high sensitivity and specificity when compared with other behavior scales 37. Madan and colleagues (2016) reported that the CSSRS had an excellent internal consistency with ordinal Î± of 0.95 38. The Receiver Operator Characteristics of the scale showed that the CSSRS performed adequately in classifying any suicide-related behavior within 6 months of discharge from the hospital with an Area Under the Curve (AUC) of 0.757, P < .001. The sensitivity and specificity of the total and summary scores from the suicidal ideation/behavior factor were 0.694 and 0.652-0.674 respectively.

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

mean signal-to-noise reliability

Signal-to-Noise: Sample size

94

Signal-to-Noise: Statistical result

.77

Signal-to-Noise: Interpretation of results

In general, a score of 0.7 or higher suggests the measure has adequate reliability. Across all providers and sites, the mean reliability estimate, with or without exclusions, is above 0.7 for, suggesting the process measure rates have good reliability.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Correlation - Spearman's Rho

Empiric Validity: Sample size

94

Empiric Validity: Statistical result

.05

Empiric Validity: Methods and findings

We tested for correlation between measure performance rate with performance on the "Depression Remission at Six Months (NQF 0711)" outcome measure at the provider level and site level. We tested a minimum threshold of 10 patients for the denominator.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

11

Face Validity: Result

10

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Inter-rater agreement for numerator, denominator, exclusions between query only specific variables versus manual abstraction from all variables in the entire dataset. 1000 records were tested for each of the three data Sources: used.

Sample Size

3000

Statistic Name

Kappa

Statistical Results

Denominator Kappas = 1, 1, 1 (DS1, DS2, DS2). Numerator Kappa = 0.2, 0.4, 0.6 (DS1, DS2, DS2). Denominator Exclusions = 1, 1, 1 (DS1, DS2, DS2).

Interpretation of results

Denominator and numerator demonstrate perfect strength in agreement. For the numerator, the strength in agreement ranged from fair to substantial. It must be noted that a limited number of additional variables were available in the dataset extracts for manual abstraction, so estimates are likely to be over or under-estimated. However, overall, it would appear the necessary data element for testing are valid.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

26.9

Median performance score

19.2

Minimum performance score

0

Maximum performance score

100

Standard deviation of performance scores

1.5

Does the performance measure use survey or patient-reported data?

Yes

Surveys or patient-reported outcome tools

Columbia-Suicide Severity Rating Scale (CSSRS)

Section 5: Measure Contact Information

Measure Steward

American Psychiatric Association

Measure Steward Contact Information

Andrew Lyzenga

800 Maine Ave SW, Suite 900

DC, District of Columbia 20024

alyzenga@psych.org

202-744-9776

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

N/A

Secondary Submitter Contact Information

N/A

Submitter Comments

Please see the attached testing report (SuicideOutcome_TSR) for full measure testing results and analyses.

Part C & D Star Rating [Medicare]

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

Program

Part C & D Star Rating [Medicare]

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

Members 18-85 years of age with diabetes (type 1 and type 2) who received both an eGFR and a uACR during the measurement year on the same or different dates of service. At least one uACR is identified by either of the following: both a urine albumin test and a urine creatinine test with service dates four days or less apart, or a uACR.

Numerator Exclusions

N/A

Denominator

Members 18-85 years of age as of December 31 of the measurement year, with diabetes (type 1 and type 2) identified during the measurement year or the year prior to the measurement year.

Denominator Exclusions

Members in hospice, palliative care or with evidence of frailty and advanced illness, end stage renal disease (ESRD), dialysis, polycystic ovarian syndrome, gestational diabetes or steroid-induced diabetes are excluded from the eligible population. Also excluded are members enrolled in an Institutional SNP (I-SNP) or living long-term in an institution any time during the measurement year.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare Advantage Beneficiaries ages 18-85

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Primary care

Measure Type

Process

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Claims Data;Other: For measure exclusions, information is also needed to determine enrollment in an Institutional Special Needs Plan (I-SNP) or long-term residence in an institution as identified by the LTI flag in the Monthly Membership Detail Data File.

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Health Plan

In which setting was this measure tested?

Ambulatory/office-based care

Multiple Scores

No

What one healthcare domain applies to this measure?

Chronic Conditions

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

N/A

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

CBE ID (CMS consensus-based entity, or endorsement ID)

Not Applicable

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

N/A

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

N/A

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program

Range of years this measure has been used by CMS Programs

N/A

What other federal programs are currently using this measure?

N/A

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

Kidney Health Evaluation MUC21-090. MUC21-090 was submitted through the MIPS program but is not currently active in MIPS. This measure is similar to but not competing with MUC2022-043 measure.

How will this measure be distinguished from other similar and/or competing measures?

The MIPS Kidney Health Evaluation measure isn't implemented in Medicare Advantage Plans.

How will this measure add value to the CMS program?

This measure will be implemented with a different population, Medicare Advantage Plan members.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Claims;Other: Data are submitted to the National Committee for Quality Assurance (NCQA) through the HEDIS Data Submission Process.

Stratification

Yes (enter here):: Plans report three separate age stratifications for individuals 18-64, 65-74, 75-85, as well as a total rate.

Feasibility of Data Elements

ALL data elements are in defined fields in electronic Sources:

Feasibility Assessment

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). Both the denominator and the numerator claims are based on Medicare claims value data sets.

Members excluded from the measure are identified through health and pharmacy claims data as well as through enrollment in an Institutional Special Needs Plan (I-SNP) or living long-term in an institution as identified by the LTI flag in the Monthly Membership Detail Data File.

Method of Measure Calculation

Claims;Other (enter here):: For measure exclusions, information is also needed to determine enrollment in an Institutional Special Needs Plan (I-SNP) or long-term residence in an institution as identified by the LTI flag in the Monthly Membership Detail Data File.

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Among Medicare plans, an average of 40.0 percent of members received a kidney health evaluation in the measurement year. Average plan performance was lower among the 18-64 age group (37.8 percent) than the 65-74 age group (21.5 percent) and the 75-85 age group (40.7 percent). Average performance across Medicare plans is normally distributed about the median, and there is large variation between plans performing at the 10th and 90th percentile (36.3 percentage point difference).

Medicare performance varied across and within geographic regions, with an average performance rate of 34.7 percent among plans in the Chicago region (the lowest performing regions) and 51.4 percent among plans in the San Francisco region (the highest performing region). The widest variation between the 10th and 90th percentiles was seen in the New York region (37.5 percentage point difference) and the lowest variation was seen in the Chicago region (34.7 percentage point difference).

Despite clinical practice recommendations, the National Kidney Foundation (NKF) states that fewer than 50% of adults with diabetes receive annual kidney health evaluation (NKF. 2016. Kidney Health Evaluation Measure. National Kidney Foundation. August 15, 2016).

Unintended Consequences

None identified

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

1

Outline the clinical guidelines supporting this measure

American Diabetes Association (ADA), Chronic kidney disease and risk management, Standards of Medical Care in Diabetes 2022. Diabetes Care 2022;45(Suppl. 1):S175 to S184

- Evidence-based
- Statement: At least annually, urinary albumin (e.g., spot urinary albumin-to-creatinine ratio) and estimated glomerular filtration rate 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. B
- Summary: The measure directly aligns with this guideline statement to assess annual uACR and eGFR in patients with type 1 and type 2 diabetes. The population approach of this guideline aligns with the measure accountable entity of health plans.
- Body of Evidence: The ADA Referemces: several randomized clinical trials of varying size to support the recommendations around kidney surveillance:
- Stockholm Creatinine Measurements project - 1,118,507 adults with creatinine tests

- Dapagliflozin Effect on Cardiovascular Events-Thrombolysis in Myocardial Infarction 58 (DECLARE-TIMI 58) - 17,160 adults with Type 2 diabetes
- Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) study - 1,441 patients with T1DM
- Evidence Grading System: ADA evidence-grading system for Standards of Medical Care in Diabetes. For this statement, the level of evidence (B) indicates supportive evidence from well-conducted cohort studies and/or supportive evidence from well-conducted case-control studies.

Additional Supporting Recommendations:

National Kidney Foundation (NKF), KDOQI Clinical Practice Guideline for Diabetes and CKD, 2007 (2012 update), Am J Kidney Dis. 2012;60(5):850-886.

- Evidence-based
- Statements: Patients with diabetes should be screened annually for DKD. Initial screening should commence: 5 years after the diagnosis of type 1 diabetes (A); or From diagnosis of type 2 diabetes (B). Screening should include: Measurements of urinary albumin-creatinine ratio (ACR) in a spot urine sample (B); Measurement of serum creatinine and estimation of GFR (B).
- Summary: The measure directly aligns with this guideline statement to assess annual uACR and eGFR in patients with type 1 and type 2 diabetes. The population approach of this guideline aligns with the measure accountable entity of health plans.
- Evidence Grading System: Grading of Recommendation Assessment, Development, and Evaluation (GRADE) approach. For these statements, the quality of evidence is High (A) and Moderate (B).
- Endocrine Society (ES), Clinical Practice Guideline, Treatment of Diabetes in Older Adults, 2019, J Clin Endocrinol Metab, May 2019, 104(5):1520 to 1574
- Evidence-based
- Statement: In patients aged 65 years and older with diabetes who are not on dialysis, we recommend annual screening for chronic kidney disease with an estimated glomerular filtration rate and urine albumin-to-creatinine ratio. (1|++++)
- Summary: The measure directly aligns with this guideline statement to assess annual uACR and eGFR in patients with type 1 and type 2 diabetes. The population approach of this guideline aligns with the measure accountable entity of health plans.
- Evidence Grading System: Grading of Recommendation Assessment, Development, and Evaluation (GRADE) framework. For this statement, strength of recommendation (1) is strong, and quality of evidence (++++) is high quality.

Name the guideline developer/entity

American Diabetes Association (ADA)

Publication year

2022

Full citation +/- URL

American Diabetes Association (ADA), Chronic kidney disease and risk management, Standards of Medical Care in Diabetes 2022. Diabetes Care 2022;45(Suppl. 1):S175 to S184

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

Yes

List the guideline statement that most closely aligns with the measure concept.

At least annually, urinary albumin (e.g., spot urinary albumin-to-creatinine ratio) and estimated glomerular filtration rate 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.

What evidence grading system did the guideline use to describe strength of recommendation?

Other (enter here):: ADA evidence-grading system for "Standards of Medical Care in Diabetes"

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

Level of evidence:

A: Clear evidence from well-conducted, generalizable randomized controlled trials that are adequately powered, including:

- Evidence from a well-conducted multicenter trial
- Evidence from a meta-analysis that incorporated quality ratings in the analysis
- Compelling nonexperimental evidence, i.e., "all or none" rule developed by the Centre for Evidence-Based Medicine at the University of Oxford

Supportive evidence from well-conducted randomized controlled trials that are adequately powered, including:

- Evidence from a well-conducted trial at one or more institutions
- Evidence from a meta-analysis that incorporated quality ratings in the analysis

B: Supportive evidence from well-conducted cohort studies

- Evidence from a well-conducted prospective cohort study or registry
- Evidence from a well-conducted meta-analysis of cohort studies

Supportive evidence from a well-conducted case-control study

C: Supportive evidence from poorly controlled or uncontrolled studies

- Evidence from randomized clinical trials with one or more major or three or more minor methodological flaws that could invalidate the results
- Evidence from observational studies with high potential for bias (such as case series with comparison with historical controls)
- Evidence from case series or case reports

Conflicting evidence with the weight of evidence supporting the recommendation

D: Expert consensus or clinical experience

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

Other (enter here):: ADA Grade B, Supportive evidence from well-conducted cohort studies or from a well-conducted case-control study

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

Other (enter here):: ADA evidence-grading system for "Standards of Medical Care in Diabetes"

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

Level of evidence:

A: Clear evidence from well-conducted, generalizable randomized controlled trials that are adequately powered, including:

- Evidence from a well-conducted multicenter trial
- Evidence from a meta-analysis that incorporated quality ratings in the analysis
- Compelling nonexperimental evidence, i.e., "all or none" rule developed by the Centre for Evidence-Based Medicine at the University of Oxford

Supportive evidence from well-conducted randomized controlled trials that are adequately powered, including:

- Evidence from a well-conducted trial at one or more institutions
- Evidence from a meta-analysis that incorporated quality ratings in the analysis

B: Supportive evidence from well-conducted cohort studies

- Evidence from a well-conducted prospective cohort study or registry
- Evidence from a well-conducted meta-analysis of cohort studies

Supportive evidence from a well-conducted case-control study

C: Supportive evidence from poorly controlled or uncontrolled studies

- Evidence from randomized clinical trials with one or more major or three or more minor methodological flaws that could invalidate the results
- Evidence from observational studies with high potential for bias (such as case series with comparison with historical controls)
- Evidence from case series or case reports

Conflicting evidence with the weight of evidence supporting the recommendation

D: Expert consensus or clinical experience

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

Other (enter here):: ADA Grade B, Supportive evidence from well-conducted cohort studies or from a well-conducted case-control study

List the guideline statement that most closely aligns with the measure concept.

At least annually, urinary albumin (e.g., spot urinary albumin-to-creatinine ratio) and estimated glomerular filtration rate 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.

Number of systematic reviews that inform this measure concept

N/A

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

N/A

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

N/A

Estimated Impact of the Measure: Estimate of Annual Denominator Size

5,000,000

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

N/A

Rationale for not using risk adjustment

Other (enter here):: Not applicable

Cost estimate completed

Yes

Cost estimate methods and results

Primary detection and management of kidney disease is an important aspect of diabetes management. Undiagnosed chronic kidney disease (CKD) can increase chances of related health problems, such as early death, heart disease, stroke, kidney failure and end-stage renal disease (ESRD). If a person is aware of their CKD, they can lower their risk for related health problems and kidney failure.

In 2016, Medicare spending was approximately $79 billion for CKD patients (2018 US Renal Data System Annual Report: Epidemiology of Kidney Disease in the United States. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases). CDC simulation studies showed that uACR screening for early detection of CKD was cost-effective in patients with diabetes, at $50 thousand per quality-adjusted life-years (Hoerger, TJ, JS Wittenborn, JE Segel, NR Burrows, K Imai, P Eggers, ME Pavcov, et al. 2010. A Health Policy Model of CKD: 2. The Cost-Effectiveness of Microalbuminuria Screening. American Journal of Kidney Diseases 55 (3): 463-73).

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

No

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

N/A

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

N/A

Meaningful to Patients: Numbers consulted

N/A

Meaningful to Patients: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians: Numbers consulted

N/A

Meaningful to Clinicians: Number indicating survey/tool is meaningful

N/A

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

N/A

Type of Testing Analysis

N/A

Testing methodology and results

N/A

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Signal-to-Noise

Signal-to-Noise: Name of statistic

Beta-binomial model

Signal-to-Noise: Sample size

578

Signal-to-Noise: Statistical result

0.995

Signal-to-Noise: Interpretation of results

Reliability was estimated by using the Beta-binomial model (Adams, 2009) for this health plan measure. Beta-binomial is appropriate for estimating the reliability of pass/fail rate measures. Reliability used here is the ratio of signal-to-noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

The values for the beta-binomial statistic for the health plan level measure is greater than 0.7, indicating the measure has very good reliability. The 10-90th percentile distribution of health plan level-reliability on this measure show most health plans exceeded the minimally accepted threshold of 0.7, and the majority of plans exceeded 0.9. Strong reliability is demonstrated since the majority of variance is due to signal and not to noise.

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

N/A

Other: Sample size

N/A

Other: Statistical result

N/A

Other: Interpretation of results

N/A

Empiric Validity

Yes

Empiric Validity: Statistic name

Construct validity was tested by using a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables: the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. Please see: MUC2022-043 - Attachment 2.

Empiric Validity: Sample size

573

Empiric Validity: Statistical result

0.314

Empiric Validity: Methods and findings

NCQA tested for construct validity of this measure by exploring whether it was correlated with other similar measures of quality hypothesized which are listed below.

Comprehensive Diabetes Care (CDC): Hemoglobin A1c (HbA1c) Control (<8.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is <8% during the measurement year.

- Hypothesis: positively correlated
- Statistical Result: 0.315 (strongest correlation)

CDC: HbA1c Poor Control (>9.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is >9% during the measurement year.

- Hypothesis: negatively correlated
- Statistical Result: -0.296

CDC: Eye Exam (Retinal) Performed: The percentage of adults 18-75 with diabetes that had an eye screening for diabetic retinal disease during the measurement year.

- Hypothesis: positively correlated
- Statistical Result: 0.201

CDC: Blood Pressure Control (<140/90 mm Hg): The percentage of adults 18-75 with diabetes whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.

- Hypothesis: positively correlated
- Statistical Result: 0.080 (weakest correlation)

These measures were chosen for construct validity because they are similarly focused on evidenced-based monitoring and treatment for patients with diabetes (type 1 and type 2). We hypothesized that a plan that does well on these measures for diabetes would also do well on this kidney health evaluation measure for patients who have diabetes.

The correlations are mild to moderate. This is not unexpected as the KED measure is new and developed to address a known gap in care where there is a lot of room for improvement in health plan performance.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

17

Face Validity: Result

17

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Measure Score - correlation with other similar measures of quality.

Sample Size

578

Statistic Name

Pearson correlation coefficient

Statistical Results

0.315

Interpretation of results

We tested for construct validity of this measure by exploring whether it was correlated with other similar measures of quality hypothesized which are listed below.

Comprehensive Diabetes Care (CDC): Hemoglobin A1c (HbA1c) Control (<8.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is <8% during the measurement year.

- Hypothesis: positively correlated
- Statistical Result: 0.315 (strongest correlation)

CDC: HbA1c Poor Control (>9.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is >9% during the measurement year.

- Hypothesis: negatively correlated
- Statistical Result: -0.296

CDC: Eye Exam (Retinal) Performed: The percentage of adults 18-75 with diabetes that had an eye screening for diabetic retinal disease during the measurement year.

- Hypothesis: positively correlated
- Statistical Result: 0.201

CDC: Blood Pressure Control (<140/90 mm Hg): The percentage of adults 18-75 with diabetes whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.

- Hypothesis: positively correlated
- Statistical Result: 0.080 (weakest correlation)

These measures were chosen for construct validity because they are similarly focused on evidenced-based monitoring and treatment for patients with diabetes (type 1 and type 2). We hypothesized that a plan that does well on these measures for diabetes would also do well on this kidney health evaluation measure for patients who have diabetes.

The correlations are mild to moderate. This is not unexpected as the KED measure is new and developed to address a known gap in care where there is a lot of room for improvement in health plan performance.

Construct validity via Pearson correlation test estimates the strength of the linear association between two continuous variables: the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an

observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

40.0

Median performance score

40.4

Minimum performance score

0.0

Maximum performance score

84.6

Standard deviation of performance scores

14.9

Does the performance measure use survey or patient-reported data?

No

Surveys or patient-reported outcome tools

N/A

Section 5: Measure Contact Information

Measure Steward

National Committee for Quality Assurance (NCQA)

Measure Steward Contact Information

Emily Hubbard
1100 13th Street NW
Washington, DC 20005
Hubbard@ncqa.org
(202) 481-1018

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Elizabeth Goldstein
7500 Social Security Boulevard
Baltimore, MD 21244
elizabeth.goldstein@cms.hhs.gov
(410) 786.6665

Secondary Submitter Contact Information

Christine Payne
7500 Social Security Boulevard
Baltimore, MD 21244
christine.payne@cms.hhs.gov
410.786.6590

Submitter Comments

There are 3 attachments to this measure submission form. 'MUC2022-043 - Attachment 1' provides background information for this measure and information which the submission form wouldn't allow. 'MUC2022-043 - Attachment 2' provides correlation graphs for the empiric validity questions. 'MUC2022-043 - Attachment 3' provides measure stratification detail.

Cross-Program Measures

These measures were submitted to multiple federal programs.

MUC2022-125 Gains in Patient Activation Measure (PAM) Scores at 12 Months

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program; Merit-based Incentive Payment System-Quality

Section 1: Measure Information

Measure Specifications and Endorsement Status

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.

Numerator

The numerator is the summary change score for the aggregate of eligible patients in that unit (e.g., patients in a primary care provider's panel, or in a clinic), expressed as the difference between the Baseline PAM score and then a second score taken within 12 months of the baseline (but not less than 6 months). In addition to the summary change score, the reporting entity should provide 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).

Numerator Exclusions

Patients who are at PAM level 4 at baseline.

Patients who are flagged with outlier scores on the PAM.

Denominator

Patients aged 14 and older with two PAM scores no less than 6 months and not more than 12 months apart who were seen for a qualifying visit at least once during the performance period. Qualifying visits include visits with CPT codes 99201-99205; 99212-99215; 99324-99337; 99341-99350; 99381-99387; 99391-99397; 99490; 99495-99496; 98966-98968, 98969-98972, 99421-99423, 99441-99443, 99444

Individual clinicians would need to have two PAM scores on at least 50% of their eligible population and a minimum of 40 patients with two PAM scores.

Denominator Exclusions

Patients who are at PAM level 4 at baseline

Children under 14

Patients with a diagnosis of dementia or cognitive impairment. ICD-10 Codes include:

Code	Code Description
G31.09	Other frontotemporal dementia
F03	Unspecified dementia
F01	Vascular dementia
F03.90	Unspecified dementia without behavioral disturbance
F03.91	Unspecified dementia with behavioral disturbance
F01.50	Vascular dementia without behavioral disturbance
G31.83	Dementia with Lewy bodies
F01.51	Vascular dementia with behavioral disturbance
F18.97	Inhalant use, unspecified with inhalant-induced persisting dementia
F02.81	Dementia in other diseases classified elsewhere with behavioral disturbance
F02.80	Dementia in other diseases classified elsewhere without behavioral disturbance
F02	Dementia in other diseases classified elsewhere
F10.97	Alcohol use, unspecified with alcohol-induced persisting dementia
F19.97	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting dementia
F19.17	Other psychoactive substance abuse with psychoactive substance-induced persisting dementia
F13.97	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting dementia
F13.27	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting dementia
F19.27	Other psychoactive substance dependence with psychoactive substance-induced persisting dementia
F01.5	Vascular dementia
F03.9	Unspecified dementia
G31.0	Frontotemporal dementia
F02.8	Dementia in other diseases classified elsewhere
F18.17	Inhalant abuse with inhalant-induced dementia
F18.27	Inhalant dependence with inhalant-induced dementia

Code	Code Description
F10.27	Alcohol dependence with alcohol-induced persisting dementia
G10	Huntington's disease
G30.0	Alzheimer's disease with early onset
G30.1	Alzheimer's disease with late onset
G30	Alzheimer's disease
G30.9	Alzheimer's disease, unspecified
G31.01	Pick's disease
G20	Parkinson's disease
A81.00	Creutzfeldt-Jakob disease, unspecified
R41.0	Disorientation, unspecified
I67.850	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
G40.909	Epilepsy, unspecified, not intractable, without status epilepticus
A81.09	Other Creutzfeldt-Jakob disease
G31.84	Mild cognitive impairment, so stated

Denominator Exceptions

Not applicable

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

The target population for the measure includes adolescents and adults > 14 years of age.

Areas of specialty the measure is aimed to, or specialties that are most likely to report this measure

Other: PAM-PM is a disease-agnostic measure meant to provide meaningful information about changes in activation across many patient populations.

Measure Type

Outcome - (PRO-PM)

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

N/A

What data sources are used for the measure?

Electronic Health Record;Standardized Patient Assessments;Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

N/A

At what level of analysis was the measure tested?

Clinician - Individual;Clinician - Group;Facility

In which setting was this measure tested?

Ambulatory surgery center;Ambulatory/office-based care;Behavioral health clinic;Dialysis facility;Home health;Inpatient rehabilitation facility;Skilled nursing facility;Veterans Health Administration facility;Other: Outpatient rehabilitation; pharmacy

Multiple Scores

No

What one healthcare domain applies to this measure?

Person-Centered Care

MIPS Quality: Identify any links with related Cost measures and Improvement Activities

The proposed quality measure assesses gains in PAM score across a defined time period as an assessment of improvements in patient activation.

This measure can be linked to the following Improvement Activity:

IA_BE_16: Promote Self-management in usual care. The Patient Activation Measure (PAM) is designated as one of the eligible improvement activities, meaning that it is expected to improve clinical care delivery and outcomes. IA_BE_16 incorporates evidence-based, culturally and linguistically tailored techniques for promoting self-management into usual care, and providing patients with tools and resources for self-management.

Is this measure in the CMS Measures Inventory Tool (CMIT)?

No

CMIT ID

N/A

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Endorsed

CBE ID (CMS consensus-based entity, or endorsement ID)

2483

If endorsed: Is the measure being submitted exactly as endorsed by NQF?

Yes

If not exactly as endorsed, specify the locations of the differences

N/A

If not exactly as endorsed, describe the nature of the differences

N/A

If endorsed: Year of most recent CDP endorsement

N/A

Year of next anticipated NQF Consensus Development Process (CDP) endorsement review

2022

Digital Measure Information

Is this measure an electronic clinical quality measure (eCQM)?

No

If eCQM, enter Measure Authoring Tool (MAT) number

N/A

If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?

N/A

If eCQM, does any electronic health record (EHR) system tested need to be modified?

N/A

Measure Use in CMS Programs

Was this measure proposed on a previous year's Measures Under Consideration list?

No

Previous Measure Information

N/A

What is the history or background for including this measure on the new measures under consideration list?

Measure currently used in a CMS program being submitted as-is for a new or different program

Range of years this measure has been used by CMS Programs

Kidney Care Choices (2022); Maternal Opioid Misuse (2021-2022)

What other federal programs are currently using this measure?

CMMI Models listed above

Is this measure similar to and/or competing with a measure(s) already in a program?

Yes

Which measure(s) already in a program is your measure similar to and/or competing with?

CMIT ID: 00371, Improvement in Management of Oral Medications (Home Health Quality Reporting/Home Health Services Compare)

CMIT ID: 00404, IRF Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients (Inpatient Rehabilitation Facility Compare/ Skilled Nursing Facility Quality Reporting)

How will this measure be distinguished from other similar and/or competing measures?

While we are aware of two current measures that assess activation, they do so using estimates of patient's ability to self-manage their health and participate in care activities that are not nearly as well researched as the PAM-based measure we are proposing. The PAM has added appeal in that it is a disease-agnostic measure, applicable and meaningful to a wide set of patients, unlike the existing measures. Please also see attachment that shows other measures that claim to measure patient activation (Attachment I, Activation Measure Comparisons)

How will this measure add value to the CMS program?

Measures are similar to PAM in that they also estimate members' ability to self-manage their conditions and effectively participate in care activities.

If this measure is being proposed to meet a statutory requirement, please list the corresponding statute

N/A

Section 2: Measure Evidence

How is the measure expected to be reported to the program?

Clinical Quality Measure (CQM) Registry

Stratification

No

Feasibility of Data Elements

All data elements are in defined fields in electronic sources.

Feasibility Assessment

All data elements to compute a PAM score and activation level can be delivered electronically. Data can be collected at the point of care in-person. Data can also be collected via IVR, through the patient portal, or via the US mail. Most EHRs can make a place for PAM data, if one is not already specified. PAM questions and scoring have been integrated into a number of electronic medical records (e.g., Epic, eClinicalWorks), and care management software (e.g. CaseTrakker, McKesson CCR/Vitals). The technical

structure also exists to provide real time scoring through a web service API for PAM questions integrated into any software application. Users of this Web service receive a PAM score and activation level for each completed assessment. PAM operationalized: Today more than 125 organizations in 40+ states, as well as national organizations, are using PAM as an outcome measure, as well as a tool to help target reSources: and tailor support to a person's level of activation of self-management ability. For example, in New York state, PAM has been mandated for use in Medicaid reform as part of the Delivery System Reform Incentive Payment (DSRIP) Program, which seeks to achieve a 25 percent reduction in avoidable hospital use over five years for state Medicaid participants. Scoring adjustments: Over time Insignia has been able to improve the scoring of PAM with the collection of both larger amounts of data, and data richer in demographic, socioeconomic and health condition insights. For example, at the end of 2013, PAM level scoring cut points were adjusted based upon data collected over the previous three years. This adjustment had the effect of slightly increasing the range of Level 3 and raising the starting point for the highest level of activation, level 4. Missing data: PAM was constructed and is scored using Rasch measurement model analysis. This is a stochastic, not deterministic, model, and thus missing data has no influence. In every analysis missing responses to an item are calibrated so that one can see if nonresponse to an item is biasing results. The calibrated difficulty structure of missing responses is universally between "disagree" and "agree". This is what we would expect if nonresponse to an item is indicating something other than nonresponse. The average (mean across thirteen items) percent of people not responding to an item, or missing data, is 3.73%. 0 & 100 Scores: Scores at either extreme are dropped from evaluation as indicators that PAM was not taken truthfully. These two scores extremes tend to account for 2% to 4% of responses Frequency of data collection: Organizations typically strive to administer PAM at least two times over 12 months. The cadence of repeat administration depends on the population (Medicare, Medicaid, Medicare), and the design of the program (frequency of interaction, modes of interaction). Typically repeat PAM administration occurs within months three and six following the baseline administration. Even a single point change in activation (there are 10 to 12 points between activation levels) has proven significant. Time to complete: Most individuals will complete PAM in 3 to 5 minutes.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

As summarized in the NQF endorsement, patient self-management and life style behaviors are important determinant of health outcomes and influence other quality metrics. Patient activation is a predictor of these self-management behaviors. Supporting patient's ability to self-manage is critical for improving outcomes. Measuring activation is a way for clinicians to know where to start with a patient, and help them move forward. Patient activation can also be increased with targeted support. There is a growing list of peer-reviewed studies (over 700 published studies using PAM as a key variable) showing it is possible to support greater activation in patients. Intervention studies show that targeted interventions can increase activation and improve outcomes (see list below).

High quality medical care should result in improvements in patient's ability to self-manage. The PAM score (and changes in PAM scores) can indicate the degree to which this is occurring. A 3-point increase

in PAM score is associated with improvements in health-related behaviors. (Fowles et al 2009; Hibbard et al 2009)

The logic model is as follows:

Assess Patient Activation >> Coaching and Support by Clinical Team >> Increased Patient Activation >> Improved Health Behaviors Improved

Health Outcomes >> Reduction in Utilization and Costs

Unintended Consequences

No unintended consequences have been observed.

Number of clinical guidelines, including USPSTF guidelines, that address this measure topic

N/A

Outline the clinical guidelines supporting this measure

N/A

Name the guideline developer/entity

N/A

Publication year

N/A

Full citation +/- URL

N/A

Is this an evidence-based clinical guideline?

N/A

Is the guideline graded?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

What evidence grading system did the guideline use to describe strength of recommendation?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe strength of recommendation in the guideline?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated strength of recommendation?

N/A

What evidence grading system did the guideline use to describe level of evidence or level of certainty in the evidence?

N/A

List all categories and corresponding definitions for the evidence grading system used to describe level of evidence or level of certainty in the evidence?

N/A

For the guideline statement that most closely aligns with the measure concept, what is the associated level of evidence or level of certainty in the evidence?

N/A

List the guideline statement that most closely aligns with the measure concept.

N/A

Number of systematic reviews that inform this measure concept

5

Briefly summarize the peer-reviewed systematic review(s) that inform this measure concept

Newland P, Lorenz R, Oliver BJ. Patient activation in adults with chronic conditions: A systematic review. J Health Psychol. 2021 Jan;26(1):103-114. doi: 10.1177/1359105320947790. Epub 2020 Aug 23. PMID: 32830587,10 studies included on patient activation, No study-specific risk of bias/quality assessment. For adults with CNCHCs [t]he literature review revealed that differing measures of self-management can be influenced using patient activation measure and HRQOL, Cuevas H, Heitkemper E, Huang YC, Jang DE, G A, Z J. A systematic review and meta-analysis of patient activation in people living with chronic conditions. Patient Educ Couns. 2021 Sep;104(9):2200-2212. doi: 10.1016/j.pec.2021.02.016. Epub 2021 Feb 10. PMID: 33610334. 32 studies included on patient activation; a meta-analysis was conducted on 7 RCTs. The quality of included studies was assessed with the Critical Appraisals Skills Programme (CASP), which includes eight unique appraisal tools to address the most common research study designs. For this study, all randomized controlled trials (RCTs) were evaluated using the CASP RCT checklist, and other intervention studies were assessed using the CASP cohort study checklist for quasi-experimental studies. For both CASP checklists, it is advised that each item be scored as "Yes," "Cannot tell," or "No," with decisions about the final scoring schema left up to the individual research team. For this study, it was decided that any study receiving a "No" on 6 or more items would be removed. Increased patient activation is associated with appropriate use of the health care system and improved self-management. Kinney RL, Lemon SC, Person SD, Pagoto SL, Saczynski JS. The association between patient activation and medication adherence, hospitalization, and emergency room utilization in patients with chronic illnesses: a systematic review. Patient Educ Couns. 2015 May;98(5):545-52. doi: 10.1016/j.pec.2015.02.005. Epub 2015 Feb 19. PMID: 25744281. 10 studies included. An assessment of methodological quality of the individual studies was conducted using a modified version of the Downs and Black criteria... For each study reviewed, a quality score was calculated by dividing the number of points received by the 18 eligible points. Higher quality was designated by a higher score. Downs and Black does not specify a cut-off threshold indicative of quality studies, however, the mid-point score of 9 has been used to distinguish between those studies of adequate vs. inadequate quality [32,33]. For this review, studies which fell below 9 points (50%) of the total score were deemed of inadequate quality and were excluded. Patients who scored in the lower PAM stages (Stages 1 and 2) were more likely to have been hospitalized. Patients who scored in the lowest stage were also more likely to utilize the

emergency room. The relationship between PAM stage and medication adherence was inconclusive in this review. Almutairi N, Hosseinzadeh H, Gopaldasani V. The effectiveness of patient activation intervention on type 2 diabetes mellitus glycemic control and self-management behaviors: A systematic review of RCTs. Prim Care Diabetes. 2020 Feb;14(1):12-20. doi: 10.1016/j.pcd.2019.08.009. Epub 2019 Sep 20. PMID: 31543458. 10 RCTs included. Only included RCTs with a sample size >120 and follow up period of >12 months; however, assessment of bias and quality was not reported. Seven [activation] interventions demonstrated a significant reduction in HbA1c, ranged from 0.36 to 0.80%. All interventions presented an improvement in at least one self-management behavior. Lin, Mei-Yu; Weng, Wei-Shih; Apriliyasari, Renny Wulan; Van Truong, Pham; Tsai, Pei-Shan, Effects of Patient Activation Intervention on Chronic Diseases: A Meta-Analysis, Journal of Nursing Research: October 2020 - Volume 28 - Issue 5 - p e116 doi: 10.1097/jnr.0000000000000387. 26 RCTs included. The two reviewers independently assessed the methodological quality of the included randomized controlled trials using the Cochrane Handbook for assessing the risk of bias (Higgins et al., 2011). We evaluated random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting. Disagreements were resolved through discussion and by consultation with the third reviewer. Patient activation interventions produced significant effects on outcomes related to physiological, psychological, behavioral, and health-related quality of life in the context of chronic diseases. The following effect sizes were obtained: (a) physiological, namely, glycated hemoglobin = -0.31 ($p < .01$), systolic blood pressure = -0.20 ($p < .01$), diastolic blood pressure = -0.80 ($p = .02$), body weight = -0.12 ($p = .03$), and low-density lipoprotein = -0.21 ($p = .01$); (b) psychological, namely, depression = -0.16 ($p < .01$) and anxiety = -0.25 ($p = .01$); (c) behavioral, namely, patient activation = 0.33 ($p < .01$) and self-efficacy = 0.57 ($p < .01$); and (d) health-related quality of life = 0.25 ($p = .01$).

Source of empirical data

Published, peer-reviewed original research

Summarize the empirical data

Studies show that targeted interventions can increase activation and improve a wide range of health outcomes. At least 20 randomized clinical trials have tested interventions that seek to increase activation as measured by PAM (see list below), and at least 29 studies have used a quasi-experimental design. Most, although not all, of the interventions tested increase activation as measured by PAM. Activation interventions have been associated with increases in PAM across different population groups including Medicaid and Medicare populations, and a wide range of conditions, including schizophrenia, diabetes, asthma, COPD, depression, arthritis and others.

A 3-point increase in PAM score is associated with improvements in health-related behaviors. (Fowles et al 2009; Hibbard et al 2009). Improvements in PAM

are also linked with better clinical outcomes and lower health care costs. Research has found that that the clinicians of patients with improved PAM scores tend to use a set of strategies that support patient behavior change (Greene, Hibbard, Alvarez et al 2016).

For a more complete list of Referemces:, please refer to Attachments A, D, and the following website:

https://s3.amazonaws.com/insigniahealth.com-assets/Research-Studies-Using-PAM.Bibliography.pdf

Name evidence type

N/A

Summarize the evidence

N/A

Does the evidence discuss a link between at least one process, structure, or intervention with the outcome?

Yes

Estimated Impact of the Measure: Estimate of Annual Denominator Size

0000

Type of Evidence to Support the Measure

Peer-Reviewed Systematic Review;Empirical data

Is the measure risk adjusted?

No

Risk adjustment variables

N/A

Patient-level demographics: please select all that apply:

N/A

Patient-level health status & clinical conditions: please select all that apply:

N/A

Patient functional status: please select all that apply:

N/A

Patient-level social risk factors: please select all that apply:

N/A

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

Rasch Analysis was used to develop the Patient Activation Measure. The analysis linking PAM with outcomes is based on multivariate (logistic and OLS regression) models that control for demographics and illness severity. These models are used to show the validity of the measure. The multivariate models are not necessary for using the PAM for a performance measure. Some of the research examines the link between PAM and outcomes for specific sub-populations, including disadvantaged populations.

For reference, see:

Hibbard JH and Cunningham P. "How Engaged Are Consumers in Their Health and Health Care, and Why Does it Matter?" Center for Studying Health Systems Change Research Brief October 2008. https://pubmed.ncbi.nlm.nih.gov/18946947/

Hibbard JH, Greene J, Overton V. "Patients With Lower Activation Associated With Higher Costs; Delivery Systems Should Know Their Patients' Scores." Health Affairs Feb. 2013. http://www.ncbi.nlm.nih.gov/pubmed/23381513

Hibbard JH, Greene J. "What the Evidence Shows about Patient Activation: Better Health Outcomes and Care Experiences; Fewer Data on Costs." Health Affairs Feb. 2013.

Research shows that a 3-point positive change in PAM is predictive of improvements in multiple health related behaviors. Improvements in PAM are also linked with better clinical outcomes and lower health care costs.

Fowles J, Terry P, Xi M, Hibbard JH, Bloom CT, Harvey L. "Measuring self-management of patients' and employees' health: Further validation of the Patient Activation Measure (PAM) based on its relation to employee characteristics." Patient Education and Counseling Vol. 77 No.2:116-122. 2009. http://www.ncbi.nlm.nih.gov/pubmed/19356881

Hibbard, JH, Mahoney E, Stock R, Tusler M. "Do Increases in Patient Activation Result in Improved Self-management Behaviors?" Health Services Research 2007; 42(4). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1955271 https://www.ncbi.nlm.nih.gov/pubmed/23381511

Rationale for not using risk adjustment

Other (enter here):: PAM is a risk measure of sorts in that it measures a person's ability to self-manage their health and care. Lower activation is predicative of poor self-management, higher healthcare utilization, and higher costs. The PAM is validated on construct validity. No risk adjustment is used. When gains in activation are used as a performance metric, then it is clear that any individual or groups of individuals can gain in activation over time. Clinicians whose patients measure low in activation are not at a disadvantage, as the outcome is measured in gains from where they started.

Cost estimate completed

Yes

Cost estimate methods and results

Research through Stanford in evaluation of CMMI grant support provides perspective as to the relationship between activation and cost for the most complex patients. Increasing activation by one level within a year lead to an 8% decline in follow up Medicare claim allowed costs. The inverse was shown as well, with a decline in PAM level associating to an 8% increase in cost. The published paper can be found here: https://pubmed.ncbi.nlm.nih.gov/30291604/. An overview paper can be found here: https://scopeblog.stanford.edu/2018/10/29/the-relationship-between-patient-self-management-and-health-care-costs/.

Health Affairs published research with a more general patient population has also documented the relationship between activation and cost. In this health system study of 33,000 patients, patient

activation was a significant predictor of cost even after adjustment for a commonly used "risk score" specifically designed to predict future costs. Patient costs were 21% higher for PAM level one patients as compared to level four patients in the evaluated follow up period.
https://pubmed.ncbi.nlm.nih.gov/23381513/

In follow research with these patients the shift between PAM levels were further evaluated. Patients who moved from 3 or 4 to 1 or 2 had projected costs that were 27 percent higher than those of the lowest-cost group (Level 4), and those who remained in 1 or 2 had costs that were 31 percent higher than those of the lowest-cost group (Level 4). https://pubmed.ncbi.nlm.nih.gov/25732493/

Section 3: Patient and Provider Perspective

Meaningful to Patients. Was input on the final performance measure collected from patient and/or caregiver?

Yes

Total number of patients and/or caregivers who responded to the question asking them whether the final performance measure helps inform care and decision making

48

Total number of patients/caregivers who agreed that the final performance measure helps inform care and decision making

45

Meaningful to Patients: Numbers consulted

39

Meaningful to Patients: Number indicating survey/tool is meaningful

39

Meaningful to Clinicians: Numbers consulted

36

Meaningful to Clinicians: Number indicating survey/tool is meaningful

36

Meaningful to Clinicians. Were clinicians and/or providers consulted on the final performance measure?

No

Total number of clinicians/providers who responded when asked if the final performance measure was actionable to improve quality of care.

N/A

Total number of clinicians/providers who agreed that the final performance measure was actionable to improve quality of care

N/A

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Construct Validity;Other (enter here):: Various

Testing methodology and results

The PAM survey measures an individual's knowledge, skills and confidence for managing their health and health care. The measure is not disease specific; it has been successfully used with a wide variety of chronic conditions, as well as with people with no medical diagnoses. As demonstrated by over 750 peer-reviewed studies, the PAM has been shown to be reliable, valid, and the de facto gold standard for measuring patient activation. The PAM is predictive of many health outcomes, including such diverse outcomes as how a patient fares after orthopedic surgery; remission of depression over time; the likelihood of hospital re-admission or ambulatory care sensitive (ACS) utilization; the trajectory of a chronic disease over time; and even the likelihood of a new chronic disease diagnosis in the coming year. A recent study indicated that PAM score changes can be used as a proxy for changes in health care costs. The study showed an inverse relationship between PAM scores and overall costs: as PAM scores increase, overall costs decrease. PAM has also been successfully used, in its entirety, as a performance metric (PAM-PM), endorsed by the National Quality Forum (see Attachment D). Health care organizations use PAM-PM to evaluate health care systems and health teams on how well they support gains in patient self-management. The change in score from baseline measurement to follow-up measurement, or the change in activation score over time, is the performance score.

Burden for Provider: Was a provider workflow analysis conducted?

No

If yes, how many sites were evaluated in the provider workflow analysis?

N/A

Did the provider workflow have to be modified to accommodate the new measure?

N/A

Section 4: Measure Testing Details

Reliability

Yes

Reliability: Type of Reliability Testing

Other (enter here): Cronbach's alpha

Signal-to-Noise: Name of statistic

N/A

Signal-to-Noise: Sample size

N/A

Signal-to-Noise: Statistical result

N/A

Signal-to-Noise: Interpretation of results

N/A

Random Split-Half Correlation: Name of statistic

N/A

Random Split-Half Correlation: Sample size

N/A

Random Split-Half Correlation: Statistical result

N/A

Random Split-Half Correlation: Interpretation of results

N/A

Other: Name of statistic

Cronbach's alpha

Other: Sample size

Per the NQF endorsement application, numerous studies were summarized to test the PROM reliability. Please see Attachment A – CMMI Memo – PAM Overview (April 2022) and Attachment B – PAM Reliability and Validity Summary for more details.

Other: Statistical result

Per the NQF endorsement application, numerous studies were summarized to test the PROM reliability. Please see Attachment A, "CMMI Memo PAM Overview (April 2022)," and Attachment B, "PAM Reliability and Validity Summary" for more details. Per the NQF endorsement application, Cronbach's alpha internal consistency reliability coefficients were computed for the 13 items of the PAM13 across a wide range of subsamples. Standard inter-item reliability in the form of Cronbach's alpha is the method used. This approach to reliability testing evaluates this core question: Are the PAM items (questions) all measuring the same construct and do 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. Appendix K summarizes these data for the ESRD population, including the availability of facility level data.

Other: Interpretation of results

The PAM 13 has very good internal consistency reliability, as suggested by many statistical Referemces:, including: Rosenthal, R., & Rosnow, R.L. (1991). Essentials of behavioral research: Methods and data analysis, 2nd edition. Boston: McGraw-Hill. Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 48(6), 1273-1296.

Empiric Validity

Yes

Empiric Validity: Statistic name

Rasch model fit

Empiric Validity: Sample size

Numerous, please see Attachments A, B, and D (MIF for NQF-endorsed PAM-PM)

Empiric Validity: Statistical result

Per the NQF endorsement application, PAM was constructed with, and is scored using the Rasch measurement model. The model is a mathematical statement of measurement, as it is known in the physical and natural sciences. The key question is: Do the data fit the model? When the data fit the model the result is a measure having the same properties as weight scale, thermometer, speedometer, etc. The measure is thus an equal interval yardstick with the "inch marks" corresponding to the item-response category combinations.

The first test of validity of measurement is item fit. Do all items fall on the single real number line representing the activation scale? All 13 items have very good fit. This has been replicated hundreds of times, see also Attachment E - PAM Validity – Rasch Const Valid Infit Outfit Samples attachment. Several differential item function analysis (e.g., Do items have the same location on the yardstick?) fail to show any DIF by subsample. The principal reason for this is that calibrations with the stochastic Rasch model are sample free (i.e., the distribution of activation scores in a sample has no effect on the calibration or fit of the PAM 13 items).

Empiric Validity: Methods and findings

Per the NQF endorsement application, construct validity is tested by examining the extent to which PAM scores or levels are related to theoretically relevant outcomes, behaviors, and underpinnings of the activation construct. Activation scores are true equal interval scores on a 0-100 scale where higher represents more activation. Using CHAIF segmentation analysis as well as Rasch variable maps we have long ago identified four distinct levels of activation. Persons falling in each level have empirically identified characteristics and each level is distinguished by different outcomes and health-related behaviors. In the construct validity testing 0-100 scores are used when the criterion variable is categorical. When the criterion variable is continuous (e.g., health care cost) activation levels are used. Each test is described below. Validity results using a sample of key studies are shown in the Attachment A, B, and D. PAM score and activation level relationships are shown for: *Lifestyle behaviors: Nutrition (consuming fruits and vegetables), regular exercise *BMI *Disposition/Attitudes: Health as a priority, feeling overwhelmed, goal setting ability *Hospitalization: Allowed costs, admits, length of stay * Physician visits * Medication: Filled prescriptions, prescription cost There is very strong evidence for the construct validity of the PAM 13 measure. We have also included a new Appendix K that descries validity data in patients with ESRD and the availability of facility level data. July 6 addendum - We previously conducted a signal-to-noise reliability analysis for two clinic systems (summary provided in Attachment L, Signal-to-Noise Analysis). Data is shown at the clinic level and rolled up for the two systems, showing good results with acceptable signal strength for more than 90% of PAM data captured. We are also attaching two relevant publications (Attachment M, Greene et al 2016 Annals of Family Medicine and Attachment N, Alvarez et al 2017 BMD Health Services Research). Of note, in the Greene et al 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 5 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. (Previous studies have demonstrated the link between increased activation and positive health and non-health outcomes.) Similarly, in the Alvarez et al analysis, primary care providers with high CS-PAM scores (a measuring indicating how much the provider sees the importance of patient self-management and patient participation in care) were significantly more likely to have patients with increased PAM scores than were primary care providers with lower CS-PAM scores. In both of these analyses patient PAM score changes linked to an individual provider were tracked over time. The data used in the articles are the same as the PAM-PM proposed here (PAM score changes over time), with one exception: The patient PAM score changes associated with each provider were continuous variables, performance was then dichotomized into "top" (average pam score change 7.5 points) and "bottom" performers (average PAM score change 3 points). Since submitting PAM-PM for consideration for inclusion in MIPS, we have collected data at a single site that we are analyzing for signal-to-noise information, at the individual provider level. That analysis will be sent as soon as possible after its completion.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

18-21 experts; 9-20 patients

Face Validity: Result

78

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between other gold standard and manual reviewer

Sample Size

7144

Statistic Name

Pearson correlation coefficient

Statistical Results

0.28

Interpretation of results

The PAM has been used in the clinical setting by individual clinicians and examined as a PAM-PM (performance measure). In a large ACO where PAM was used, we found that 7,144 patients had PAM scores at two points in time and were also linked with their individual clinicians. These clinicians were

all primary care providers and included physicians, nurse practitioners and physician assistants. We examined the degree of change in PAM scores for each clinician. The clinicians who had the patients with the highest average PAM score gains showed an average increase in scores of 7.5 for their patients; clinicians whose patients had the least gains still averaged gains of 3.1 points on the 0-100 scale 5. In a follow up study, we found that the PAM-PM was significantly linked with clinician behaviors with regard to supporting the patient role and clinician beliefs about the importance of the patient role in the care process (r=.28 , p,<.05)6. For example, clinicians whose patients were making gains in PAM scores were more likely to problem solve with patients about overcoming obstacles, they were more likely to partner with patients in finding small steps changes, and more likely to show support for patient progress. The findings indicate that when clinicians are more supportive of the patient role, their patients are more likely to have greater gains in PAM score over time. These studies together provide evidence that the PAM- PM is valid at the individual clinician level. The findings also suggest that clinicians can learn the skills and behaviors that support gains in patient activation.

Measure performance – Type of Score

Continuous Variable – Mean

Measure Performance Score Interpretation

Higher score is better

Mean performance score

57.4-68.2

Median performance score

0

Minimum performance score

0

Maximum performance score

0

Standard deviation of performance scores

10.8-11.7

Does the performance measure use survey or patient-reported data?

Yes

Surveys or patient-reported outcome tools

The performance measure is based on the Patient Activation Measure (PAM) , used as originally specified. A copy of the PAM is included as an attachment to the application, as Attachment F - PAM-13 and Attachment G - PAM-10 . We recommend the 13-item PAM with populations that are economically or educationally disadvantaged. Reliabilities are somewhat lower with these populations, so the PAM-13 is better with those groups. The Patient Activation Measure (PAM) is a 10- or 13- item questionnaire that assesses an individual's knowledge, skills and confidence for managing their health and health care. It is a patient/consumer survey that can be administered by any entity (health plan, health system, hospital or clinic, researcher) across modes (paper, IVR, online, phone, online, pad/smart phone) and

over 35 languages. 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, see attachments A, B, D, and H (PAM Score Descriptives). The PAM is a proprietary measure; its survey and tools, including training and data quality and integrity supports, are available for use with a valid license. We are committed to ensuring broad and equitable access to the PAM and welcome the opportunity to collaborate with CMS to ensure the PAM can be readily adopted by MIPS participants.

Section 5: Measure Contact Information

Measure Steward

Insignia Health, LLC, a wholly owned subsidiary of Phreesia

Measure Steward Contact Information

Hilary Hatch

434 Fayetteville Street, Suite 440,

Raleigh, NC 27601

hhatch@phreesia.com

646-234-4130

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Zeeshan Butt
434 Fayetteville Street, Suite 4400
Raleigh, NC 27601
zbutt@phreesia.com
(773) 209-8183

Secondary Submitter Contact Information

Zeeshan Butt
434 Fayetteville Street, Suite 4400
Raleigh, NC 27601
zbutt@phreesia.com
(773) 209-8184

Submitter Comments

In response to the question -- What one area of specialty the measure is aimed to, or which specialty is most likely to report this measure -- we selected "Other" because the PAM-PM is a disease-agnostic measure meant to provide meaningful information about changes in activation across many patient populations. The NQF PAM-PM measure is scheduled for maintenance review in 2022, reassigned from original review dates in 2021. We have attached an e-mail from NQF summarizing this change, Attachment J - PAM-PM NQF Maintenance Review schedule - e-mail. More detailed review of Measure Performance is available in Attachment D For some of the Measure Score Level queries, we were unable to input a summary response because more than one estimate available, but we provide more info on those analyses in Attachments A, B, and D. For the Empiric Validity: Statistical Result, while unable to provide a single numeric response, per the NQF endorsement application, PAM was constructed with, and is scored using the Rasch measurement model. The model is a mathematical statement of measurement, as it is known in the physical and natural sciences. The key question is: Do the data fit the model? When the data fit the model the result is a measure having the same properties as weight scale, thermometer, speedometer, etc. The measure is thus an equal interval yardstick with the "inch marks" corresponding to the item-response category combinations. The first test of validity of measurement is item fit. Do all items fall on the single real number line representing the activation scale? All 13 items have very good fit. This has been replicated hundreds of times, see also Attachment E - PAM Validity - Rasch Const Valid Infit Outfit Samples attachment. Several differential item function analysis (e.g., Do items have the same location on the yardstick?) fail to show any DIF by subsample. The principal reason for this is that calibrations with the stochastic Rasch model are sample free (i.e., the distribution of activation scores in a sample has no effect on the calibration or fit of the PAM 13 items).

On 6-3-2022, in response to a reviewer request, we attempted to add information related to facility level findings to support review under the ESRD program, which was accidentally left out of the original submission; however, we were unable to add an additional Appendix with that information. We are attempting to include that information within the existing fields but would welcome the opportunity to upload the file in its entirety.