NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Applications Partnership (MAP) Hospital Workgroup: 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.

Table of Contents

Ambulatory Surgical Center Quality Reporting Program..... 4

MUC2022-028 ASC Facility Volume Data on Selected Surgical Procedures (formerly ASC-7)..... 5

End-Stage Renal Disease (ESRD) Quality Incentive Program.....19

MUC2022-075 Standardized Modality Switch Ratio for Incident Dialysis Patients (SMoSR) 20

MUC2022-076 Standardized Fistula Rate for Incident Patients 36

MUC2022-079 Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities 52

Hospital Inpatient Quality Reporting Program.....67

MUC2022-018 Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Hospital Level – Inpatient) 68

MUC2022-032 Geriatrics Surgical Measure.....84

MUC2022-055 Hybrid Hospital-Wide All-Cause Risk Standardized Readmission Measure 100

MUC2022-057 Hybrid Hospital-Wide All-Cause Risk Standardized Mortality Measure.....115

MUC2022-058 Hospital Disparity Index (HDI).....132

MUC2022-112 Geriatrics Hospital Measure 146

Hospital Outpatient Quality Reporting Program 162

MUC2022-020 Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Hospital Level – Outpatient) 163

MUC2022-030 Hospital Outpatient Department Volume Data on Selected Outpatient Surgical Procedures (formerly OP-26) 179

Hospital Value-Based Purchasing Program.....191

MUC2022-082 Severe Sepsis and Septic Shock: Management Bundle.....192

Inpatient Psychiatric Facility Quality Reporting Program.....210

MUC2022-078 Psychiatric Inpatient Experience Measurement.....211

Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program 225

MUC2022-120 Documentation of Goals of Care Discussions Among Cancer Patients 226

Rural Emergency Hospital Quality Reporting Program (REHQRP).....241

MUC2022-039 Median Time from emergency department (ED) Arrival to ED Departure for Discharged ED Patients.....242

MUC2022-066 Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy 255

MUC2022-067 Risk-standardized hospital visits within 7 days after hospital outpatient surgery ... 268

MUC2022-081 Abdomen Computed Tomography (CT) Use of Contrast Material.....281

Cross-Program Measures 293

MUC2022-024 Hospital Harm - Acute Kidney Injury 294

MUC2022-026 Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA PRO-PM) in the HOPD or ASC Setting..... 316

MUC2022-027 Facility Commitment to Health Equity..... 335

MUC2022-050 Screen Positive Rate for Social Drivers of Health..... 350

MUC2022-053 Screening for Social Drivers of Health..... 369

MUC2022-064 Hospital Harm - Pressure Injury..... 387

MUC2022-084 COVID-19 Vaccination Coverage among Healthcare Personnel (HCP) (2022 revision)..... 407

MUC2022-125 Gains in Patient Activation Measure (PAM) Scores at 12 Months..... 426

Ambulatory Surgical Center Quality Reporting Program

MUC2022-028 ASC Facility Volume Data on Selected Surgical Procedures (formerly ASC-7)

Program

Ambulatory Surgical Center Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Structural measure of facility capacity collects surgical procedure volume data on selected categories of procedures frequently performed in the ASC setting Categories include: Eye, Gastrointestinal, Genitourinary, Musculoskeletal, Nervous, Respiratory, Skin, and Other

Numerator

All-patient, all payer surgical procedure volume data for six categories of procedures frequently performed in the ASC setting (gastrointestinal, eye, nervous system, musculoskeletal, skin, and genitourinary) within a one-year performance period.

Numerator Exclusions

none

Denominator

N/A

Denominator Exclusions

N/A

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

Other: No Specialty

Measure Type

Structure

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

Facilities will report the count of performed surgical procedures per category. Data entry will be achieved through the secure side of QualityNet.cms.gov via an online tool available to authorized users.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Ambulatory surgery center

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

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

N/A

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

Yes

CMIT ID

01587-C-ASCQR

Alternate Measure ID

ASC7

What is the endorsement status of the measure?

Never Submitted

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

99999

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, but the measure is undergoing substantial change

Range of years this measure has been used by CMS Programs

This measure was first adopted in the CY2012 OPPS/ASC final rule. It was finalized for removal in CY2018 OPPS/ASC final rule. (2012-2019)

What other federal programs are currently using this measure?

This measure was first adopted in the CY2012 OPPS/ASC final rule. It was finalized for removal in CY2018 OPPS/ASC final rule.

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 aligned with OP-26, which was removed and is being added to the MUC list this year.

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

This measure is different because it occurs in a different setting

How will this measure add value to the CMS program?

This measure adds value because it is setting specific to Ambulatory Surgical Centers.

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

Yes (enter here):: The counts are stratified by surgical codes

Feasibility of Data Elements

ALL data elements are in defined fields in electronic sources

Feasibility Assessment

Because this is an aggregate account of procedures performed, facilities should be able to readily obtain and submit these counts through QualityNet.

Method of Measure Calculation

Hybrid;Other (enter here):: This measure strictly counts the volume of procedures performed.

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

There is evidence that the volume of specific procedures and the overall volume of procedures performed at ASCs is increasing. For example, according to one study, the volume of orthopedic procedures (1) performed at ASCs on Medicare beneficiaries increased by 8.6% between 2012 and 2017. Studies have also found that the volume of otolaryngologic procedures performed on Medicare beneficiaries at ASCs increased by 1.8% between 2010 and 2017 and decreased by 6% at HOPDs. (2) Another study in a smaller set of ASCs in one geographic region in the United States found that between 2016 and 2019 average orthopedic surgical volume increased by about 38% and the average ASC volume

overall increased by 19% (3). Finally, another study found that ASC volume for anterior cervical discectomy and fusion increased by 185% between 2015 and 2017 (4).

In addition, according to a MedPAC analysis, between 2012 and 2017 the number of ASCs increased, and the volume of services per Medicare FFS beneficiary increased on average by 1.2 percent per year and by and 1.7 percent in 2017 (5).

Finally, CMS has now allowed some previously high-volume inpatient procedures, such as hip and knee replacement, to be performed in the ASC setting. Elective primary THA and TKA procedures were removed from the inpatient-only (IPO) list and added to the ASC Covered Procedures List (CPL): TKA was removed from the IPO for CY 2018 and added to the ASC CPL in CY2020; THA was removed from the IPO for CY 2020 and added to the ASC CPL in CY2021 [84 FR 61142; 85 FR 85866].

Lopez CD, Boddapati V, Schweppe EA, Levine WN, Lehman RA, Lenke LG. Recent Trends in Medicare Utilization and Reimbursement for Orthopaedic Procedures Performed at Ambulatory Surgery Centers. J Bone Joint Surg Am. 2021 Aug 4;103(15):1383-1391. doi: 10.2106/JBJS.20.01105. PMID: 33780398.

Kondamuri NS, Miller AL, Rathi VK, Miller L, Bergmark RW, Patel TS, Gray ST. Trends in Ambulatory Surgery Center Utilization for Otolaryngologic Procedures among Medicare Beneficiaries, 2010-2017. Otolaryngol Head Neck Surg. 2020 Jun;162(6):873-880. doi: 10.1177/0194599820914298. Epub 2020 Apr 14. PMID: 32283985.

Shukla D, Patel S, Clack L, Smith TB, Shuler MS. Retrospective analysis of trends in surgery volumes between 2016 and 2019 and impact of the insurance deductible: Cross-sectional study. Ann Med Surg (Lond). 2021 Feb 23;63:102176. doi: 10.1016/j.amsu.2021.02.022. PMID: 33732449; PMCID: PMC7937670.

Lopez CD, Boddapati V, Lombardi JM, Sardar ZM, Dyrszka MD, Lehman RA, Riew KD. Recent trends in medicare utilization and reimbursement for anterior cervical discectomy and fusion. Spine J. 2020 Nov;20(11):1737-1743. doi: 10.1016/j.spinee.2020.06.010. Epub 2020 Jun 18. PMID: 32562771.

Medicare Payment Advisory Commission. March 2019. Report to the Congress: Medicare and the health care delivery system. Chapter 5. Washington, DC: MedPAC

There is evidence that an increasing number of procedures are being performed at ASCs. For example, according to one study, the volume of orthopedic procedures (1) performed at ASCs on Medicare beneficiaries increased by 8.6% between 2012 and 2017. Studies have also found that the volume of otolaryngologic procedures performed on Medicare beneficiaries at ASCs increased by 1.8% between 2010 and 2017 and decreased by 6% at HOPDs. (2) Another study in a smaller set of ASCs in one geographic region in the United States found that between 2016 and 2019 average orthopedic surgical volume increased by about 38% and the average ASC volume overall increased by 19% (3). Finally, another study found that ASC volume for anterior cervical discectomy and fusion increased by 185% between 2015 and 2017 (4).

In addition, according to a MedPAC analysis, between 2012 and 2017 the number of ASCs increased, and the volume of services per Medicare FFS beneficiary increased on average by 1.2 percent per year and by and 1.7 percent in 2017 (5).

Finally, CMS has now allowed some previously high-volume inpatient procedures, such as hip and knee replacement, to be performed in the ASC setting. Elective primary THA and TKA procedures were removed from the inpatient-only (IPO) list and added to the ASC Covered Procedures List (CPL): TKA was removed from the IPO for CY 2018 and added to the ASC CPL in CY2020; THA was removed from the IPO for CY 2020 and added to the ASC CPL in CY2021 [84 FR 61142; 85 FR 85866].

Lopez CD, Boddapati V, Schweppe EA, Levine WN, Lehman RA, Lenke LG. Recent Trends in Medicare Utilization and Reimbursement for Orthopaedic Procedures Performed at Ambulatory Surgery Centers. J Bone Joint Surg Am. 2021 Aug 4;103(15):1383-1391. doi: 10.2106/JBJS.20.01105. PMID: 33780398.

Kondamuri NS, Miller AL, Rathi VK, Miller L, Bergmark RW, Patel TS, Gray ST. Trends in Ambulatory Surgery Center Utilization for Otolaryngologic Procedures among Medicare Beneficiaries, 2010-2017. Otolaryngol Head Neck Surg. 2020 Jun;162(6):873-880. doi: 10.1177/0194599820914298. Epub 2020 Apr 14. PMID: 32283985.

Shukla D, Patel S, Clack L, Smith TB, Shuler MS. Retrospective analysis of trends in surgery volumes between 2016 and 2019 and impact of the insurance deductible: Cross-sectional study. Ann Med Surg (Lond). 2021 Feb 23;63:102176. doi: 10.1016/j.amsu.2021.02.022. PMID: 33732449; PMCID: PMC7937670.

Lopez CD, Boddapati V, Lombardi JM, Sardar ZM, Dyrszka MD, Lehman RA, Riew KD. Recent trends in medicare utilization and reimbursement for anterior cervical discectomy and fusion. Spine J. 2020 Nov;20(11):1737-1743. doi: 10.1016/j.spinee.2020.06.010. Epub 2020 Jun 18. PMID: 32562771.

(1) Medicare Payment Advisory Commission. March 2019. Report to the Congress: Medicare and the health care delivery system. Chapter 5. Washington, DC: MedPAC

Unintended Consequences

N/A

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

00000

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

As noted above, procedures that were previously performed on an inpatient or hospital outpatient basis are now migrating to the ASC setting. This underscores the need to address the potential for poorer outcome for beneficiaries who are treated at low volume providers, for procedures that have a volume-outcome relationship. In addition, better understanding the volume of procedures from an all-payer perspective will allow CMS to target and prioritize future quality measure development

There are scores of systematic reviews that examine the volume outcome relationship for surgery. Most of the reviews that have been published support a volume-outcome relationship, but the relationship is weak for some procedures, and stronger for others. In addition, most studies have addressed procedures that are performed in the inpatient setting (although in some cases, like for knee replacement surgery those procedures are migrating to the outpatient space); studies differ in if they examined surgeon volume vs. hospital volume. A recent scoping review of the volume/outcome

relationship examined 403 studies that addressed 90 types of surgery. Study authors found that most (about 87%) of the studies had a significant volume-outcome relationship; there were 61 different types of outcomes that were examined in these studies. About half of the studies addressed cancer-related surgery (6).

Below we summarize the one systematic review that has addressed outpatient surgery.

One systematic review published in 2020 (7) examined outpatient surgery using international data, analyzed data from eight retrospective studies that addressed seven procedures: anterior cruciate ligament reconstruction, cataract surgery, meniscectomy, thyroidectomy, primary hip arthroscopy, open carpal tunnel release, and rotator cuff repair. Study authors found a volume outcome relationship for all but carpal tunnel release and thyroidectomy, however the results did not allow the study authors to recommend clear volume thresholds for these procedures.

Levaillant, M., Marcilly, R., Levaillant, L. et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. BMC Med Res Methodol 21, 204 (2021). https://doi.org/10.1186/s12874-021-01396-6

Stanak, M., Strohmaier, C. Minimum volume standards in day surgery: a systematic review. BMC Health Serv Res 20, 886 (2020). https://doi.org/10.1186/s12913-020-05724-2

As noted above, procedures that were previously performed on an inpatient or hospital outpatient basis are now migrating to the ASC setting. This underscores the need to address the potential for poorer outcome for beneficiaries who are treated at low volume providers, for procedures that have a volume-outcome relationship.

There are scores of systematic reviews that examine the volume outcome relationship for surgery. Most of the reviews that have been published support a volume-outcome relationship, but the relationship is weak for some procedures, and stronger for others. In addition, most studies have addressed procedures that are performed in the inpatient setting (although in some cases, like for knee replacement surgery those procedures are migrating to the outpatient space); studies differ in if they examined surgeon volume vs. hospital volume. A recent scoping review of the volume/outcome relationship examined 403 studies that addressed 90 types of surgery. Study authors found that most (about 87%) of the studies had a significant volume-outcome relationship; there were 61 different types of outcomes that were examined in these studies. About half of the studies addressed cancer-related surgery (6).

Below we summarize the one systematic review that has addressed outpatient surgery.

One systematic review published in 2020 (7) examined outpatient surgery using international data, analyzed data from eight retrospective studies that addressed seven procedures: anterior cruciate ligament reconstruction, cataract surgery, meniscectomy, thyroidectomy, primary hip arthroscopy, open carpal tunnel release, and rotator cuff repair. Study authors found a volume outcome relationship for all but carpal tunnel release and thyroidectomy, however the results did not allow the study authors to recommend clear volume thresholds for these procedures.

Levaillant, M., Marcilly, R., Levaillant, L. et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. BMC Med Res Methodol 21, 204 (2021). https://doi.org/10.1186/s12874-021-01396-6

(1) Stanak, M., Strohmaier, C. Minimum volume standards in day surgery: a systematic review. BMC Health Serv Res 20, 886 (2020). https://doi.org/10.1186/s12913-020-05724-2

Source of empirical data

Published, peer-reviewed original research

Summarize the empirical data

There have been many studies that have examined the relationship between surgeon/facility volume of procedures and procedural outcomes. As noted above, volume-outcome relationships are commonly found across surgeries (6) however there has been less focus on this relationship for outpatient surgery. One systematic review found a volume-outcome relationship for five of seven outpatient procedures (7).

(6) Levaillant, M., Marcilly, R., Levaillant, L. et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. BMC Med Res Methodol 21, 204 (2021). https://doi.org/10.1186/s12874-021-01396-6

(7) Stanak, M., Strohmaier, C. Minimum volume standards in day surgery: a systematic review. BMC Health Serv Res 20, 886 (2020). https:/doi.org/10.1186/s12913-020-05724-2

Name evidence type

N/A

Summarize the evidence

There have been many studies that have examined the relationship between surgeon/facility volume of procedures and procedural outcomes. As noted above, volume-outcome relationships are commonly found across surgeries (6) however there has been less focus on this relationship for outpatient surgery. One systematic review found a volume-outcome relationship for five of seven outpatient procedures (7).

(6) Levaillant, M., Marcilly, R., Levaillant, L. et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. BMC Med Res Methodol 21, 204 (2021). https://doi.org/10.1186/s12874-021-01396-6

(7) Stanak, M., Strohmaier, C. Minimum volume standards in day surgery: a systematic review. BMC Health Serv Res 20, 886 (2020). https://doi.org/10.1186/s12913-020-05724-2

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

00000

Type of Evidence to Support the Measure

Peer-Reviewed Systematic Review;Empirical data;Other (enter here):: Peer-Reviewed original research

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):: Raw count

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

No

Reliability: Type of Reliability Testing

N/A

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

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

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: Raw count

Measure Performance Score Interpretation

Other (enter here):: raw count

Mean performance score

00000

Median performance score

00000

Minimum performance score

1

Maximum performance score

00000

Standard deviation of performance scores

00000

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

CMS

Measure Steward Contact Information

Janis Grady
7500 Security Boulevard
Baltimore, MD 21244
janis.grady@cms.hhs.gov
(410) 786-7217

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

N/A

End-Stage Renal Disease (ESRD) Quality Incentive Program

MUC2022-075 Standardized Modality Switch Ratio for Incident Dialysis Patients (SMoSR)

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The standardized modality switch ratio (SMoSR) is defined to be the ratio of numbers of observed modality switches (from in-center to home dialysis- peritoneal or home hemodialysis) that occur for adult incident ESRD dialysis patients treated at a particular facility, to the number of modality switches (from in-center to home dialysis- peritoneal or home hemodialysis) that would be expected given the characteristics of the dialysis facility's patients and the national norm of dialysis facilities. The measure includes only the first durable switch that is defined as lasting 30 continues days or longer.

Numerator

Observed number of switches from in-center hemodialysis to a home dialysis modality (peritoneal dialysis or home hemodialysis) among eligible patients at the facility during the time period.

Numerator Exclusions

N/A

Denominator

Expected number of switches from in-center hemodialysis to a home dialysis modality (peritoneal dialysis or home hemodialysis) among eligible patients at the facility during the time period, given the national average of modality switches, and patient case-mix at the facility.

Denominator Exclusions

Patient's time at risk under hospice care, patient's time at risk when in a nursing home and on home hemodialysis, pediatric patients (less than 18 years of age), patients with no CMS-2728 Medical Evidence form (i.e., AKI patients on dialysis but not designated as ESRD).

Patients who are attributed to clinics with fewer than 1 expected modality switch are not excluded from the measure. All patients who meet the denominator inclusion criteria are included and used to model a given facilities expected switch rate to home dialysis. If that switch rate is <1, then the facility is excluded from reporting outcomes.

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

Nephrology

Measure Type

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?

Claims Data;Registries

If applicable, specify the data source

N/A

Description of parts related to these sources

Data are derived from an extensive national ESRD patient database, which is primarily based on the Renal Management Information System (REMIS), EQRS facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form and patient tracking data), the Medicare Enrollment Database (EDB), and Medicare dialysis claims data (primarily outpatient). In addition, the database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), and data from the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC) (which includes Provider and Survey and Certification data from Automated Survey Processing Environment (ASPEN)), and the Dialysis Facility Compare (DFC). Hospice information is obtained from Medicare Part A Hospice claims submitted by Hospice providers. Tracking by dialysis provider and treatment modality is available for all patients including those with only partial or no Medicare coverage

At what level of analysis was the measure tested?

Facility

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?

Submitted

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

3696

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?

Claims;Other: EQRS

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 combination of electronic sources.

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

N/A

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

Published, peer-reviewed original research

Summarize the empirical data

Home dialysis rates remain low in the United States compared with many other countries, hovering around 12% (Briggs 2019). Because there are not formal randomized controlled trials of modality uptake, the evidence for SMoSR is based on a large body of observational studies in the U.S. as well as outside the U.S. such as Canada, several European countries, and Australia and New Zealand.

We evaluated studies that examined the epidemiology and characteristics of home dialysis uptake; educational interventions and processes to support shared-decision making; and studies comparing or assessing outcomes (mortality; hospitalization) between a home dialysis modality (i.e., peritoneal dialysis) and in-center hemodialysis, or the association of home modalities with comorbidities and other health outcomes.

Clinical, operational, economic and patient factors have been identified as barriers to uptake of home dialysis modalities (Chan 2019). Clinical factors include lack of physician competency in prescribing home dialysis modalities; operational include lack of clinician and staff training; economic obstacles include lack of sufficient housing or storage space for dialysis supplies; and patient barriers include lack of

adequate education. Studies also have identified demographic characteristics of black race, male sex, older age, and comorbidities as predictors of low uptake of home dialysis; while small dialysis facility size and low physician and nurse experience with home dialysis are facility level barriers.

Studies that examine the role and impact of education on home modality uptake show that about 30% of chronic dialysis patients have reported their modality selection was not really their choice or did not feel as though they made an informed choice, and that this percentage is higher among in-center hemodialysis (ICHD) patients (Dahlerus 2016; Van Biesen 2014; Song 2013; Winterbottom 2012). Studies have also found that there is a mismatch between stated preference for dialysis modality (i.e., home dialysis) and the actual modality on which patients start. The preferred modality was a home therapy but in many cases patients started on in-center hemodialysis (Pyart 2018; Keating 2014; Liebman 2012). This suggests existing educational efforts fall short of supporting decision making by the patient. Specifically, decision-making efficacy and satisfaction of modality selection has been reported as greater among PD vs in-center HD patients (Zee 2018)

Because of the lack of RCTs comparing dialysis modalities and outcomes, the current evidence is observational in nature. Some studies have shown a survival advantage associated with PD as an initial modality however evidence is mixed about the longer term outcomes and survival benefit for PD versus in-center hemodialysis. As such, in-center and home dialysis are generally considered equivalent with respect to hospitalization rates and mortality. In one meta-review, some differences were observed in physical and mental quality of life domains between patients on PD versus in-center hemodialysis (Budhram 2020)

The evidence indicates that persistently low rates of home dialysis use are associated with both patient and facility level factors. Education and shared decision making interventions suggest an opportunity to improve uptake of home dialysis. Moreover, home modalities offer patients potential flexibility and independence.

Collectively these studies support the construct of the SMoSR which is an indicator of successful education by the facility to facilitate a decision to switch to a home modality, through on-going educational efforts after a patient starts on in-center hemodialysis.

References:

Briggs V, Davies S, Wilkie M. International Variations in Peritoneal Dialysis Utilization and Implications for Practice Am J Kidney Dis. 2019 Jul;74(1):101-110. doi: 10.1053/j.ajkd.2018.12.033. Epub 2019 Feb 22.

Budhram B, Sinclair A, Komenda P, Severn M, Sood MM. A Comparison of Patient-Reported Outcome Measures of Quality of Life By Dialysis Modality in the Treatment of Kidney Failure: A Systematic Review Can J Kidney Health Dis. 2020 Oct 19;7:2054358120957431. doi: 10.1177/2054358120957431. eCollection 2020.

Chan CT, Blankestijn PJ, Dember LM, Gallieni M, Harris DCH, Lok CE, Mehrotra R, Stevens PE, Wang AY, Cheung M, Wheeler DC, Winkelmayer WC, Pollock CA; Conference Participants. Dialysis initiation, modality choice, access, and prescription: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference Kidney Int. 2019 Jul;96(1):37-47. doi: 10.1016/j.kint.2019.01.017. Epub 2019 Apr 13.

Dahlerus C, Quinn M, Messersmith E, Lachance L, Subramanian L, Perry E, Cole J, Zhao J, Lee C, McCall M, Paulson L, Tentori F. Patient Perspectives on the Choice of Dialysis Modality: Results From the Empowering Patients on Choices for Renal Replacement Therapy (EPOCH-RRT) Study Am J Kidney Dis. 2016 Dec;68(6):901-910. doi: 10.1053/j.ajkd.2016.05.010. Epub 2016 Jun 21

Keating PT, Walsh M, Ribic CM, Brimble KS. The impact of patient preference on dialysis modality and hemodialysis vascular access BMC Nephrol. 2014 Feb 22;15:38. doi: 10.1186/1471-2369-15-38.

Liebman SE, Bushinsky DA, Dolan JG, Veazie P. Differences between dialysis modality selection and initiation Am J Kidney Dis. 2012 Apr;59(4):550-7. doi: 10.1053/j.ajkd.2011.11.040. Epub 2012 Feb 2.

Pyart R, Donovan K, Carrington C, Roberts G. Peritoneal Dialysis: Turning Choice Into Reality Perit Dial Int. 2018 Sep-Oct;38(5):328-333. doi: 10.3747/pdi.2018.00011. Epub 2018 Jul 10.

Song MK, Lin FC, Gilet CA, Arnold RM, Bridgman JC, Ward SE. Patient perspectives on informed decision-making surrounding dialysis initiation Nephrol Dial Transplant. 2013 Nov;28(11):2815-23. doi: 10.1093/ndt/gft238. Epub 2013 Jul 30.

Van Biesen W, van der Veer SN, Murphey M, Loblova O, Davies S. Patients' perceptions of information and education for renal replacement therapy: an independent survey by the European Kidney Patients' Federation on information and support on renal replacement therapy PLoS One. 2014 Jul 31;9(7):e103914. doi: 10.1371/journal.pone.0103914. eCollection 2014.

Winterbottom AE, Bekker HL, Conner M, Mooney AF. Patient stories about their dialysis experience biases others' choices regardless of doctor's advice: an experimental study Nephrol Dial Transplant. 2012 Jan;27(1):325-31. doi: 10.1093/ndt/gfr266. Epub 2011 Jun 3.

Zee J, Zhao J, Subramanian L, Perry E, Bryant N, McCall M, Restovic Y, Torres D, Robinson BM, Pisoni RL, Tentori F. Perceptions about the dialysis modality decision process among peritoneal dialysis and in-center hemodialysis patients BMC Nephrol. 2018 Oct 29;19(1):298. doi: 10.1186/s12882-018-1096-x.

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

117,942

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

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment;Severity of Illness

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

The C-Statistic=0. 674, which suggests good predictive ability of the risk model.

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), Profile Inter Unit Reliability (PIUR)

Signal-to-Noise: Sample size

6779-7220

Signal-to-Noise: Statistical result

0.605-0.606

Signal-to-Noise: Interpretation of results

The IUR is moderate (0.605) and indicates that the measure can detect differences in performance scores across facilities. As noted above, the PIUR (0.606) measures reliability in terms of reflagging rates but is places on the same scale as IUR. A PIUR that is larger than the IUR indicates that the measure has a higher reliability for identifying extreme values. In this case, the IUR and PIUR are nearly the same, so the IUR also is descriptive of the measures usefulness in identifying extreme values.

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

Gamma statistic

Empiric Validity: Sample size

6779-7220

Empiric Validity: Statistical result

0.29

Empiric Validity: Methods and findings

Validity of the Standardized Modality Switch Ratio was assessed using several different statistical tests to examine the relationship with other facility level quality measures: Standardized Mortality Ratio (SMR), First-Year Standardized Mortality Ratio (FYSMR), Standardized Hospitalization Ratio (SHR), Standardized Waitlist Ratio-Incident Dialysis Patients (SWR), ICH-CAHPS "Providing information to patients" and the percentage of home dialysis patients at the facility.

Spearman's rho Correlations with Quality Outcome Performance Measures:

We started by calculating Spearman's rho coefficient to examine the correlation of SMoSR with SMR, FYSMR, SHR, and SWR. Spearman's correlation coefficient, which is a rank-based correlation metric, was chosen for its robustness against potential extreme providers and tied providers. The peer-reviewed literature is mixed in regard to whether home dialysis compared to in-center dialysis offers better survival or lower hospitalization rates. Therefore, we hypothesized no or weak correlations of SMoSR with SMR, FYSMR, and SHR. However, facility processes of care that support robust modality education should result in higher referral for transplant evaluation and subsequent waitlisting. Therefore, we hypothesized a positive correlation between SMoSR and SWR. Table 1 reports the estimated Spearman's rho correlations.

Gamma Tests for Concordance Analysis with Performance Classification:

Next, we performed gamma tests to examine the concordance of facility level SMoSR flagging classifications (Better than Expected, "As Expected", and "Worse than Expected" with 2019 SWR. The choice of gamma tests in the analysis is due to the fact that these performance categories are naturally ordered in a descending order.

A positive Gamma coefficient would indicate a concordance in flagging categories between SMoSR and an existing performance measure. In contrast, a negative Gamma signifies a discordant relationship. The null hypothesis of Gamma=0 is set up to test for a significant correlation. The higher a Gamma value the stronger the relationship. We hypothesized that there would be moderate agreement in facility classification of performance between the SMoSR and the first year SWR. The estimated magnitude of concordance is provided in Table 2.

Association with patient reported outcomes: ICH-CAHPS "Providing information to patients"

The In-Center Hemodialysis Consumer Assessment of Healthcare Provider and Systems (ICH-CAHPS)1 is a patient reported experience of care survey to measure in-center hemodialysis patients perspectives on the care they receive at dialysis facilities. This measure is reported on Dialysis Facility Care Compare. We computed a Pearson correlation (rho) to assess the association between the ICH-CAHPS mean scores for the 9 question composite measure on "providing information to patients" and SMoSR performance classifications of "better than expected", "as expected" and "worse than expected."

Collectively the ICH-CAHPS linearized top box score for "providing information" indicates how well the facility is doing providing information on safety as well as all renal replacement modalities, including home dialysis and transplant. Since this facility process of modality education is a critical step for many patients to understand their treatment choices, we expect a higher proportion of patients reporting "on facilities always providing information" will be associated with a better performance classification on SMoSR. Please see Table 3 below for this association and the Pearson's correlation r statistic.

Association between the percentage of home dialysis patients and performance on SMoSR:

We computed a Pearson correlation rho to assess the association between the different SMoSR performance classifications and the percentage of home dialysis patients at a facility. The proportion of home dialysis patients at a facility reflects the processes that are in place to provide effective modality education and then facilitate a transfer from in-center to home dialysis. We expect a better SMoSR performance classification to be associated with a higher percentage of patients on home dialysis at a facility. Table 4 reports these results and the Pearson correlation r statistic.

Two-part Semi-continuous Model:

A challenge with the analysis for the association between SMoSR and the percentage of home dialysis patients at a facility is that some facilities have no home program resulting in zero patients on home dialysis. This cluster "zero-patient" facilities will distort the correlation calculation due to the significant amount of ties. One option is to delete these facilities from the calculation. However, such an approach would then be based on a selective sub-sample which may introduce bias. To avoid this, we used a two-part semi-continuous regression model that accommodates data that have both a spike at zero and continuous values over the nonzero part (Atchison 1995). In the first part, we used a logistic regression model to predict the propensity of observing facilities with zero (vs. nonzero) percentage of home dialysis patients as a function of the SMoSR, adjusted for a set of facility characteristics. For the second part of the model, a linear regression is fit only among the subset of facilities with non-zero number of home dialysis patients using SMoSR as the predictor for the percentage of home dialysis patients. We adjusted for the same set of facility characteristics as the binary part. The two models are connected formally through a mixture structure, where the mixing proportion is estimated from the data.

For the logistic model, we expect a higher SMoSR value to be associated with lower odds of facilities having zero home dialysis patients; whereas for the linear model, we expect a positive association between SMoSR and the percentage of home dialysis patients. These results are presented in Table 5 below.

In addition to the above mentioned statistical tests, the validity of the measure is also based on face validity. The SMoSR was reviewed by a TEP in 2021 which supported the measure construct and provided input on the SMoSR risk adjustment and exclusion methodology.

References:

Aitchison J. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of The American Statistical Association 1955; 50: 901

University of Michigan Kidney Epidemiology and Cost Center. Effective Availability and Utilization of Home Dialysis Technical Expert Panel Summary Report, Prepared for The Centers for Medicare and Medicaid Services. June, 2021.

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

1.07

Median performance score

.84

Minimum performance score

0.0

Maximum performance score

15.285

Standard deviation of performance scores

1.00

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

Golden Horton
7500 Security Boulevard
Baltimore, MD 21244
Golden.horton@cms.hhs.gov
(410) 786-4024

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Jaclyn George
1415 Washington Heights
Ann Arbor, MI 48109
jaclynrg@med.umich.edu
(734) 837-7237

Secondary Submitter Contact Information

Jennifer Sardone
1415 Washington Heights
Ann Arbor, MI 48109
jmsto@med.umich.edu
(734) 837-7237

Submitter Comments

For the estimated impact of the measure, 117,942 is the number of patients in 2019 that were included in the calculations in the testing form. The MJF should be referenced for greater detail about validity testing and risk adjustment.

MUC2022-076 Standardized Fistula Rate for Incident Patients

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The Standardized Fistula Rate (SFR) for Incident Patients is based on the prior SFR (NQF #2977) that included both incident and prevalent patients. This measure was initially endorsed in 2016, but as part of measure maintenance review by the NQF Standing Committee in 2020, concerns were raised about the strength of evidence supporting the prior measure. Namely, recent updates to the KDOQI guidelines downgraded the evidence supporting fistula as the preferred access type and instead focus on catheter avoidance and developing an individualized ESKD Life plan.

Numerator

The numerator is the adjusted count of adult incident patient-months using an AVF as the sole means of vascular access as of the last hemodialysis treatment session of the month.

Numerator Exclusions

N/A

Denominator

All patient-months for patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD) and became ESRD within the prior 12 months for the entire reporting month at the same facility.

Denominator Exclusions

Exclusions that are implicit in the denominator definition include:

Patient-months after 12 months of starting ESRD

Pediatric patients (<18 years old)

Patients-months on Peritoneal Dialysis

Patient-months with in-center or home hemodialysis for less than a complete reporting month at the same facility

In addition, the following exclusions are applied to the denominator:

Patients with a catheter that have limited life expectancy:

Patients under hospice care in the current reporting month

Patients with metastatic cancer in the past 12 months

Patients with end stage liver disease in the past 12 months

Patients with coma or anoxic brain injury in the past 12 months

The denominator is defined at the patient level not facility level. The reason this rule is applied is to comport with how measures are implemented for public reporting. Due to small cell size and potentially identifiable data, facilities with <11 patients do not receive a score.

As stated in the measure description and rationale, this is a measure of incident patients only. Dialysis patients in their first 12 months of ESRD are more likely to be using a catheter for vascular access and in turn are at higher risk for CVC related infections. The measure focus is on the first 12 months of dialysis since this is the most active time of vascular access creation and where the potential benefit is greatest relative to treatment with a CVC.

Patient attribution to facilities is already described - Patients are required to have been treated by the same facility for the complete month in order to be assigned to that facility for the reporting month.

When a patient is not treated in a single facility for a span of 30 days (for instance, if there were two facility transfers within 30 days of each other), we do not attribute that patient to any facility for that month. Therefore, transient treatment at a facility due to either travel or a temporary clinical condition do not impact the fistula rate of that facility.

Patients with a catheter (of any duration) AND one or more of the limited life expectancy exclusions are excluded from the denominator.

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

Nephrology

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?

Claims Data;Registries

If applicable, specify the data source

N/A

Description of parts related to these sources

Data are derived from an extensive national ESRD patient database, which is primarily based on the Renal Management Information System (REMIS), EQRS facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form and patient tracking data), the Medicare Enrollment Database (EDB), and Medicare claims data. In addition the database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), and data from the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC) (which includes Provider and Survey and Certification data from Automated Survey Processing Environment (ASPEN)), and the Dialysis Facility Compare (DFC). The database is comprehensive for Medicare patients not enrolled in Medicare Advantage. Medicare Advantage patients are included in all sources but their Medicare payment records are limited to inpatient claims. Non-Medicare patients are included in all sources except for the Medicare payment records. Tracking by dialysis provider and treatment modality is available for all patients including those with only partial or no Medicare coverage. Past-year comorbidity data are obtained from multiple claim types (inpatient, outpatient, home health, hospice, skilled nursing facility claims). EQRS is the data source for establishing the vascular access type used to determine the numerator.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Dialysis 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

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?

Submitted

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

3659

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 Fistula Rate, ESRD QIP

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

This measure focuses on incident patients; the current SFR includes both incident and prevalent patients.

How will this measure add value to the CMS program?

This measure is a refinement of the SFR, and would serve as a replacement of that measure (they would not be reported together)

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: EQRS

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

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

Three main sources of input were used to provide the evidence base for this measure: 1. In 2015, UM-KECC held a Technical Expert Panel (TEP) to seek input for development of a standardized fistula measure. The TEP agreed that AVF are the preferred access for most patients, and that AVG were still preferred relative to a vascular catheter. The TEP recommended that the AVF measure should be adjusted for conditions where an AVG may be an acceptable alternative such as: older age, diabetes, vascular disease, and BMI. Of note, three of our TEP members went on to author/edit the revised KDOQI Vascular Access Guidelines that were published in 2020. KDOQI Vascular Access Guidelines: In general, the evidence for the guidelines has been rated as either low or moderate, with many of the guidelines relying on expert opinion. The evidence review team focused on 16 studies and noted that bloodstream infections were significantly lower among patients who started HD with an AV fistula or AV graft versus a catheter. The workgroup refrained from recommending AV fistula on the basis of lower mortality compared to catheter use, and instead relied on the evidence indicating lower blood stream infections.

The two guidelines that are most directly relevant to this measure are the following: 1. KDOQI suggests that if sufficient time and patient circumstances are favorable for a mature, usable AVF, such a functioning AVF is preferred to an AVG in incident HD patients due to fewer longterm vascular access events (eg, thrombosis, loss of primary patency, interventions) associated with unassisted AVF use. (Conditional Recommendation, Low Quality of Evidence) 2. KDOQI suggests that most incident HD patients starting dialysis with a CVC should convert to either an AVF or AVG, if possible, to reduce their risk of infection/bacteremia, infection-related hospitalizations, and adverse consequences. (Conditional Recommendation, Very Low-Moderate Quality of Evidence) 3. From the peer reviewed literature presented in the revised guidelines, the core evidence that the Workgroup was most compelled by centered around the lower rates of bloodstream infection associated with AVF compared to CVCs. This

measure relies on those studies that highlight lower infection risk with AVF that have withstood the enhanced scrutiny of the evidence review team, who noted many vascular access studies were observational in nature and thus had a potential risk of bias. 3. Recent peer-reviewed literature continues to highlight benefits of AVF over CVCs particularly for incident patients who are in their first year of dialysis. Notable recent findings include: AVF associated with lower risk of access-related hospitalizations: One study using USRDS data that focused on elderly patients who started dialysis with a catheter and had an AVF or AVG created within the first 6 months of dialysis noted that AVF creation was associated with a lower risk of access-related hospitalization. AVF has lower rates of blood stream infection and sepsis compared to AVG or CVC: In a study of 2352 incident dialysis patients, after adjusting for confounders, AVF use was associated with 61% lower risk of blood stream infections compared with CVC or AVG use. In a separate study, based on the vascular access used at initiation of dialysis, patients with AVG (HR 1.35) and CVC (HR 1.80) were more likely to develop sepsis (both P < .001). Additionally, in patients who developed sepsis, mortality at 1 year after sepsis was 21% higher in AVG and nearly doubled in CVC when compared to AVF. A third study of patients over the age of 67 who start dialysis with a catheter and went on to have either an AVF or AVG placed in the first 6 months reports that rates of all-cause infection-related hospitalization (RR 0.93, P=0.01) and bacteremia/septicemia-related hospitalization (RR 0.90; P=0.02) were lower in the AVF group versus AVG group. AVF have lower maintenance interventional requirements compared to AVG: Using USRDS data and accounting for patient characteristics, one study reported that during maturation of the AV access, interventions for both AVFs and AVGs were relatively common and similar between the two types of access. However, once successfully matured, AVFs had lower maintenance interventional requirements. Catheter dependence after AVF or AVG placement among elderly incident dialysis patients is complex: for many younger hemodialysis patients, creation of an AVF, compared with an AVG, is associated with longer initial catheter dependence, but then longer access survival and lower long-term catheter dependence. In patients 67 years of age, similar increased catheter dependence was found at 1 and 3 months after AVF creation, compared to AVG, but lower catheter dependence at 12 and 36 months. However, creation of AVF in the older population was associated with greater cumulative catheter-dependent days (80 vs 55 days per person-year) after 3 years of follow up.

From a cost perspective, Hall and colleagues report11 that based on Markov models of hypothetical patients starting dialysis with a CVC, the AVF option was cost effective compared with continued catheter use for all age and life expectancy groups, except for 85-89 year olds in the lowest life expectancy quartile. The AVF option was more cost effective than the AVG option for all quartiles of life expectancy among the 65- to 69-year-old age group. For older age groups, differences in cost-effectiveness between the strategies were attenuated, and the AVF option tended to only be cost effective in patients with life expectancy >2 years. These findings highlights that not all elderly patients will realize the benefit of catheter independence from AVF creation and specific patient characteristics and shared decision making remain critical in appropriate vascular access selection. AVF is associated with higher health-related quality of life (HRQOL) and less depression compared to CVC in the first year of dialysis: A prospective cohort study of 1461 patients who initiated dialysis reported that patients with an AVF had higher KDQOL-36 scores and lower Beck Depression Inventory scores at 3 months and 12 months after the initiation of dialysis compared to those with CVC. Furthermore, in a survey conducted by the American Association of Kidney Patients, satisfaction with current vascular access was 90% with AVF, 79% with AVG, and 67% with CVC. The factors most frequently reported as important in influencing the selection of vascular access modality included infection risk (87%), physician recommendation

(84%), vascular access durability (78%), risk of complications involving surgery (76%), and impact on daily activities (73%). As we navigate vascular access decisions that embrace shared decision making and respect patient choice, these two studies highlight that the majority of patients who choose an AVF are satisfied with that decision and may enjoy better health-related quality of life. In summary, the recently revised KDOQI guidelines for vascular access continue to support AV fistula as the preferred vascular access for most patients on dialysis, although with less emphasis than in prior iterations. Long-term catheters are still viewed as the least desirable vascular access, primarily due to the increased risk of blood-stream infections, with increased recognition of certain patient characteristics and scenarios where this access type may be the most appropriate. Given that over 80% of new hemodialysis patients start with a CVC, the additional studies noted above that were published after the updated KDOQI guidelines suggest that attempts to create AVF are still warranted.

Summarize the evidence

There are three major healthcare processes associated with achieving AVF creation:

Patient education interventions: Providing kidney disease education is associated with 1.78 increased odds of starting dialysis with an AVF and a 0.51 odds of starting dialysis with a CVC alone. Patient education can enhance motivation and potentially lead to improved health outcomes.

Vascular Access Coordinator/Program: In one study, an organized dialysis access program resulted in a 82% decrease in the number of central venous catheter days which lead to a concurrent reduction in central line-associated bloodstream infection and deaths. As a result of creating an access program, central venous catheter rates decreased from an average rate of 45% to 8%.

Surgeon Selection: Several studies have suggested that there is significant variation in likelihood of AVF, as opposed to AVG, creation based on the vascular access surgeon. Using a national claims database to identify patients initiating hemodialysis with a CVC, and adjusting for demographic and comorbid conditions, the individual surgeon identifier had the greatest magnitude of effect on access type (AVF or AVG) created, with some surgeons more than twice as likely to create AVF as other surgeons. Thus, surgeon selection by the dialysis facility is an important component in efforts to maximize creation of AVF in otherwise eligible patients.

Comorbidity adjustment : One frequently cited barrier to successful AVF creation has been the burden of comorbidities at the dialysis facility level. A recent study noted that after adjustment for facility-level comorbidity burden, only small differences in facility rates of AVF use were seen except in the extremes of high or low levels of comorbidity burden. This suggests that dialysis facilities with a relatively high patient comorbidity burden can achieve similar fistula rates as facilities with healthier patients if the above care processes are employed.

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

1,871,951

Type of Evidence to Support the Measure

Other (enter here):: EQRS

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment;Severity of Illness

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

The C-statistic was 0.748. This indicates 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), Profile IUR

Signal-to-Noise: Sample size

6355-6659

Signal-to-Noise: Statistical result

0.705-0.970

Signal-to-Noise: Interpretation of results

The result of IUR (0.705) and PIUR (0.970) testing suggests 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

Poisson regression models

Empiric Validity: Sample size

6355-6659

Empiric Validity: Statistical result

1.13

Empiric Validity: Methods and findings

Validity was assessed using Poisson regression models to measure the association between facility level quintiles of performance scores and the 2018-2019 Standardized Mortality Ratio (SMR, NQF 0369), 2018-2019 Standardized Hospitalization Ratio (SHR, NQF 1463), and 2018 First Year Standardized Mortality Ratio (SMR) respectively. Facility-level performance scores were divided into quintiles (Q1 to Q5), and the relative risk (RR) for SMR (and SHR and first year SMR, separately) was calculated for each quintile, using Q5 as the reference group. A RR>1.0 would indicate a higher relative risk of mortality or hospitalization, compared to the lowest performance score quintiles.

For the all-cause hospitalization rate and vascular access infection related hospitalization rate, we used linear regression to test the association between the SFR quintiles and the 2018-2019 all-cause hospitalization rate, and 2018-2019 vascular access related infection hospitalization rate, respectively. For all-cause hospitalization and vascular access related infection hospitalization, the respective rate was calculated for each quintile and a trend test of the rates was performed.

SMR: We expect a negative association with SMR since successfully creating an AVF is generally seen as representing a robust process to coordinate care outside of the dialysis facility. AVFs are also associated with lower risk of infection which may reduce the risk of a life threatening infection or other poor outcomes that place patients at higher risk of mortality. Higher standardized fistula rates will be negatively associated with SMR.

SHR: We expect a negative association with SHR since successfully creating an AVF is generally seen as representing a robust process to coordinate care outside of the dialysis facility, and potentially reduces the risk for patients at such facilities going to hospital due to infections or other acute clinical events. Higher standardized fistula rates will be negatively associated with SHR.

First Year SMR: We expect a negative association with the first year SMR as many incident patients begin with a catheter, and therefore face higher risk for infection compared to patients with an AVF. AVFs are associated with lower risk of infection which may reduce the risk of a life threatening infection or other poor outcomes that place patients at higher risk of mortality particularly in their first year of dialysis. Higher standardized fistula rates will be negatively associated with the first year SMR.

All-cause hospitalization rate: We expect a negative association between all-cause hospitalization rates and higher AVF rates given the known risk of infection and other complications related to long-term catheter dependence, particularly in incident patients.

Vascular access related infection hospitalization rate: We expect a negative association between access related infection hospitalizations and AVF rates because of the higher rates of catheter in patients in the first year of dialysis, which creates a higher risk of a catheter related infection.

Cut-points for the quintiles of the performance scores were defined as follows:

Q1: 0% - <30.8%

Q2: 30.8% - <38.3%

Q3: 38.3% - <44.6%

Q4: 44.6 - <52.1%

Q5: 52.1% - <99.0% as Reference

Results from the Poisson model indicated that the percent of patient-months with a fistula was significantly associated with the risks of mortality and hospitalization.

For the 2018-2019 SMR, the relative risk of mortality increased as the performance measure quintile decreased from the reference group (Q5) with the highest risk in quintile 1. For quintile 4, RR=1.02 (95% CI: 1.00, 1.04; p<0.001), quintile 3, RR=1.06 (95% CI: 1.04, 1.08; p<0.001), quintile 2, RR=1.08 (95% CI: 1.06, 1.10; p<0.001), and quintile 1, RR=1.13 (95% CI: 1.11, 1.15; p<0.001).

Similarly for 2018-2019 SHR, the relative risk of hospitalization increased as the performance measure quintile decreased from the reference group (Q5) with the highest risk in quintile 1. For quintile 4, RR=1.06 (95% CI: 1.05, 1.06; p<0.001), quintile 3, RR=1.07 (95% CI: 1.06, 1.07; p<0.001), quintile 2, RR=1.11 (95% CI: 1.10, 1.12; p<0.001), and quintile 1, RR=1.15 (95% CI: 1.14, 1.15; p<0.001).

For the 2018 first year SMR, the relative risk of mortality increased as the performance measure quintile decreased from the reference group (Q5)with the highest risk in quintile 1. For quintile 4, RR=1.08 (95% CI: 1.03, 1.14; p=0.002), quintile 3, RR=1.11 (95% CI: 1.05, 1.16; p<0.001), quintile 2, RR=1.17 (95% CI: 1.12, 1.23; p<0.001), and quintile 1, RR=1.53 (95% CI: 1.46, 1.60; p<0.001).

For the 2018-2019 all-cause hospitalization, the hospitalization rate decreased as the performance measure quintile increased. Hospitalization rates for quintiles 1 to 5 are 1.06, 0.99, 0.95, 0.93, and 0.87 patient-years respectively (trend test p<0.001).

For the 2018-2019 vascular access related infection hospitalization, the hospitalization rate decreased as the performance measure quintile increased. Hospitalization rates for quintiles 1 to 5 are 0.22, 0.18, 0.17, 0.16, and 0.15 respectively (trend test p<0.001).

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

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

41.4

Median performance score

41.6

Minimum performance score

0.0

Maximum performance score

99.0

Standard deviation of performance scores

12.7

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

Golden Horton
7500 Security Boulevard
Baltimore, MD 21244
Golden.horton@cms.hhs.gov
(410) 786-4024

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Jaclyn George
1415 Washington Heights
Ann Arbor, MI 48109
jaclynrg@med.umich.edu
(734) 837-7237

Secondary Submitter Contact Information

Jennifer Sardone
1415 Washington Heights
Ann Arbor, MI 48109
jmsto@med.umich.edu
(734) 837-7237

Submitter Comments

For estimated impact of the measure, 1,871,951 is the number of patient months in the reporting period used in the calculations in the testing form. The MJF should be referenced for greater detail about validity testing and risk adjustment.

MUC2022-079 Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The Standardized Emergency Department Encounter Ratio is defined to be the ratio of the observed number of emergency department (ED) encounters that occur for adult Medicare ESRD dialysis patients treated at a particular facility to the number of encounters that would be expected given the characteristics of the dialysis facility's patients and the national norm for dialysis facilities. Note that in this document an emergency department encounter always refers to an outpatient encounter that does not end in a hospital admission. This measure is calculated as a ratio but can also be expressed as a rate.

Numerator

The observed number of outpatient Emergency Department encounters during the reporting period among eligible adult Medicare patients at a facility.

Numerator Exclusions

N/A

Denominator

The expected number of Emergency Department encounters among eligible Medicare patients at the facility during the reporting period adjusted for the characteristics of the patients at the facility.

Denominator Exclusions

Exclusions that are implicit in the denominator definition include time at risk while a patient:

Has Medicare Advantage coverage

Has had ESRD for 90 days or less

Is less than 18 years of age

The denominator also excludes patient time at risk for calendar months in which a patient is:

Actively enrolled in hospice at any time during the calendar month

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Medicare FFS

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

Nephrology

Measure Type

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?

Claims Data;Registries

If applicable, specify the data source

N/A

Description of parts related to these sources

Data are derived from an extensive national ESRD patient database, which is primarily based on the Renal Management Information System (REMIS), EQRS facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form and patient tracking data), the Medicare Enrollment Database (EDB), and Medicare claims data. In addition the database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), and data from the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC) (which includes Provider and Survey and Certification data from Automated Survey Processing Environment (ASPEN)), and the Dialysis Facility Compare (DFC). The database is comprehensive for Medicare patients not enrolled in Medicare Advantage. Medicare Advantage patients are included in all sources but their Medicare payment records are limited to inpatient claims. Non-Medicare patients are included in all sources except for the Medicare payment records. Tracking by dialysis provider and treatment modality is available for all patients including those with only partial or no Medicare coverage. Past-year comorbidity data are obtained from multiple claim types (inpatient, outpatient, home health, hospice, skilled nursing facility claims). EQRS is the data source for establishing the vascular access type used to determine the numerator.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Dialysis facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Seamless Care Coordination

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

N/A

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

Yes

CMIT ID

05676

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Endorsed

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

3565

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

2020

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

2024

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?

Claims;Other: EQRS

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

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

Published, peer-reviewed original research

Summarize the empirical data

Among Medicare beneficiaries, 30% of hospital admissions that originate in the ED are for diagnoses that are often dialysis related such as complications of vascular access, congestive heart failure/fluid overload, septicemia, and hyperkalemia [1]. Recent research points to many additional opportunities to further reduce unnecessary ED use in this population. Programs developed to impact dialysis provider practices have been shown to improve intermediate outcomes (reduced catheter vascular access [3], small solute adequacy, anemia management), hospitalization, and mortality.

Cohen and colleagues [9] reported that missed dialysis treatments are associated with an over two-fold higher risk of an ED visit, suggesting an opportunity for dialysis facilities to establish or strengthen

facility practices that can help to reduce skipped treatments through increased communication, care coordination, and patient education. This in turn has the potential to reduce avoidable ED visits. Given the association between missed dialysis treatments and increased risk of an ED visit [4], dialysis facility interventions that improve adhearance to the treatment schedule would be expected to decrease ED utilization. Other interventions, such as telehealth, have been demonstrated to reduce ED utilization in high-risk dialysis patients [5].

Zhang and colleagues [10] reported that rates of ED visits among patients on thrice weekly in-center hemodialysis vary by dialysis schedule (Mon/Weds/Fri; Tues/Thurs/Sat) and by day of week. For example the ED visit rate (without hospital admission) was highest on the day following the longer interdialytic interval over the weekend (Mondays), suggesting an association with facility structure and treatment schedule.

In the general population, outpatient ED visits were reported to have increased more slowly for Medicare patients being treated by patient-centered medical home practices when compared to non-patient-centered medical homes[6]. A comparable example that may hold promise of reducing ED use among ESRD dialysis patients is the current CMS Centers for Medicare and Medicaid Innovation Comprehensive End Stage Renal Disease (ESRD) Care model that emphasizes care coordination as a central feature of care delivery in order to reduce utilization and improve outcomes. During the second performance year, the original Wave 1 cohort of ESCOs (ESRD Seamless Care Organizations) experienced about a 3% reduction in ED use relative to the period before the CEC model was launched [11].

Finally, low health literacy has been associated with increased use of ED services [7] and some studies have indicated that patient education interventions can reduce ED utilization [8].

1. Lovasik, B.P., et al., Emergency Department Use and Hospital Admissions Among Patients With End-Stage Renal Disease in the United States. JAMA Intern Med, 2016. 176(10): p. 1563-1565.
2. Centers for Disease Control and Prevention. National hospital ambulatory medical care survey: 2011 emergency department summary tables. http://www.cdc.gov/nchs/fastats/injury.htm 2011 [cited 2017 January 9].
3. Ng LJ, Chen F, Pisoni RL, Krishnan M, Mapes D, Keen M, Bradbury BD. Hospitalization risks related to vascular access type among incident US hemodialysis patients. Nephrol Dial Transplant. 26(11):3659-66, 2011
4. Chan, K. E.;Thadhani, R. I.;Maddux, F. W. Adherence barriers to chronic dialysis in the United States. J Am Soc Nephrol. 2014 25(11):2642-8 doi:10.1681/asn.2013111160
5. Minatodani, D. E.;Berman, S. J. Home telehealth in high-risk dialysis patients: a 3-year study. Telemed J E Health. 2013 19(7):520-2 doi:10.1089/tmj.2012.0196
6. Pines, J. M.;Keyes, V.;van Hasselt, M.;McCall, N. Emergency department and inpatient hospital use by Medicare beneficiaries in patient-centered medical homes. Ann Emerg Med. 2015 65(6):652-60 doi:10.1016/j.annemergmed.2015.01.002
7. Green, J. A.;Mor, M. K.;Shields, A. M.;Sevick, M. A.;Arnold, R. M.;Palevsky, P. M.;Fine, M. J.;Weisbord, S. D. Associations of health literacy with dialysis adherence and health resource utilization in patients receiving maintenance hemodialysis. Am J Kidney Dis. 2013 62(1):73-80 doi:10.1053/j.ajkd.2012.12.014

8. Morgan, S. R.;Chang, A. M.;Alqatari, M.;Pines, J. M. Non-emergency department interventions to reduce ED utilization: a systematic review. Acad Emerg Med. 2013 20(10):969-85 doi:10.1111/acem.12219
9. Cohen D, Gray K, Colson C, Van Wyck D, Tentori F, and Brunell S. Impact of Rescheduling a Missed Hemodialysis Treatment on Clinical Outcomes. Kidney Med. 2(1):12-19.Published online December 11, 2019
10. Zhang S, Morgenstern H, Albertus P, Nallamothu B, He K, and Saran R. Emergency department visits and hospitalizations among hemodialysis patients by day of the week and dialysis schedule in the United States. PLOS ONE. https://doi.org/10.1371/journal.pone.0220966 August 15, 2019.
11. Marrufo G, Negrusa B, Ullman D, Hirth R, Messana J, Maughan B, Nelson J, Lindsey N, Gregory D, Svoboda R, Melin C, Chung A, Dahlerus C, Nahra T, Jiao A, McKeithen K, and Gilfix Z. Comprehensive End-Stage Renal Disease Care (CEC) Model. Performance Year 2 Annual Evaluation Report. Prepared for: Centers for Medicare & Medicaid Services. September 2019. https://innovation.cms.gov/Files/reports/cec-annrpt-py2.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

383,414

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

Patient-level demographics: please select all that apply:

Age;Sex

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

Case-Mix Adjustment;Severity of Illness

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

The estimate of the C-statistic for the SEDR is 0.61.

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), Profile Inter Unit Reliability (PIUR)

Signal-to-Noise: Sample size

6056-6691

Signal-to-Noise: Statistical result

0

Signal-to-Noise: Interpretation of results

The value obtained for the IUR is moderate in size. The PIUR is larger and demonstrates that the SEDR is effective at detecting outlier facilities and statistically meaningful differences in performance scores across dialysis facilities.

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

Classification of SEDR and mean facility performance scores for Related Measures

Empiric Validity: Sample size

6659

Empiric Validity: Statistical result

0

Empiric Validity: Methods and findings

To validate SEDR we first stratified facilities into the better than/as expected and worse than expected categories of SEDR. Next we calculated mean performance scores for several quality measures: Standardized Mortality Ratio (SMR), Standardized Transfusion Ratio (STrR), Standardized Fistula Rate (SFR), Percentage of Prevalent Patients Waitlisted (PPPW), Standardized Hospitalization Ratio (SHR), and Emergency Department Visit within 30 days of discharge (ED30). We then compared mean performance scores across the two strata of better than/as expected and worse than expected categories for SEDR.

Statistically significant outliers (i.e., better and worse than expected) were determined using the method described in section 2b4.1 to flag facilities as better than expected and worse than expected based on the national average, at the $p<0.05$ level.

We expect better mean performance on the above quality measures for facilities classified as better than/as expected for SEDR compared to facilities classified as worse than expected. Compared to facilities that perform worse than expected, facilities that perform better than/as expected on SEDR are likely to have more successful care coordination and other processes of care in place that may help patients avoid an ED visit:

SMR: We expect to observe a lower mean standardized mortality ratio for facilities in the better than/as expected category for SEDR compared to facilities classified as worse than expected. Facilities with a higher rate of ED utilization may not have care processes in place to support management of acute care.

STrR: We expect to observe a lower mean standardized transfusion event ratio for facilities in the better than/as expected category for SEDR compared to facilities classified as worse than expected. Facilities that have a lower STrR likely have processes of care in place to support robust anemia management and other care processes compared to facilities with a higher STrR.

Standardized Fistula Rate (SFR): We expect to observe a higher mean standardized fistula rate for facilities in the better than/as expected category for SEDR compared to facilities classified as worse than expected. AVFs are typically considered to be the preferred vascular access due to lower risk of infection and potential need for hospitalization or other acute care. Higher standardized fistula rates suggests facilities are successful at creating AVFs due to more robust processes to coordinate care outside of the dialysis facility. Facilities that do a better job at care coordination reduce the likelihood that patients will experience a preventable and unscheduled acute event resulting in an ED visit.

PPPW: We expect to observe a higher mean standardized percentage of prevalent patients on the waitlist for facilities in the better than/as expected category for SEDR compared to facilities classified as worse than expected. Facilities that have a higher standardized percentage of patients on the transplant waitlist suggest they may have more robust processes to coordinate care outside of the dialysis facility with other providers and the transplant center, compared to facilities with lower percentages. This includes the facility taking steps to ensure patients maintain sufficient health status in order to be placed on the waitlist. Therefore, facilities that have higher standardized waitlist percentages are likely deploying effective care coordination and other care processes that may reduce the likelihood of patients getting preventable and unscheduled acute care from the ED.

SHR: We expect that facilities classified as worse than expected for SEDR will have a standardized hospitalization ratio that is close to the national norm. SEDR only captures outpatient ED visits that do not result in an admission which, by definition, is a different patient subpopulation than SHR. Patients that require acute care from the ED without an admission likely have lower acuity medical needs that can be handled in an outpatient setting without admission. Therefore we do not expect SEDR flagging to be related to how facilities perform on SHR.

ED30: We expect to observe a lower mean ED30 ratio for facilities classified as better than/as expected for SEDR compared to facilities classified as worse than expected since both measures are a reflection of

outpatient ED use. However the measures represent two different aspects of dialysis patients emergency department use that assess complementary elements of facility care. A low SEDR, corresponding to low overall emergency department encounter rates, indicates that the facility has processes (e.g. patient/staff education, assistance with primary care, frequent evaluation of target weight) in place to avoid the need for unscheduled acute care. A low ED30 indicates that a facility is successful in managing the transition of care (e.g. medication reconciliation, evaluation of target weight, assistance with follow up appointments) that occurs after a hospital discharge.

See attached MJF for results and discussion.

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

Lower score is better

Mean performance score

1.00

Median performance score

1.00

Minimum performance score

0.0

Maximum performance score

4.30

Standard deviation of performance scores

0.34

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

Golden Horton
7500 Security Boulevard
Baltimore, MD 21244
Golden.horton@cms.hhs.gov
(410) 786-4024

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Jaclyn George
1415 Washington Heights
Ann Arbor, MI 48019
jaclynrg@med.umich.edu
(734) 837-7237

Secondary Submitter Contact Information

Jennifer Sardone
1415 Washington Heights

Ann Arbor, MI 48109
jmsto@med.umich.edu
(734) 837-7237

Submitter Comments

For the estimated impact of the measure, 383,414 is the number of patients in 2017 that were included in the calculations in the testing form. The MJF should be referenced for greater detail about validity testing and risk adjustment.

Hospital Inpatient Quality Reporting Program

MUC2022-018 Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Hospital Level – Inpatient)

Program

Hospital Inpatient Quality Reporting Program

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 hospital 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 re 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?

Facility

In which setting was this measure tested?

Community hospital;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

N/A - not a MIPS measure

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 ID: 3663e

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: 1074 (QDM version) and 1075FHIR (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?

There are no Hospital IQR measures addressing CT or radiation doses. There are three process measures related to CT in hospital outpatient settings, but none directly addresses radiation dose: (1) Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke who Received Head CT or MRI Scan Interpretation Within 45 Minutes of ED Arrival, CMIT 918; (2) Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac Low-Risk Surgery, CMIT 1367; and Abdomen Computed Tomography (CT) Use of Contrast Material, CMIT 2599. Three existing MIPS 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 7 different EHR systems reflecting 15 inpatient hospital sites [Epic (N=5), Cerner (N=1), Allscripts (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 at 15 inpatient hospital sites – the rates of out-of-range exams varied 16%-43% by site. 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. (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 $ 3250 per hospital 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: We estimate the implementation costs at $3250 per hospital based on the time and costs reported by our field-testing sites (see feasibility attachment for more information). This estimate is likely conservative, as our testing partners noted that the work of assembling the relevant data decreased over time. For hospitals that are part of large health systems, the cost may be incurred at the health system level rather than at the level of the individual hospital.

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.

15

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

15

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?

15

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.

Random Split-Half Correlation: Sample size

15

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 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, the measure is reliable at the hospital level. Given the high volume of CT, virtually no hospitals would fall below the minimum denominator to achieve ICC >0.90.

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

15

Face Validity: Result

15

Patient/Encounter Level Testing

Yes

Type of Analysis

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

Sample Size

11,585

Statistic Name

Percent agreement

Statistical Results

0.92

Interpretation of results

See the Patient/Encounter Level Validity Testing Attachment.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Lower score is better

Mean performance score

0.31

Median performance score

0.34

Minimum performance score

0.16

Maximum performance score

0.43

Standard deviation of performance scores

0.07

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

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

Measure Steward Contact Information

Nate Mazonson
550 16th Box 0560
San Franciso, CA 94044
nate@alaracare.com
(650) 520-6649

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Rebecca Smith-Bindman
550 16th 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.edu
(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. 100% 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.

MUC2022-032 Geriatrics Surgical Measure

Program

Hospital Inpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

This programmatic measure assesses hospital commitment to improving surgical outcomes for patients greater than or equal to 65 years of age through patient-centered competencies aimed at achieving quality of care and safety for all older adult surgical patients. The measure will include 11 attestation-based questions across 7 domains representing a comprehensive framework required for optimal care of the older surgical patient. A hospital will receive a point for each domain where they attest to all items from at least one question (for a total of 7 points). Note that "patients" in all elements refers to surgical patients greater than or equal to 65 years of age at time of operation.

Numerator

This programmatic measure assesses hospital commitment to improving surgical outcomes for patients 65 years of age and older through patient-centered competencies aimed at achieving quality of care and safety for all older adult surgical patients. The measure will include 11 attestation-based question across 7 domains representing a comprehensive framework required for optimal care of the older surgical patient. A hospital will receive a point for each domain where they attest to all items from at least one question (for a total of 7 points). Note that "patients" in all elements refers to surgical patients 65 years of age and older at time of operation.

Domain 1: Identifying Goals of Care

Question 1: Advance Care Planning. Please attest that your hospital provides education to patients and providers regarding advance care planning and ensures that advance care planning preferences are captured, updated, and available for review in the medical record.

Question 2: Patient Goals. Please attest that your hospital provides education regarding goal concordant care and has established protocols for ensuring patient goals and decision making is documented in the medical record.

Domain 2: Medication Management

Question 3: Inappropriate Medications. Please attest that your hospital flags medications that may be inappropriate for older surgical patients and has established protocols for reviewing drug and non-drug alternatives to identified substances.

Question 4: Pain Management. Please attest that your hospital employs opioid sparing multimodal pain management strategies where possible and has protocols for capturing these regimens in the medical record.

Domain 3: Cognition and Delirium

Question 5: Delirium and Cognition Screening. Please attest that your hospital performs delirium and cognition screens and implements protocols for flagging high risk patients and implementing appropriate management plans for those with positive screens.

Domain 4: Function and Mobility

Question 6: Function and Mobility Screening. Please attest that your hospital performs pre-operative function and mobility screens and implements protocols to flagging high risk patients and implementing appropriate management plans for those with positive screens.

Domain 5. Social Determinants of Health

Question 7: Social Determinants of Health. Please attest that your hospital performs preoperative screens for psychosocial risk factors and establishes protocols for identifying at risk patients and employing appropriate management plans.

Domain 6: Care Transitions

Question 8: Identifying Needs at Hospital Discharge Please attest that your hospital elicits discussion between providers and patients regarding discharge care and establishes protocols to ensure that discharge summaries contain management plans for all identified post-discharge needs.

Question 9: Post-Acute Care. Please attest that your hospital has protocols for establishing two-way communication between providers and post-acute care facilities and tracks the quality of care at post-acute care facilities upon discharge.

Domain 7: Ensuring Quality Care for High Risk Patients

Question 10: Identification and Management of Seriously Ill Patients. Please attest that your hospital employs multidisciplinary evaluation of older patients and provides appropriate management, including the early utilization of palliative care consultations, for those with serious illness.

Question 11: Geriatric Surgery Leader and Quality Framework. Please attest that your hospital designates a geriatric surgery point person to oversee all aspects of this measure and establishes a framework for ongoing quality improvement regarding the care for patients.

Numerator Exclusions

N/A

Denominator

The denominator for each hospital is 7 which represents the total number of domains with at least one complete attestation.

The measure is calculated as the number of complete attestations / total number of domains. There is no partial credit for any question. Attestation of at least one element in all 7 domains is required to qualify for the measure numerator.

Denominator Exclusions

N/A

Denominator Exceptions

N/A

State of development

Field (Beta) Testing

State of Development Details

Testing was conducted during the pilot site visits.

What is the target population of the measure?

65 and older

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

Geriatric medicine

Measure Type

Other: Other

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

Programmatic Measure

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data;Electronic Clinical Data (non-EHR);Electronic Health Record;Paper Medical Records;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?

Facility

In which setting was this measure tested?

Community hospital;Hospital inpatient acute care facility;Veterans Health Administration facility

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

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)

N/A

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?

Other: Web-based tool in CMS quality reporting portal.

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

Feasibility assessments have been conducted; findings have demonstrated feasibility at the hospital level for all components. Attestation components are defined and feasible to address with reliability testing accomplished (see reliability). This has been accomplished via clinical registries, EHR, claims data, and ICD 9 coding. It is important to recognize that not all sites need to use the same method of data collection, given benchmarking is not being executed across sites but conducted longitudinally. Hence, individual sites need to remain consistent in their data source.

Method of Measure Calculation

Other (enter here):: Attestation using a web-based tool within the HQR system.

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

See "GSV Measure Evidence" attachment for performance gap information.

Unintended Consequences

Potential known unintended consequences in the geriatric population would be due to efforts around function/mobility. An anticipated increase in falls might occur as patients are encouraged to ambulate and/or management plans focus on efforts to mitigate deconditioning. If/when patients fall, ambulation efforts might be halted, which can then have the unintended consequences of deconditioning, restraint use, and/or pressure ulcers.

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

4

Outline the clinical guidelines supporting this measure

See "GSV Measure Evidence" attachment for clinical guidelines supporting this measure.

Name the guideline developer/entity

The American Geriatrics Society (AGS)

Publication year

2019

Full citation +/- URL

Boyd C, Smith CD, Masoudi FA, Blaum CS, Dodson JA, Green AR, Kelley A, Matlock D, Ouellet J, Rich MW, Schoenborn NL, Tinetti ME. Decision Making for Older Adults With Multiple Chronic Conditions: Executive Summary for the American Geriatrics Society Guiding Principles on the Care of Older Adults With Multimorbidity. J Am Geriatr Soc. 2019 Apr;67(4):665-673. doi: 10.1111/jgs.15809. Epub 2019 Mar 10. PMID: 30663782.

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

No

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

The recommended MCC Actions include: (1) identify and communicate patients' health priorities and health trajectory; (2) stop, start, or continue care based on health priorities, potential benefit vs harm and burden, and health trajectory; and (3) align decisions and care among patients, caregivers, and other clinicians with patients' health priorities and health trajectory.

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 recommended MCC Actions include: (1) identify and communicate patients' health priorities and health trajectory; (2) stop, start, or continue care based on health priorities, potential benefit vs harm and burden, and health trajectory; and (3) align decisions and care among patients, caregivers, and other clinicians with patients' health priorities and health trajectory.

Number of systematic reviews that inform this measure concept

15

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

See "GSV Measure Evidence" attachment.

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

See "GSV Measure 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

4,000,000

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?

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

Not conceptually or empirically indicated (enter here):: Not conceptually or empirically indicated (enter here): This is a programmatic measure for a facility to attest to specific standards for delivering high quality care to elderly patients. There are

Cost estimate completed

Yes

Cost estimate methods and results

No additional cost, can use existing resources

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

3

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

3

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.

70

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

70

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?

100

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

Other (enter here): Test-Retest

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

Kappa - Interrater Reliability (IRR)

Other: Sample size

240

Other: Statistical result

0.98

Other: Interpretation of results

From Pilot and current site evaluations, there is high Interrater reliability (IRR). IRR amongst the two raters had 99% agreement amongst 2 raters.

Empiric Validity

Yes

Empiric Validity: Statistic name

Modified Delphi Method (modified version of the RAND-UCLA Appropriateness Methodology)

Empiric Validity: Sample size

8

Empiric Validity: Statistical result

8

Empiric Validity: Methods and findings

1.) JAHF grant awarded to ACS to develop the Geriatric Surgery Verification and Quality Improvement Program (now known as the Geriatric Surgery Verification (GSV) Program). CQGS Core Development Team (CDT), Advisory Panel and multidisciplinary stakeholder groups convened to set the standards for geriatric surgical care. To achieve this objective and set the program standards, key benchmarks were established/achieved:

1.1.) Gather Literature and Develop Preliminary Standards: The CQGS CDT completed a comprehensive literature review in February 2016 and released the preliminary standards to stakeholders in April 2016.

1.2.) Conduct Preliminary Field Visits: Between December 2015 and April 2016, the CDT conducted 11 field visits in 7 cities across the country to measure the current state of surgery in older adults, evaluate scalability of the program, identify best practices already in place serving older adult surgery patients and gain perspectives/opinions from frontline staff. CDT reported on the field visit findings to CQGS stakeholders in May 2016.

1.3.) Refine and Vet Standards w/ Stakeholders:

1.3.1.) CQGS Preliminary Standards written in 2016 by the CDT, and based on ratings analysis, standards were repackaged from 308 standards into 92 Alpha Standards in 2017 [using a modified version of the RAND-UCLA Appropriateness Methodology, 44 of the 58 CQGS stakeholders performed two ratings on the validity and feasibility of the standards. Based on the ratings analysis, became 92 Alpha Standards and implemented into Alpha Pilot].

1.3.2.) The CQGS invited 15 hospitals to participate in a survey to determine which of the standards were already implemented, understand how easy or difficult the standards would be to implement if not already in practice, and identify and record standards that were confusing or difficult to interpret.

1.3.3.) From the results of Alpha Pilot, the 92 Alpha Standards were further condensed to 30 Beta Standards, categorized into 6 chapters: Program management, Goals and Decision-Making, Preoperative Optimization, Immediate Preoperative and Intraoperative Clinical Care, Postoperative Clinical Care, and Transitions of Care. The Beta Standards were released in November 2017.

1.3.4.) 8 hospitals participated in the Beta Pilot in December 2017 to determine the feasibility of implementation for each beta standard, collect "best practices" for standards implementation, learn barriers to implementation in hospitals of varying size, facility-type, and location, identify issues in the wording of the standards leading to misinterpretation, continue to record rationale for or against the standards, and resolve difficulties in the verification process. Beginning in June of 2018, site visits to each pilot hospital were performed and report was released end of 2018.

2.) Using the data from the Beta Pilot, the CQGS team finalized the standards and developed educational and supportive materials to aid in the official GSV Program launch in July 2019 with release of Optimal Resources for Geriatric Surgery 2019 Standards.

3.) GSV Program began accepting applications for enrollment in October 2019. To date: 53 hospitals enrolled in the program in various stages of implementation. 3 hospitals have completed verification process.

4.) In 2021, ACS conveyed a team of experts in geriatric surgical care to evaluate the GSV standards to determine feasibility as a measure that collectively improves care for older adults.

5.) In January 2022, 9 unique hospitals committed to providing optimal geriatric surgical care (as demonstrated through their established centers for geriatrics, participation in age-friendly health initiatives, or enrollment in the GSV Program, or any combination of these elements), were surveyed to evaluate implementation of key measures that collectively improve care for older adults. The key measure domains were determined through the combination of both the established ACS quality framework, and the processes, resources, and infrastructures necessary for the optimal care of the older adult surgical patient as determined by JAHF grant work.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

70

Face Validity: Result

44

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

6

Median performance score

5

Minimum performance score

0

Maximum performance score

7

Standard deviation of performance scores

0000

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 College of Surgeons

Measure Steward Contact Information

Clifford Ko
633 N Saint Clair St
Chicago, IL 60611
geriatricsurgery@facs.org
(312) 202-5518

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Michael Bencur
633 N Saint Clair St
Chicago, IL 60611
mbencur@facs.org
(312) 202-5120

Secondary Submitter Contact Information

Jill Sage
20 F Street, NW
Suite 1000
Washington, DC 20001
jsage@facs.org
(202) 672-1507

Submitter Comments

Michelle Schreiber, MD
Deputy Director for Quality and Value, Center for Clinical Standards and Quality
Centers for Medicare & Medicaid Services
200 Independence Ave. SW
Washington, DC 20201

RE: Programmatic Geriatric Surgery Hospital Measure

Dear Dr. Michelle Schreiber:

On behalf of the over 84,000 members of the American College of Surgeons (ACS), we appreciate the opportunity to propose a new type of measure--a programmatic measure--that values the full quality program needed to care for geriatric surgical patients.

The US population is rapidly aging and is one of the fastest-growing demographics in the country. Based on 2019 US Census data, the 65-and-older population grew by over a third since 2010, and by 2030 this population is estimated to grow to 72 million (20% of the total population). (1-3) Currently, over 4 million high risk operations (procedures with a mortality rate over 1%) are performed on adults over 65 annually. (3) Hospitals are increasingly faced with older patients who have complex medical, physiological, and psychosocial needs that are often inadequately addressed by the current healthcare infrastructure. Despite this growing need, our healthcare system has not comprehensively rethought care for the complex geriatric population since the creation of Medicare more than 50 years ago.

Part of what is needed in rethinking care for the older adult population is programmatic facility-level geriatric measurement. This solution is different from the current types of CMS measures, this programmatic measure incentivizes team-based care organized around the geriatric surgical patient to meet the challenges unique to geriatric surgical patients. Although existing quality metrics have improved both the rate and reporting of clinical outcomes (falls, appropriate use of anticoagulants, etc.) that are important to older individuals, these measures can be narrow in scope and may have limited long term effectiveness due to ceiling effects. Rather than simply addressing individual clinical issues in isolation, optimizing care for older patients with multifaceted vulnerability profiles will require a holistic approach with the goal of reframing the entire care pathway to better serve the needs of this unique population. The Geriatric Surgery Measure was developed with the Modified Delphi method, receiving input from more than 50 organizations, including the ACS. The multistakeholder group identified clinical frameworks based on evidence and best practices that provide goal-centered, clinically effective care for older patients. As a result, this programmatic measure consists of structural and process measures which address all 6 Institute of Medicine domains (safe, effective, patient-centered, timely, efficient, equitable), and is comprehensive across the full spectrum of geriatric surgical care.

This programmatic measure differs from the classical NQF structural measures or process measures which are usually singular structural components, or a simple process tied to a transaction/patient visit. The challenge with classical structure or processes measures is that care is not a single structural element or process. It is the collection of all these components orchestrated across the continuum of care for the entire team in a patient-centered manner. Together, these become a patient-centered program of care. When the components are properly tied together, care becomes well-coordinated, complex aspects of care are more reliably delivered, harms are minimized, and outcomes are optimized. The elements in the program are focused around care delivery, coordination, data, and data-driven improvement activities.

When CMS and NQF consider what we know as traditional structural and process measures, we agree that singular elements within one or another transaction as part of a series of transactions in health are "check-the-box" measures. These have limited impact on quality or improvement. However, within clinical domains of care such as geriatric surgical care, there are crucial structures and processes of care that reach across multiple transactions and link the care team's efforts together. The Geriatric Surgery Hospital Measure is based on key standards within the ACS Geriatric Surgical Verification Program (GSV) program, which follows the ACS Quality Model--the framework used across all ACS Quality programs,

including the Trauma Center Verification Program, the Commission on Cancer (CoC), and the Metabolic and Bariatric Surgery Verification program, and so on. Orchestrating all these elements result in better outcomes and improving their implementation would be an essential first step in surgical outcomes. (4-10)

Currently, CMS quality programs consist of a large, extremely costly universe of measures in multiple different payment programs. They often lack the consideration for focusing a surgical team in a patient-centered way. Such sporadic measurement creates a massive amount of burden and overhead but has a limited impact on improving the quality of overall care--this is especially evident in the MIPS program. (11-12) As a result, these efforts fail to create accountability to patients for the care they receive. Measuring a surgeon with sporadic metrics and disjointedly measuring anesthesia services, pathology, radiology and facility care with disparate measure sets does not create the alignment needed. The development of individual measures and the subsequent combination of these measures into payment incentive programs may be useful for fee-for-service payment. However, value-based payments need a more condition, patient-type approach. This proposed approach is programmatic within a clinical domain. Its implementation will create a team-based approach to optimizing the patient's chance at achieving their desired outcome. While this measure may appear long in its specification, it is a yearly attestation measure, and therefore the reporting burden is less burdensome on the surgical team and facility when compared to most CMS measures that require regular reporting of individual events included in the numerator.

This programmatic approach also offers useful information that patients will find beneficial when deciding where to seek care. An important standard to apply in evaluating payment quality incentives is their effectiveness in providing patients with knowledge of where to find high quality care in their community. Twenty years of NQF and payer actions in quality have not produced reliable public knowledge or a public-facing website that informs patients about where to get the care they need for the condition they have. Information on the comprehensiveness of a quality program, along with comparable information on the price of that care, are the prerequisites for a valid depiction of the value of care. In assessing the effectiveness of our measures, we wonder if the patient had this information, would it enable them to easily find information on a website for the types of care they seek, for a safety and equitability profile and for personal goal attainment.

This programmatic measure is the first step needed to build the foundation to care for the rapidly aging Medicare population. At this time, this effort does not meet all of the goals toward value-based care, but it is designed to be added to appropriate condition or procedure specific cost measures, which would help patients determine the affordability of the care they desire. Combining quality and price for care is a key step in establishing value.

Sincerely,

The ACS Geriatric Surgery Verification Quality Program and Health Policy Team

Clifford Ko, MD, MS, MSHS, FACS

Frank Opelka, MD, FACS

Jill Sage, MPH

MUC2022-055 Hybrid Hospital-Wide All-Cause Risk Standardized Readmission Measure

Program

Hospital Inpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Hospital-level, risk-standardized readmission rate (RSRR) of all-cause 30-day unplanned readmission after admission for any eligible condition within 30 days of hospital discharge. The measure, based on NQF #2879, uses enrollment data, inpatient claims, and electronic health record data. Hospitals receive a single summary RSRR, derived from the volume-weighted results of five specialty cohorts. Conditionally supported by the MAP pending NQF endorsement and currently in the IQR Program (voluntary reporting 7/1/2021, mandatory reporting beginning 7/1/2023). This MUC submission expands the cohort from Medicare fee-for-service (FFS) patients to include Medicare Advantage patients age 65 & older.

Numerator

The outcome for this measure is 30-day unplanned readmission. We define readmission as an inpatient admission for any cause, except for certain planned readmissions, within 30 days from the date of discharge from an eligible index admission. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days.

Numerator Exclusions

N/A

Denominator

The measure includes admissions for beneficiaries enrolled in Medicare FFS and/or Medicare Advantage for the 12 months prior to the date of index admission, on the date of the index admission, and the 30 days following discharge of the index admission; aged 65 or over; discharged alive from a non-federal short-term acute care hospital; and not transferred to another acute care facility.

Denominator Exclusions

The measure excludes index admissions for patients:

1. Admitted to a Prospective Payment System (PPS)-exempt cancer hospital;

2. Without at least 30 days post-discharge enrollment in FFS Medicare or Medicare Advantage;

3. Discharged against medical advice (AMA);

4. Admitted for primary psychiatric diagnoses;

5. Admitted for rehabilitation; or

6. Admitted for medical treatment of cancer.

Note: Patients who do not have a full 30 days of post-discharge enrollment in Medicare FFS or Medicare Advantage due to death are eligible for inclusion in the cohort. Thus, if a patient had an unplanned readmission and later died, all within 30 days of discharge from the index admission, the case would be captured in the outcome, assuming they met all inclusion/exclusion criteria.

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, Medicare Advantage

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

Geriatric medicine

Measure Type

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?

Administrative Data (non-claims);Electronic Health Record

If applicable, specify the data source

N/A

Description of parts related to these sources

Medicare Inpatient Claims: The index dataset contains administrative inpatient hospitalization data for Medicare FFS and/or MA beneficiaries. Hospital inpatient claims are also used to characterize comorbidities as documented during the index admission and in the year before the index admission to capture a comprehensive view of patients medical histories. Readmissions are identified by subsequent hospital inpatient claims for short-term acute care and critical access hospitals. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. It was also used to determine

hospice enrollment. EHR: The measure utilizes 13 lab and vital signs extracted from the EHR for risk adjustment, which are routinely captured during the course of normal care.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Hospital inpatient acute care facility

Multiple Scores

No

What one healthcare domain applies to this measure?

Seamless Care Coordination

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

N/A

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

Yes

CMIT ID

05746

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Endorsed

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

2879

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

No

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

Denominator

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

The only change to the hybrid measure is the addition of Medicare Advantage admissions to the cohort that previously included only FFS admissions. Some nominal tweaks to description and target population will also have to be considered in order to ncorporate MA beneficiaries more accurately in the measure descriptions. These changes will allow the hybrid measures to capture the target population of all Medicare beneficiaries.

If endorsed: Year of most recent CDP endorsement

2020

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

2023

Digital Measure Information

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

Yes

If eCQM, enter Measure Authoring Tool (MAT) number

Although this is not an eCQM, we utilize the MAT to specify the EHR-portion of the specifications for this Hybrid measure. CMS 529

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?

No

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?

2014

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

x-3701

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

Hospital, 2015

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

2015 HIQR

What was the MAP recommendation in this year?

2015; CMS Programs: HIQR MAP recommendation: Encourage continued development

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

MAP recommendation: Encourage continued development While the claims version of this measure is NQF-endorsed (1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)) is already a part

of the Hospital Inpatient Quality Reporting, the e-Measure version of this measure is in alpha testing. The MAP supported continued development of this measure, noting the potential to improve the measure risk adjustment model by including available clinical data. Further, members highlighted the need to review the conceptual and empirical relationship between SDS factors and this outcome and include such variables if appropriate. Further, MAP noted that CMS should review the empirical and conceptual relationship between SDS factors and hospital-wide readmissions and seek endorsement on this version of the measure by the relevant NQF standing committee.

MAP report page number being referenced for this year:

MAP 2015 Considerations for Selection of Measures for Federal Programs: Hospitals Page:9

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

Measure currently used in a CMS program, but the measure is undergoing substantial change

Range of years this measure has been used by CMS Programs

Hospital Inpatient Quality Reporting Program, voluntary reporting beginning July 1, 2021 through June 30, 2022 and mandatory reporting beginning July 1, 2023 through June 30, 2024, impacting the FY 2026 payment.

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program

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?

Claims-only hospital-wide readmission measure (HWR)

- Hospital-Level 30-Day Risk-Standardized Readmission Rate following Acute Myocardial Infarction (NQF 0505)
- Hospital-Level 30-Day Risk-Standardized Readmission Rate following Chronic Obstructive Pulmonary Disease (NQF 1891)
- Hospital-Level 30-Day Risk-Standardized Readmission Rate following Heart Failure (NQF 0330)
- Hospital-Level 30-Day Risk-Standardized Readmission Rate following Pneumonia (NQF 0506)
- Hospital-Level 30-Day Risk-Standardized Readmission Rate Following Coronary Artery Bypass Graft Surgery (NQF 2515)
- Hospital-Level 30-Day Risk-Standardized Readmission Rate Following Elective Primary Total Hip Arthroplasty and/or Total Knee Arthroplasty (NQF 1551)

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

This Hybrid measure utilizes EHR clinical data for risk adjustment in addition to claims data. The Hybrid measure is identical to the claims-only HWR measure, except for the addition of the important clinical data added for use in risk adjustment. The cohort expansion of the Hybrid HWR measure to include MA

admissions is the only material change to the Hybrid HWR measure (currently in voluntary reporting) being considered

How will this measure add value to the CMS program?

Specifically, this Hybrid HWR measure will complement the existing CMS Hospital-Wide All-Cause Risk-Standardized Mortality Measure to allow assessment of trends in hospital performance for both outcomes, similar to other complementary pairs of readmission and mortality measures for specific conditions and procedures. Further, the Hybrid HWR measure will provide annually updated performance estimates for a larger proportion of the nations hospitals, allowing significant performance outliers to be identified. We would also like to note that the hospital-wide readmission measures cannot be adopted into the HRRP program as measures in HRRP must be condition- or procedure-specific. The existing Hybrid HWR measure is scheduled to replace the currently implemented claims-only HWR measure in 2025 reporting, with the addition of Medicare Advantage patients pending pre-rulemaking and rulemaking.

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;Claims

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

This measure uses data from claims and the EHR.

We tested the feasibility and validity of electronic extraction of these critical data elements as part of a more comprehensive evaluation of a larger set of core clinical data elements (CCDEs). The CCDE are a set of 21 EHR data elements that are captured on most adults admitted to acute care hospitals, are easily extracted from EHRs, and can be used to risk adjust hospital outcome measures for a variety of conditions and procedures. All of the critical data elements used in the Hybrid HWR measure are included in the CCDE. Testing of the CCDE involved three phases: 1) identification of potentially feasible clinical data through qualitative assessment, 2) empirical feasibility testing of several clinical data elements electronically extracted from two large multi-facility health systems, and 3) validity testing of the CCDE at an additional health system. Data capture criteria for defining the 21 CCDE included: obtained consistently under current practice; captured in a standard definition; entered in a structured field; encoded consistently; extractable from the EHR; and exported with metadata. Scores from the data element feasibility scorecard were 3 for Data Availability; 3 for Data Accuracy; 3 for Data Standards and 3 for Workflow, with 3 indicating the highest level of feasibility. This feasibility testing was

conducted in hospitals with both Medicare FFS and MA patients. See attached Data Element Feasibility Scorecard for additional detail.

Administrative claims data used in this measure are routinely captured as part of the billing process and there are no fees associated with collecting the data.

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Claims;Hybrid: eCQM

Evidence of Performance Gap

Studies have estimated the rate of preventable readmissions to be as low as 12% and as high as 76%. Given that studies have shown readmissions to be related to quality of care, and that interventions have been able to reduce 30-day readmission rates, it is reasonable to consider an all-condition readmission rate as a quality measure.

Please see the supplemental file and evidence attachment for additional details.

Unintended Consequences

We have not identified any unintended consequences during measure development, testing and use. We are committed to monitoring this measures use and assessing potential unintended consequences over time, such as the inappropriate shifting of care or coding/billing practices, increased patient morbidity and mortality, and other negative unintended consequences for patients.

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 and publicly available reports (e.g., from agencies);Internal data analysis

Summarize the empirical data

Evidence that hospitals have been able to reduce readmission rates through quality-of-care initiatives illustrates the degree to which hospital practices can affect readmission rates. Successful randomized trials have reduced 30-day readmission rates by 20-40%. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions, applying lessons learned from clinical trials. Several have been notably successful in reducing readmissions. The strongest evidence supporting the efficacy of improved discharge processes and enhanced care at transitions is a randomized controlled trial by the Project RED (Re-Engineered Discharge) intervention, in which a nurse was assigned to each patient as a discharge advocate, responsible for patient education, follow-up, medication reconciliation, and preparing individualized discharge instructions sent to the patients

primary care provider and there was a follow-up phone call from a pharmacist within four days of discharge, which demonstrated a 30% reduction in 30-day readmissions. Please see evidence attachment for additional details and citations.

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

11,029,470

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

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment

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 testing the impact of cohort expansion to include MA admissions, we used a one-year dataset containing combined FFS + MA inpatient claims for patients discharged between July 1, 2018, through June 30, 2019, from acute care hospitals.

Discrimination Statistics: The range of c-statistics from 0.60 to 0.69 showed good discrimination across the specialty cohort models. Please see the supplemental file: Medicare Advantage: Evaluation of the Claims-Based Hospital-Wide Readmission Measure with the addition of Medicare Advantage to the Fee-for service FFS cohort for additional details.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

We expect the burden associated with reporting of the Hybrid HWR measure to 10 minutes per measure, per quarter. Therefore, using the estimate of 10 minutes per measure per quarter (10 minutes, one measure, four quarters = 40 minutes), we estimate a burden increase of 40 minutes (0.67 hours) per hospital per year. We estimate that the impact of this proposed change is a total collection of information burden increase of 2,211 hours and a total cost increase of approximately $83,266 for all participating IPPS hospitals (3,300) annually. We do not anticipate any increases in hospital burden as a result of increasing the cohort size to include MA admissions

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?

Yes

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

24

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

Median signal-to-noise reliability

Signal-to-Noise: Sample size

4563

Signal-to-Noise: Statistical result

.700

Signal-to-Noise: Interpretation of results

In a claims-only dataset, median signal-to-noise reliability for the combined FFS and MA cohort was higher than for the FFS only cohort among each of the specialty cohorts. Median reliability for the total cohort was 0.700 for combined FFS and MA admissions, as compared to 0.602 for FFS only. We have not yet tested CCDE elements in the combined MA+FFS cohort, but we do expect the pattern of increased reliability for the combined FFS and MA cohort will be similar for the hybrid measure. The signal-to-noise analyses to compare the measure-score reliability of the HWR measure with combined FFS and MA patients indicates strong agreement. For the hospital event rate based on the patient binomial outcomes like readmission (Yes/No), a value of 0-0.2 indicates poor agreement; 0.3-0.4 indicates fair agreement; 0.5-0.6 indicates moderate agreement; 0.7-0.8 indicates strong agreement; and >0.8 indicates almost perfect agreement. The value of 0.70 is strong (Landis & Koch, 1977).

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, measuring the strength of relationship between two measure scores.

Empiric Validity: Sample size

4767

Empiric Validity: Statistical result

-0.656

Empiric Validity: Methods and findings

We have not tested empirical validity in the combined Medicare FFS + MA population. However, correlation analysis was performed between the claims-only HWR measure and three external measures of hospital quality, assessed on a national public reporting sample of almost 7 million patients and 4,767 hospitals. The hybrid HWR measure uses the same concept, cohort, outcome, and claims-only risk adjustment variables as the claims-only measure. The only difference between the hybrid HWR and claims-only HWR measures is that CCDE data are added as risk adjustment variables in addition to the claims-only risk adjustment variables for the hybrid measure. There is no conceptual reason to believe that the results from the claims-only measure would be significantly dissimilar to results from the hybrid measure. Empiric validity testing was completed by examining the relationship of performance the claims-only HWR measure scores (RSRRs) with each of these external measures of hospital quality as measured by Pearson correlation coefficients: Hospital Star Rating readmission group score, Overall Hospital Star Rating summary score, and HCAHPS. There is no reason to expect that the addition of Medicare Advantage beneficiaries to the cohort would impact these validity results. As expected, the claims-only HWR measure score was moderately, negatively correlated with both the Star Rating Standardized Readmission Group Score (-0.656), and the Summary Score (-0.486), meaning that higher scores (better performance) on the comparator measures was associated with lower scores (better performance) on the HWR measure. This is expected because the star ratings quality measures focus on, or contain a portion of, the same domain of quality as the HWR measure (readmission). The HCAHPS measures related to transitions of care, communication about medications, doctor and nurse communication, and discharge instructions, were also correlated with HWR in the expected direction. For example, the HCAHPS discharge information linear mean score had a Pearson correlation coefficient of -0.305, indicating that better performance on the discharge measure was correlated with lower measure scores (better performance) on the HWR measure. Likewise, the HCAHPS score for No, staff did not give patients information about help after discharge was positively correlated with HWR, meaning that worse performance on the discharge measure was associated with worse performance on HWR or higher rates of readmission. The results showed the expected correlation between the claims-only HWR measure score and the three external measures of quality, which provides external support for measure score validity. See Testing Attachment for additional detail.

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: Proportion and Ratio

Measure Performance Score Interpretation

Lower score is better

Mean performance score

15.65

Median performance score

15.61

Minimum performance score

10.37

Maximum performance score

47.22

Standard deviation of performance scores

1.43

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

Jennifer Robinson
7500 Security Blvd
Baltimore, MD 21244
jennifer.robinson@cms.hhs.gov
(443) 729-6368

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

N/A

MUC2022-057 Hybrid Hospital-Wide All-Cause Risk Standardized Mortality Measure

Program

Hospital Inpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Hospital-level, risk-standardized 30-day all-cause mortality rate (RSMR) for Medicare fee-for-service (FFS) and Medicare Advantage (MA) patients (65 to 94). The measure, based on NQF #3502, uses enrollment data, inpatient claims, and electronic health data to identify 30-day all-cause mortality outcome, and adjust for comorbidities based on the ICD-10 diagnosis/procedure codes and clinical risk factors from electronic health data for the measure score calculation. This measure, previously conditionally supported for use in IQR and planned for use by CMS for voluntary reporting in IQR, is being expanded to include Medicare Advantage patients in addition to FFS patients in the cohort.

Numerator

The outcome for this measure is 30-day all-cause mortality. Mortality is defined as death for any reason within 30 days after the index admission date, including in-hospital deaths.

Numerator Exclusions

N/A

Denominator

The cohort includes inpatient admissions for patients aged 65-94 years old, for patients enrolled in Medicare (Fee-for-Service [FFS] and/or Medicare Advantage [MA]) for the 12 months prior to the date of admission and during the index admission. If a patient has more than one admission in the year, one hospitalization is randomly selected for inclusion in the measure. Cohort includes index admissions for patients:

- Who have not been transferred from another inpatient facility
- Admitted for acute care (does not include principal discharge diagnosis of psychiatric disease, or rehabilitation care
- Not enrolled in hospice within 12 months prior to the index admission
- Without a principal diagnosis of cancer and also enrolled in hospice during their index admission
- Without any diagnosis of metastatic cancer
- Not enrolled in hospice within two days of admission
- Without a principal discharge diagnosis, or a secondary diagnosis that is present on admission (POA) for a condition for which hospitals have limited ability to influence survival

Denominator Exclusions

The measure excludes admissions for patients:

- With inconsistent or unknown vital status

- Discharged against medical advice
- With admissions for crush injury, burn, intracranial injury, spinal cord injury, skull and face fractures, or open wounds of head, neck, and trunk
- With an admission in a low volume CCS, defined as less than or equal to 100 patients with that principal discharge diagnosis per service-line division across all hospitals.

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, Medicare Advantage

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

Geriatric medicine

Measure Type

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?

Administrative Data (non-claims);Electronic Health Record

If applicable, specify the data source

N/A

Description of parts related to these sources

Medicare Inpatient Claims: The index dataset contains administrative inpatient hospitalization data for Medicare FFS and/or MA beneficiaries. Hospital inpatient claims are also used to characterize comorbidities as documented during the index admission and in the year before the index admission to capture a comprehensive view of patients medical histories. Readmissions are identified by subsequent hospital inpatient claims for short-term acute care and critical access hospitals. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. The mortality outcome included in-hospital deaths as well as all-cause mortality within 30 days of the index admission based on the National Death Index (NDI). It was also used to determine hospice enrollment. EHR: The measure

utilizes 10 lab and vital signs extracted from the EHR for risk adjustment, which are routinely captured during the course of normal care.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

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

N/A

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

Yes

CMIT ID

06031

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Endorsed

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

3502

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

No

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

Denominator

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

The only change to the hybrid measure calculation is the addition of Medicare Advantage admissions to the cohort that previously included only FFS admissions.

If endorsed: Year of most recent CDP endorsement

2019

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

2023

Digital Measure Information

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

Yes

If eCQM, enter Measure Authoring Tool (MAT) number

Although this is not an eCQM, we utilize the MAT to specify the EHR-portion of the specifications for this Hybrid measure. CMS 844

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?

No

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?

2017

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

MUC17-196

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

Hospital, 2018

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

2017; MUC17-196. CMS Program: HIQR

What was the MAP recommendation in this year?
2017; Measure ID: MUC17-196. CMS Program: HIQR MAP Recommendation: Conditionally Support

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

N/A

MAP report page number being referenced for this year:

MAP 2018 Considerations for Implementing Measures in Federal Programs: Hospitals, Page numbers Page numbers 7,8,9.

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

Measure currently used in a CMS program, but the measure is undergoing substantial change

Range of years this measure has been used by CMS Programs

Hospital Inpatient Quality Reporting Period, voluntary reporting began with July 2022-June 2023 performance data for 2024 confidential reporting; mandatory reporting will begin with July 2023-June 2024 performance data for 2025 public reporting (FY 2026 payment year).

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program

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?

Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate Following Pneumonia Hospitalization; Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate Following Acute Ischemic Stroke; Hospital 30-day, All-Cause, Risk-Standardized Mortality Rate Following Heart Failure (HF) Hospitalization; Hospital 30-day, All-Cause, Risk-Standardized Mortality Rate Following Acute Myocardial Infarction (AMI) Hospitalization; Hospital 30-day, All-Cause, Risk-Standardized Mortality Rate Following Coronary Artery Bypass Graft (CABG) Surgery; Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization; CMS Death Rate among Surgical Inpatients with Serious Treatable Complications.

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

This measure is a hospital-wide measure, as opposed to a condition- or procedure-specific mortality measure.

How will this measure add value to the CMS program?

Death is a finite event, easy to measure accurately, and easily understood by patients and providers. For the majority of Medicare beneficiaries admitted to acute care hospitals in the US, the goal is to avoid short-term mortality. By measuring Hospital-Wide Mortality (HWM), CMS can ensure that efforts to reduce other outcomes, such as readmissions and resource utilization, are not resulting in unintended consequences. Specifically, this HWM measure will complement the existing CMS Hospital-Wide All-Cause Risk-Standardized Readmission Measure (NQF #1789) to allow assessment of trends in hospital performance for both outcomes, similar to other complementary pairs of readmission and mortality measures for specific conditions and procedures. Further, the HWM measure will provide CMS with annually updated performance estimates for a larger proportion of the nation's hospitals, allowing significant performance outliers to be identified.

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;Claims

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

This measure uses beneficiary enrollment data, claims data, and the EHR data.

We tested the feasibility and validity of electronic extraction of these critical data elements as part of a more comprehensive evaluation of a larger set of core clinical data elements (CCDEs). The CCDE are a set of 21 EHR data elements that are captured on most adults admitted to acute care hospitals, are easily extracted from EHRs, and can be used to risk adjust hospital outcome measures for a variety of conditions and procedures. All of the critical data elements used in the measure are included in the CCDE. Testing of the CCDE involved three phases: 1) identification of potentially feasible clinical data through qualitative assessment, 2) empirical feasibility testing of several clinical data elements electronically extracted from two large multi-facility health systems, and 3) validity testing of the CCDE at an additional health system. Data capture criteria for defining the 21 CCDE included: obtained consistently under current practice; captured in a standard definition; entered in a structured field; encoded consistently; extractable from the EHR; and exported with metadata. Scores from the data element feasibility scorecard were 3 for Data Availability; 3 for Data Accuracy; 3 for Data Standards and 3 for Workflow, with 3 indicating the highest level of feasibility. This feasibility testing was conducted in hospitals with both Medicare FFS and MA patients. See attached Data Element Feasibility Scorecard for additional detail.

Administrative claims data used in this measure are routinely captured as part of the billing process and there are no fees associated with collecting the data.

Method of Measure Calculation

Hybrid

Hybrid measure: Methods of measure calculation

Hybrid: Claims;Hybrid: eCQM

Evidence of Performance Gap

The variation in performance between the lowest-performing hospitals (RSMR of 3.95%) and the highest performing hospitals (RSMR of 8.7%) in the claims-only data, which we estimate will be similar to the Hybrid results on a national sample, shows there is a clear quality gap.

In terms of performance compared to the median (6.93%), some hospitals can achieve substantially lower overall risk-standardized mortality rates than the average-performing hospital, while other hospitals are performing substantially worse than an average performer.

Specifically, the best performing hospital (RSMR of 3.95%) is performing 43% better than an average performer (or has about 30 fewer deaths per 1000 patients compared to the average performer), while the worst performing hospital (8.70%) is performing 25% worse than an average performer (or has 18 more deaths per 1000 patients). Note that the average performer refers to hospital with the same case and service-line mix, performing at the average (median).

Unintended Consequences

We have not identified any unintended consequences during measure development, testing and use. We are committed to monitoring this measures use and assessing potential unintended consequences over time, such as the inappropriate shifting of care or coding/billing practices, increased patient morbidity and mortality, and other negative unintended consequences for patients.

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 and publicly available reports (e.g., from agencies);Internal data analysis

Summarize the empirical data

Mortality is an unwanted outcome for the overwhelming majority of patients admitted to US hospitals. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to, complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients conditions at the time of hospital admission and then evaluate patient outcomes. This mortality measure identifies institutions whose performance is better or worse than would be expected based on their patient case-mix, and therefore can both promote hospital quality improvement, and better inform consumers about care quality. According to internal analyses, from July 2016 to June 2017, there were about 10 million inpatient admissions among Medicare Fee-for-Service (FFS) beneficiaries between the age of 65 and 94, at 4,700 US hospitals. The observed 30-day mortality rate was 8.17%. This is especially relevant as, while the current condition- and procedure-specific mortality measures address the most common and morbid healthcare conditions as identified by MedPAC (1) in the most recent three-year public reporting period, together they captured only 4.8 million Medicare FFS beneficiary admissions; a HWM measure is likely to capture about 6.5 million admissions across 4,700 hospitals. Using acute myocardial infarction as an example, which has seen the greatest declines in mortality, the median hospital risk-standardized mortality rate (RSMR) following admission for acute myocardial infarction has declined from 16.4% in 2006 to 13.1% in 2016 (July 2015-July 2016 data). (2,3) If development and reporting of this HWM measures produces even a tenth as much impact, this would translate into nearly 14,000 deaths averted in a one-year period. Furthermore, if all hospitals performed as well as hospitals in the 10th percentile for RSMR, about 100,000 deaths

would be averted, compared to if all hospitals were performing at the median. For some conditions and diagnoses, evidence supports that optimal medical care reduces mortality. (4,5) We know from ongoing improvements in condition- and procedure-specific mortality rates that interventions to improve these outcomes are feasible. (2) Multiple organizations, including the Institute for Healthcare Improvement (IHI), promote a range of evidence-based strategies to reduce hospital mortality. (6) These strategies include: 1. Adoption of strategies shown to reduce ventilator-associated pneumonia (7-9) 2. Delivery of reliable, evidence-based care for acute myocardial infarction (10,11) 3. Prevention of adverse drug events though medication reconciliation (12) 4. Prevention of central line infections through evidence-based guideline-concordant care (13) 5. Prevention of surgical site infections through evidence-based guideline-concordant care (14,15) To reduce mortality, the IHI further encourages hospitals to use multidisciplinary rounds to improve communication, employ Rapid Response Teams to attend to patients at the first sign of clinical decline, identify high-risk patients on admission and increase nursing care and physician contact accordingly, standardize patient handoffs to avoid miscommunication or gaps in care, and establish partnerships with community providers to promote evidenced-based practices to reduce hospitalizations before patients become critically ill. (16) The IHIs 100,000 Lives Campaign, which was created to enlist hospitals in a coordinated effort to adopt the above interventions, led to an estimated more than 120,000 lives saved over the first 18 months of the campaign. (17)

References:

1. MedPAC. March 2011 Report to the Congress: Medicare Payment Policy. 2011; https://www.medpac.gov/document/march-2011-report-to-the-congress/ Accessed January 20, 2016

2. Medicare Hospital Quality Chartbook 2010 Performance Report on Outcomes Measures for Acute Myocardial Infarction, Heart Failure, and Pneumonia September 29, 2010. Prepared by Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation. Medicare Hospital Quality Chartbook 2010: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/HospitalChartBook.pdf Accessed February 27, 2019.

3. Trends in mortality rates following admission for acute myocardial infarction, chronic obstructive pulmonary disease, heart failure, pneumonia, and acute ischemic stroke. Prepared by Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation. Medicare Hospital Quality Chartbook 2017: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/2017-Chartbook.zip, Accessed February 27, 2019.

4. To Err Is Human: Building a Better Health System, Institute of Medicine (IOM), National Academy Press, Washington, DC (1999)

5. Classen DC, Resar R, Griffin F, et al. Global trigger tool shows that adverse events in hospitals may be ten times greater than previously measured. Health affairs. 2011;30(4):581-589

6. Berwick DM, Calkins DR, McCannon CJ, Hackbarth AD. The 100,000 lives campaign: Setting a goal and a deadline for improving health care quality. JAMA. 2006;295(3):324-327.

7. Tablan O, Anderson L, Besser R, Bridges C, Hajjeh R. CDC; Healthcare Infection Control Practices Advisory Committee. Guidelines for preventing health-care-associated pneumonia, 2003:

Recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee. MMWR Recommendation Reports. 2004;53(RR-3):1-36.

8. American Thoracic Society, Infectious Diseases Society of America. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. American Journal of Respiratory Critical Care Medicine. 2005;171:388-416.

9. Resar R, Pronovost P, Haraden C, Simmonds T, Rainey T, Nolan T. Using a bundle approach to improve ventilator care processes and reduce ventilator-associated pneumonia. Joint Commission Journal on Quality and Patient Safety. 2005;31(5):243-248.

10. Antman EM, Anbe DT, Armstrong PW, et al. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction; A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (committee to revise the 1999 guidelines for the management of patients with acute myocardial infarction). Journal of the American College of Cardiology. 2004;44(3):E1-e211.

11. Centers for Medicare & Medicaid Services. Hospital Quality Initiative Overview. 2005; http://www.allhealth.org/briefingmaterials/HospitalQualityInitiativeOverview-CMS-512.pdf. Accessed January 20, 2016.

12. Joint Commission. 2005 Joint Commission National Patient Safety Goals: Practical Strategies and Helpful Solutions for Meeting These Goals. 2005

13. O'Grady NP, Alexander M, Dellinger EP, et al. Guidelines for the prevention of intravascular catheter related infections. Clinical Infectious Diseases. 2002;35(11):1281-1307.

14. Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR, Committee HICPA. Guideline for prevention of surgical site infection, 1999. American Journal of Infection Control. 1999;27(2):97-134.

15. The Joint Commission. Surgical Care Improvement Project. 2005; http://www.jointcommission.org/surgical_care_improvement_project/. Accessed January 20, 2016.

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

11029470

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

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment

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

Please see the testing form attachment for further details. The range of c-statistic results is 0.8228 to 0.9587 across 15 divisions which is consistent with or better than results we have seen for other 30-day mortality measures.

Rationale for not using risk adjustment

N/A

Cost estimate completed

Yes

Cost estimate methods and results

We expect the burden associated with reporting of the Hybrid HWM measure to be similar to our estimates for reporting the Hybrid HWR measure, that is, 10 minutes per measure, per quarter. Therefore, using the estimate of 10 minutes per measure per quarter. Therefore, using the estimate of 10 minutes per measure per quarter (10 minutes one measure four quarters = 40 minutes), we estimate a burden increase of 40 minutes (0.67hours) per hospital per year. Beginning with the 2022 through 2023 reporting period, we estimate an annual burden increase of 2,200 hours across participating IPPS hospitals (0.67 hours 3,300 IPPS hospitals). We estimate this to represent a cost increase of $90,200

across IPPS hospitals ($41 2,200 hours). We do not anticipate any increases in hospital burden as a result of increasing the cohort size.

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.

4

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

3

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?

21

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

Adjusted intraclass correlation coefficient

Random Split-Half Correlation: Sample size

84825

Random Split-Half Correlation: Statistical result

.7748

Random Split-Half Correlation: Interpretation of results

As a metric of agreement, we calculated the ICC [1,2]. To calculate the ICC, we used 84,825 admissions across 21 hospitals, using 15 months of data. The agreement between the two independent assessments of the RSMR for each hospital was 0.6826, and the adjusted ICC (which estimates the ICC if we had been able to use one full year of data in each split sample) [3,4], is 0.7748. This adjusted ICC is considered substantial reliability, according to conventional standards [1]. To note, the datasets used for

development and testing of the original Hybrid HWM measure and for these analyses were from the Kaiser Permanente Northern California (KPNC) health system. The dataset includes matched administrative claims and electronic health record (EHR) data for index acute care admission. The KPNC dataset includes both Medicare Advantage and Medicare FFS patients. 1. Landis J, Koch G. The measurement of observer agreement for categorical data, Biometrics 1977;33:159-174. 2. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428. 3. Brown W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296 322. 4. Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271 295.

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

Qualitative trend analysis - Qualitative assessment of an association between this measure and 3 other external assessments.

Empiric Validity: Sample size

4581

Empiric Validity: Statistical result

00000

Empiric Validity: Methods and findings

The external empiric validity was not directly tested in the Hybrid HWM measure due to lack of availability of EHR data from a nationally representative set of hospitals. Instead, we report results of testing done in the claims-only HWM measure. Because of the homology between the two measures, there is no reason to suspect that the results of analyses done for the claims-only measure would differ in any significant way from results of analyses for a nationally representative hybrid measure. Based on our discussions with our Technical Workgroup and with other experts, we concluded that there is no single analysis that is sufficient to validate the measure because there is no gold standard exists for the validation of a hospital-wide quality measure. With this limitation in mind, we used the three empiric external analyses to demonstrate a trend of validity using different metrics: 1. Nurse to bed ratio. 2. Hospital Star Rating mortality group score. 3. Overall Hospital Star Rating. Our approach was to select three separate assessments against which we could compare the measure score with the hypothesis

that a trend toward correlation with these external assessments would support a conclusion of high measure score validity. For each external measure of quality, the comparison showed a trend toward better performance on the HWM measure with better performance on the comparator measure, as expected. For example, when comparing the claims-only HWM measure to the nurse to bed ratio, as the number of nurses per bed increases (more nurses in the hospital) across quartiles of nurse-to-bed ratio, the median overall HWM mortality rate is lower (better). Likewise, better performance on the HWM measure is associated with better Star Rating mortality group scores across quartiles of mortality group score performance. Finally, HWM performance improves across the Star Rating category in the expected direction: HWM scores are better (lower) as the Star Rating category improves (increases from 1, to 5 Stars).

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

6

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: Proportion and Ratio

Measure Performance Score Interpretation

Lower score is better

Mean performance score

4.61

Median performance score

4.61

Minimum performance score

3.98

Maximum performance score

5.43

Standard deviation of performance scores

0.33

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

Jennifer Robinson
7500 Security Blvd
Baltimore, MD 21244
jennifer.robinson@cms.hhs.gov
(443) 729-6368

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

This measure includes most acute care specialties with the exception of psychiatry.

MUC2022-058 Hospital Disparity Index (HDI)

Program

Hospital Inpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The HDI is a prototype method for a single score that summarizes several measurements of disparity in care at a hospital. This score will summarize existing results of the Centers for Medicare and Medicaid Services (CMS) Disparity Methods (stratified measure results) across a range of measures and social and demographic risk factors, to provide more accessible information about variations in healthcare disparity across hospitals.

Numerator

The HDI is a composite score and does not have a typical numerator. We are using this field to describe those hospitals that will obtain a score. The HDI includes hospitals that have patient populations which allow for calculation of both Within and Across Disparity Method for dual enrollment in Medicare and Medicaid (DE) and Within and Across Disparity Method results for at least one race and ethnicity group (Black, Hispanic, and Asian/Pacific Islander patients). This is operationalized as including at least one patient with the risk factor and one without.

Numerator Exclusions

Hospitals without at least one patient with the risk factor and one patient without the risk factor will not be eligible for disparity evaluation because we cannot examine disparities.

Denominator

The HDI does not have a traditional numerator and denominator. We use this field to define currently included measures for which Within and Across Disparity Method results are calculated and combined for an overall HDI score. Currently CMS reports Within and Across Disparity Methods for dual enrollment in Medicare and Medicaid (DE) for seven measures. All are readmission measures, including six condition specific measures and one hospital wide readmission measure:

Hospital 30- Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Acute Myocardial Infarction (AMI) Hospitalization, AMI Readmission measure, NQF ID# 0505, CMI ID# 80;

Hospital 30- Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Coronary Artery Bypass Graft (CABG) Surgery, CABG Readmission measure, NQF ID# 2515, CMIT ID# 1426;

Hospital 30- Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization, COPD Readmission measure, NQF ID# 1891, CMIT ID# 1455;

Hospital 30-Day, All-Cause, Risk- Standardized Readmission Rate (RSRR) Following Heart Failure (HF) Hospitalization, HF Readmission measure, NQF ID# 0330, CMIT ID# 78;

Hospital 30-Day, All-Cause, Risk-Standardization Readmission Rate (RSRR) Following Pneumonia (PN) Hospitalization, PN Readmission Measure, NQF ID# 0506, CMIT ID# 83;

Hospital- Level 30-Day, All-Cause, Risk- Standardized Readmission Rate (RSRR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA), THA/TKA Readmission measure, NQF ID# 1551, CMIT ID# 899; and

Hospital-Wide 30-Day, All-Cause, Risk-Standardized Readmission Rate Following Hospitalization, HWR Measure, NQF ID# 1789, CMIT ID# 2710. For the development and testing of the HDI we used the results of these measures for Reporting Year (RY) 2022.

Denominator Exclusions

N/A

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Hospitals serving Medicare Fee for Service patients.

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

Other: No Specialty

Measure Type

Outcome

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?

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?

Facility

In which setting was this measure tested?

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

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

N/A

Section 2: Measure Evidence

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

Other: The HDI is being proposed as a prototype for evaluation; CMS plans to use this Index in IQR to promote high quality care to beneficiaries.

Stratification

No

Feasibility of Data Elements

ALL data elements are in defined fields in electronic sources

Feasibility Assessment

As described above, the HDI is a prototype method for a single score that summarizes several measurements of disparity in care at a hospital that will already be reported. Select measures will be reported stratified by dual enrollment status confidentially beginning in Fall 2022. Codes included in measure stratification, as well as in individual measures in the HDI methodology, are the same as they appear in the original measure, and which have been reviewed by the MUC and recommended by the MAP previously.

The variables identified for stratification as part of the HDI method are available through claims and public sources. Dual enrollment status is available through claims data, race and ethnicity is imputed via the Medicare Bayesian Improved Surname Geocoding (MBISG) method which relies on claims data, and data from the US Census, and ASI score is calculated using American Community Survey data linked to patient zip code.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

The goal of all facilities is to minimize disparities in health care. With strong variation between the best and worst performing hospitals, there is evidence many facilities can improve how equitable their care is across patients of differences races, SES, and dual enrollment status.

Unintended Consequences

Composite measures, like the HDI, are beneficial in that they summarize detailed information to give a high-level picture of multiple and intersecting variables at play; however, they can be confusing for facilities to interpret in the absence of more explanation. CMS will take this into account in providing feedback reports to hospitals and aims to be responsive to any other issues that arise in use of such a 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

Significant and persistent inequities in health care outcomes exist in the United States. Belonging to a racial or ethnic minority group, living with a disability, being a member of the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) community, living in a rural area, or being near or below the poverty level, is often associated with worse health outcomes.1,2,3,4,5,6,7,8 Numerous studies have shown that among Medicare beneficiaries, racial and ethnic minority individuals often receive lower quality of hospital care, report worse experiences of care, and experience more frequent hospital readmissions and procedural complications.9,10,11,12,13,14 Readmission rates for the most common conditions in the Hospital Readmissions Reduction Program are higher for Black Medicare beneficiaries and higher for Hispanic Medicare beneficiaries with Congestive Heart Failure and Acute Myocardial Infarction.15,16,17,18,19 To ensure that all patients receive excellent care when hospitalized regardless of their individual characteristics, measurement and reporting of disparities is essential.

1 Joynt KE, Orav E, Jha AK. Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care. JAMA. 2011;305(7):675-681.

2 Lindenauer PK, Lagu T, Rothberg MB, et al. Income Inequality and 30 Day Outcomes After Acute Myocardial Infarction, Heart Failure, and Pneumonia: Retrospective Cohort Study. British Medical Journal. 2013;346.

3 Trivedi AN, Nsa W, Hausmann LRM, et al. Quality and Equity of Care in U.S. Hospitals. New England Journal of Medicine. 2014;371(24):2298-2308.

4 Polyakova, M., et al. Racial Disparities In Excess All-Cause Mortality During The Early COVID-19 Pandemic Varied Substantially Across States. Health Affairs. 2021; 40(2): 307-316.

5 Rural Health Research Gateway. Rural Communities: Age, Income, and Health Status. Rural Health Research Recap. November 2018.

6 https://www.minorityhealth.hhs.gov/assets/PDF/Update_HHS_Disparities_Dept-FY2020.pdf

7 https://www.cdc.gov/mmwr/volumes/70/wr/mm7005a1.htm

8 Poteat TC, Reisner SL, Miller M, Wirtz AL. COVID-19 Vulnerability of Transgender Women With and Without HIV Infection in the Eastern and Southern U.S. Preprint. medRxiv. 2020;2020.07.21.20159327. Published 2020 Jul 24. doi:10.1101/2020.07.21.20159327

9 Martino, SC, Elliott, MN, Dembosky, JW, Hambarsoomian, K, Burkhart, Q, Klein, DJ, Gildner, J, and Haviland, AM. Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage. Baltimore, MD: CMS Office of Minority Health. 2020.

10 Guide to Reducing Disparities in Readmissions. CMS Office of Minority Health. Revised August 2018. Available at: https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/OMH_Readmissions_Guide.pdf

11 Singh JA, Lu X, Rosenthal GE, Ibrahim S, Cram P. Racial disparities in knee and hip total joint arthroplasty: an 18-year analysis of national Medicare data. Ann Rheum Dis. 2014 Dec;73(12):2107-15.

12 Rivera-Hernandez M, Rahman M, Mor V, Trivedi AN. Racial Disparities in Readmission Rates among Patients Discharged to Skilled Nursing Facilities. J Am Geriatr Soc. 2019 Aug;67(8):1672-1679.

13 Joynt KE, Orav E, Jha AK. Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care. JAMA. 2011;305(7):675-681

14 Tsai TC, Orav EJ, Joynt KE. Disparities in surgical 30-day readmission rates for Medicare beneficiaries by race and site of care. Ann Surg. Jun 2014;259(6):1086-1090.

15 Rodriguez F, Joynt KE, Lopez L, Saldana F, Jha AK. Readmission rates for Hispanic Medicare beneficiaries with heart failure and acute myocardial infarction. Am Heart J. Aug 2011;162(2):254-261 e253.

16 Centers for Medicare and Medicaid Services. Medicare Hospital Quality Chartbook: Performance Report on Outcome Measures; 2014.

17 Guide to Reducing Disparities in Readmissions. CMS Office of Minority Health. Revised August 2018. Available at: https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/OMH_Readmissions_Guide.pdf

18 Prieto-Centurion V, Gussin HA, Rolle AJ, Krishnan JA. Chronic obstructive pulmonary disease readmissions at minority-serving institutions. Ann Am Thorac Soc. Dec 2013;10(6):680-684.

19 Joynt KE, Orav E, Jha AK. Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care. JAMA. 2011;305(7):675-681

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

2,999

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

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment

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

The HDI aggregates existing stratified measures; those stratified measures adjust for a range of patient demographic and clinical factors using separate models for each measure, social risk factor, and stratification method (Within facility, Across facility). The prototype HDI incorporates the results of 66 risk models.

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

No

Reliability: Type of Reliability Testing

N/A

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

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

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

Continuous Variable – Mean

Measure Performance Score Interpretation

Higher score is better

Mean performance score

0.00

Median performance score

0.00

Minimum performance score

-2.82

Maximum performance score

2.35

Standard deviation of performance scores

.33

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

Jennifer Robinson
7500 Security Blvd
Baltimore, MD 21244
jennifer.robinson@cms.hhs.gov
(443) 729-6368

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

N/A

MUC2022-112 Geriatrics Hospital Measure

Program

Hospital Inpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

This structural measure assesses hospital commitment to improving outcomes for patients greater than or equal to 65 years of age through patient-centered competencies aimed at achieving quality of care and safety for all older patients. The measure will include 14 attestation-based questions across 8 domains representing a comprehensive framework required for optimal care of older patients admitted to the hospital or being evaluated in the emergency department. A hospital will receive a point for each domain where they attest to at least one corresponding statement (for a total of 8 points). For each item, attestation of all elements is required to qualify for the measure numerator.

Numerator

This programmatic measure assesses hospital commitment to improving outcomes for patients 65 years of age and older through patient-centered competencies aimed at achieving quality of care and safety for all older patients. The measure will include 14 attestation-based questions across 8 domains representing a comprehensive framework required for optimal care of older patients admitted to the hospital or being evaluated in the emergency department. A hospital will receive a point for each domain where they attest to at least one corresponding statement (for a total of 8 points). For each item, attestation of all elements is required to qualify for the measure numerator. Note that "patients" in all elements refers to patients 65 years of age and older at time of admission to the hospital or emergency department.

Domain 1: Identifying Goals of Care

Question 1: Advance Care Planning. Please attest that your hospital provides education to patients and providers regarding advance care planning and ensures that advance care planning preferences are captured, updated, and available for review in the medical record.

Question 2: Patient Goals. Please attest that your hospital provides education regarding goal concordant care and has established protocols for ensuring patient goals and decision making is documented in the medical record.

Domain 2: Medication Management

Question 3 Inappropriate Medications. Please attest that your hospital flags medications that may be inappropriate for older patients and has established protocols for reviewing drug and non-drug alternatives to identified substances.

Question 4: Pain Management. Please attest that your hospital employs opioid sparing multimodal pain management strategies where possible and has protocols for capturing these regimens in the medical record.

Domain 3: Cognition and Delirium

Question 5: Delirium and Cognition Screening. Please attest that your hospital performs delirium and cognition screens and assessments and implements appropriate management plans for those with delirium.

Domain 4 : Preventing Delirium Related Events.

Question 6: Delirium Prevention. Please attest that your hospital establishes protocol for minimizing delirium for patients in the hospital through environment modifications, delirium screens, and timely discharge/transfer of patients.

Domain 5: Function and Mobility

Question 7: Function and Mobility Screening. Please attest that your hospital performs function and mobility assessments and implements appropriate management plans to promote mobility.

Question 8: Assistance with Activities of Daily Living (ADLs) / Instrumental Activities of Daily Living (IADLs): Please attest that your hospital screens older patients for ADL/IADL needs and establishes protocols for management of patients with identified deficiencies.

Domain 6. Social Determinants of Health

Question 9: Social Determinants of Health. Please attest that your hospital assesses patients for psychosocial risk factors and employs appropriate management plans.

Question 10: Elder Abuse, Neglect, and Exploitation: Please attest that your hospital assesses older patient for potential abuse and has protocols for intervention for positive assessments including appropriate reporting and involvement of social services.

Domain 7: Care Transitions

Question 11: Identifying Needs at Hospital Discharge Please attest that your hospital elicits discussion between providers and patients regarding discharge care and establishes protocols to ensure that discharge summaries contain management plans for all identified post-discharge needs.

Question 12: Post-Acute Care. Please attest that your hospital has protocols for establishing two-way communication between providers and post-acute care facilities and tracks the quality of care at post-acute care facilities upon discharge.

Domain 8: Ensuring Quality Care for High-Risk Patients

Question 13: Identification and Management of Seriously Ill Patients: Please attest that your hospital employs multidisciplinary evaluation of older patients and provides appropriate management, including the early utilization of palliative care consultations, for those with serious illness.

Question 14: Geriatric Leader and Quality Framework. Please attest that your hospital designates a geriatric champion to oversee all aspects of this measure and establishes a framework for ongoing quality improvement regarding the care for older patients.

Numerator Exclusions

N/A

Denominator

The denominator for each hospital is 8 which represents the total number of domains with at least one complete attestation.

The measure is calculated as the number of complete attestations / total number of domains. There is no partial credit for any question. Attestation of at least one element in all 8 domains is required to qualify for the measure numerator.

Denominator Exclusions

N/A

Denominator Exceptions

N/A

State of development

Specification

State of Development Details

Testing is currently being conducted in the participating programs.

What is the target population of the measure?

65 and older

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

Geriatric medicine

Measure Type

Other: Other

Is the measure a composite or component of a composite?

Not a composite or component of a composite measure

If Other, Please Specify

Programmatic Measure

What data sources are used for the measure?

Administrative Data (non-claims);Claims Data;Electronic Clinical Data (non-EHR);Electronic Health Record;Paper Medical Records;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?

Facility

In which setting was this measure tested?

Community hospital;Hospital inpatient acute care facility;Veterans Health Administration facility

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

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)

N/A

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?

Other: Web-based tool in CMS quality reporting portal.

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

Feasibility assessments have been conducted; findings have demonstrated feasibility at the hospital level for all components. Attestation components are defined and feasible to address with reliability testing accomplished (see reliability). This has been accomplished via clinical registries, EHR, claims data, and ICD 9 coding. It is important to recognize that not all sites need to use the same method of data collection, given benchmarking is not being executed across sites but conducted longitudinally. Hence, individual sites need to remain consistent in their data source.

Method of Measure Calculation

Other (enter here):: Attestation using a web-based tool within the HQR system.

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

See "GSV IHI GEDA Measure Evidence" attachment for performance gap information.

Unintended Consequences

Potential known unintended consequences in the geriatric population would be due to efforts around function/mobility. An anticipated increase in falls might occur as patients are encouraged to ambulate and/or management plans focus on efforts to mitigate deconditioning. If/when patients fall, ambulation efforts might be halted, which can then have the unintended consequences of deconditioning, restraint use, and/or pressure ulcers.

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

9

Outline the clinical guidelines supporting this measure

See the "GSV IHI GEDA Measure Evidence" attachment for clinical guidelines supporting this measure.

Name the guideline developer/entity

The American Geriatrics Society (AGS)

Publication year

2019

Full citation +/- URL

Boyd C, Smith CD, Masoudi FA, Blaum CS, Dodson JA, Green AR, Kelley A, Matlock D, Ouellet J, Rich MW, Schoenborn NL, Tinetti ME. Decision Making for Older Adults With Multiple Chronic Conditions: Executive Summary for the American Geriatrics Society Guiding Principles on the Care of Older Adults With Multimorbidity. J Am Geriatr Soc. 2019 Apr;67(4):665-673. doi: 10.1111/jgs.15809. Epub 2019 Mar 10. PMID: 30663782.

Is this an evidence-based clinical guideline?

Yes

Is the guideline graded?

No

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

The recommended MCC Actions include: (1) identify and communicate patients' health priorities and health trajectory; (2) stop, start, or continue care based on health priorities, potential benefit vs harm and burden, and health trajectory; and (3) align decisions and care among patients, caregivers, and other clinicians with patients' health priorities and health trajectory.

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 recommended MCC Actions include: (1) identify and communicate patients' health priorities and health trajectory; (2) stop, start, or continue care based on health priorities, potential benefit vs harm and burden, and health trajectory; and (3) align decisions and care among patients, caregivers, and other clinicians with patients' health priorities and health trajectory.

Number of systematic reviews that inform this measure concept

18

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

See the "GSV IHI GEDA Measure Evidence" attachment.

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

See the "GSV IHI GEDA Measure 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

4,000,000

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?

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

Not conceptually or empirically indicated (enter here):: This is a programmatic measure for a facility to attest to specific standards for delivering high quality care to elderly patients. There are no clinical or social factors that would require risk adjustment and all facilities should be held to the same standard for caring for geriatric patients.

Cost estimate completed

Yes

Cost estimate methods and results

No additional cost, can use existing resources.

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

3

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

3

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.

70

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

70

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?

760

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

Other (enter here): Test-Retest

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

Kappa - Interrater Reliability (IRR)

Other: Sample size

240

Other: Statistical result

0.98

Other: Interpretation of results

From Pilot and current site evaluations, there is high Interrater reliability (IRR). IRR amongst the two raters had 99% agreement amongst 2 raters.

Empiric Validity

Yes

Empiric Validity: Statistic name

Modified Delphi Method (modified version of the RAND-UCLA Appropriateness Methodology)

Empiric Validity: Sample size

8

Empiric Validity: Statistical result

8

Empiric Validity: Methods and findings

1.) JAHF grant awarded to ACS to develop the Geriatric Surgery Verification and Quality Improvement Program (now known as the Geriatric Surgery Verification (GSV) Program). CQGS Core Development Team (CDT), Advisory Panel and multidisciplinary stakeholder groups convened to set the standards for geriatric surgical care. To achieve this objective and set the program standards, key benchmarks were established/achieved:

1.1.) Gather Literature and Develop Preliminary Standards: The CQGS CDT completed a comprehensive literature review in February 2016 and released the preliminary standards to stakeholders in April 2016.

1.2.) Conduct Preliminary Field Visits: Between December 2015 and April 2016, the CDT conducted 11 field visits in 7 cities across the country to measure the current state of surgery in older adults, evaluate scalability of the program, identify best practices already in place serving older adult surgery patients and gain perspectives/opinions from frontline staff. CDT reported on the field visit findings to CQGS stakeholders in May 2016.

1.3.) Refine and Vet Standards w/ Stakeholders:

1.3.1.) CQGS Preliminary Standards written in 2016 by the CDT, and based on ratings analysis, standards were repackaged from 308 standards into 92 Alpha Standards in 2017 [using a modified version of the RAND-UCLA Appropriateness Methodology, 44 of the 58 CQGS stakeholders performed two ratings on the validity and feasibility of the standards. Based on the ratings analysis, became 92 Alpha Standards and implemented into Alpha Pilot].

1.3.2.) The CQGS invited 15 hospitals to participate in a survey to determine which of the standards were already implemented, understand how easy or difficult the standards would be to implement if not already in practice, and identify and record standards that were confusing or difficult to interpret.

1.3.3.) From the results of Alpha Pilot, the 92 Alpha Standards were further condensed to 30 Beta Standards, categorized into 6 chapters: Program management, Goals and Decision-Making, Preoperative Optimization, Immediate Preoperative and Intraoperative Clinical Care, Postoperative Clinical Care, and Transitions of Care. The Beta Standards were released in November 2017.

1.3.4.) 8 hospitals participated in the Beta Pilot in December 2017 to determine the feasibility of implementation for each beta standard, collect "best practices" for standards implementation, learn barriers to implementation in hospitals of varying size, facility-type, and location, identify issues in the wording of the standards leading to misinterpretation, continue to record rationale for or against the standards, and resolve difficulties in the verification process. Beginning in June of 2018, site visits to each pilot hospital were performed and report was released end of 2018

2.) Using the data from the Beta Pilot, the CQGS team finalized the standards and developed educational and supportive materials to aid in the official GSV Program launch in July 2019 with release of Optimal Resources for Geriatric Surgery 2019 Standards.

3.) GSV Program began accepting applications for enrollment in October 2019. To date: 53 hospitals enrolled in the program in various stages of implementation. 3 hospitals have completed verification process.

4.) In 2021, ACS conveyed a team of experts in geriatric surgical care to evaluate the GSV standards to determine feasibility as a measure that collectively improves care for older adults.

5.) In January 2022, 9 unique hospitals committed to providing optimal geriatric surgical care (as demonstrated through their established centers for geriatrics, participation in age-friendly health initiatives, or enrollment in the GSV Program, or any combination of these elements), were surveyed to evaluate implementation of key measures that collectively improve care for older adults. The key measure domains were determined through the combination of both the established ACS quality framework, and the processes, resources, and infrastructures necessary for the optimal care of the older adult surgical patient as determined by JAHF grant work.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

70

Face Validity: Result

44

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

7

Median performance score

6

Minimum performance score

0

Maximum performance score

8

Standard deviation of performance scores

0000

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 College of Surgeons (ACS) (primary steward); American College of Emergency Physicians (ACEP); Institute for Healthcare Improvement (IHI)

Measure Steward Contact Information

Clifford Ko
633 N Saint Clair St
Chicago, IL 60611
geriatricsurgery@facs.org
(312) 202-5518

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Michael Bencur
633 N Saint Clair St

Chicago, IL 60611
mbencur@facs.org
(312) 202-5120

Secondary Submitter Contact Information

Jill Sage
20 F Street, NW
Suite 1000
Washington, DC 20001
jsage@facs.org
(202) 672-1507

Submitter Comments

The Geriatric Hospital Measure is a new type of measure--a programmatic measure--one that that values the full quality program needed to care for geriatric patients.

The US population is rapidly aging and across the country, hospitals are increasingly faced with older patients who have complex medical, physiological, and psychosocial needs that are often inadequately addressed by the current healthcare infrastructure. Although existing quality metrics have improved both the rate and reporting of clinical outcomes (falls, appropriate use of anticoagulants, etc.) that are important to older individuals, these measures can be narrow in scope and may have limited long term effectiveness due to ceiling effects. Rather than simply addressing individual clinical issues in isolation, optimizing care for older patients with multifaceted vulnerability profiles will require a holistic approach with the goal of reframing the entire care pathway to better serve the needs of this unique population. Despite this growing need, our healthcare system has not comprehensively rethought care for the complex geriatric population since the creation of Medicare more than 50 years ago. To this end, multiple organizations including the American College of Surgeons (ACS), the Institute for Healthcare Improvement (IHI), and the American College of Emergency Physicians (ACEP) have identified clinical frameworks based on evidence-based best practices that provide goal-centered, clinically effective care for older patients.

Part of what is needed in rethinking care for the older adult population is programmatic facility-level geriatric measurement. This solution is different from the current types of CMS measures, this programmatic measure incentivizes team-based care organized around the geriatric patient to meet the challenges unique to their needs. The measure differs from the classical NQF structural measures or process measures that are usually singular structural components or a simple process tied to a transaction/visit. The challenge is care is not a single structural element or process. It is the collection of all these components orchestrated across the continuum of care for the entire team in a patient centered manner. Together, these become a patient-centered programs of care. When the components are properly tied together, care becomes well-coordinated, complex aspects of care are more reliably delivered, harms are minimized, and outcomes are optimized. The elements in the program are focused around care delivery, coordination, data, and data-driven improvement activities.

When CMS and NQF consider traditional structural and process measures, we agree that singular elements within one or another transaction as part of a series of transactions in health are check-the-box measures. These have limited impact on quality or improvement. However, within clinical domains

of care such as geriatric care, there are crucial structures and processes of care that reach across multiple transactions and link the care team's efforts together. Orchestrating all these elements result in better outcomes and improving their implementation would be an essential first step to improve geriatric outcomes.

CMS quality programs currently consist of a large, extremely costly universe of measures in multiple different payment programs. They often lack the consideration for focusing a care team in a patient-centered way. Such sporadic measurement creates a massive amount of burden and overhead but has a limited impact on improving the quality of overall care and fails to create accountability to patients for the care they receive. The development of individual measures and the subsequent combination of these measures into payment incentive programs may be useful for fee-for-service payment. Value-based payments need a more condition, patient-type approach. This proposed approach is programmatic within a clinical domain. Its implementation will create a team-based approach to optimizing the patient's chances to achieve their desired outcome. Disjointedly measuring members of a care team with disparate measure sets does not create the alignment needed.

Another benefit of this programmatic approach is the usefulness of this information for patients deciding where to seek care. Twenty years of NQF and payer actions in quality have not produced reliable public knowledge or a public-facing website that informs patients about where to get the care they need for the condition they have. This information, along with comparable information on the price of that care, are the prerequisites for a valid depiction of the value of care.

An important standard to apply in evaluating payment quality incentives is their effectiveness in providing patients with knowledge of where to find high quality care in their community. In assessing the effectiveness of our measures, we wonder if the patient had this information, would it enable them to easily find information on a website for the types of care they seek, for a safety and equitability profile and for personal goal attainment. Our measures do not meet all these objectives since these are the initial steps. They are also designed to be added to appropriate condition or procedure specific cost measures which would help patients determine the affordability of the care they desire. Combing quality and price for care is a key step in establishing value.

Hospital Outpatient Quality Reporting Program

MUC2022-020 Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Hospital Level – Outpatient)

Program

Hospital Outpatient Quality Reporting Program

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 hospital outpatient care settings (including emergency 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?

Facility

In which setting was this measure tested?

Community hospital;Emergency department;Hospital outpatient department (HOD)

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

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

N/A - not a MIPS measure

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 ID: 3663e

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: 1074 (QDM version) and 1075FHIR (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?

There are three process measures related to CT in hospital outpatient settings, but none directly addresses radiation dose: (1) Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke who Received Head CT or MRI Scan Interpretation Within 45 Minutes of ED Arrival, CMIT 918; (2) Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac Low-Risk Surgery, CMIT 1367; and Abdomen Computed Tomography (CT) Use of Contrast Material, CMIT 2599. There are no measures addressing CT or radiation doses in the hospital inpatient reporting program. Three existing MIPS 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 7 different EHR systems reflecting 16 outpatient and emergency hospital settings [Epic (N=5), Cerner (N=1), Allscripts (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 outpatient hospital sites (including emergency settings), the rates of out-of-range exams varied 20%-45% by site. 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. (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 $3250 per hospital 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: We estimate the implementation costs at $3250 per hospital based on the time and costs reported by our field-testing sites (see feasibility attachment for more information). This estimate is likely conservative, as our testing partners noted that the work of assembling the relevant data decreased over time. For hospitals that are part of large health systems, the cost may be incurred at the health system level rather than at the level of the individual hospital.

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.

15

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

15

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.

Random Split-Half Correlation: Sample size

16

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 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, the measure is reliable at the hospital level. Given the high volume of CT, virtually no hospitals would fall below the minimum denominator to achieve ICC >0.90.

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

15

Face Validity: Result

15

Patient/Encounter Level Testing

Yes

Type of Analysis

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

Sample Size

22,587

Statistic Name

Percent agreement

Statistical Results

0.91

Interpretation of results

See the Patient/Encounter Level Validity Testing Attachment.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Lower score is better

Mean performance score

0.31

Median performance score

0.30

Minimum performance score

0.20

Maximum performance score

0.45

Standard deviation of performance scores

0.08

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 Box 0560
San Franciso, CA 94044
nate@alaracare.com
(650) 520-6649

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Rebecca Smith-Bindman
550 16th 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.edu
(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. 100% 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.

MUC2022-030 Hospital Outpatient Department Volume Data on Selected Outpatient Surgical Procedures (formerly OP-26)

Program

Hospital Outpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Structural measure of facility capacity collects surgical procedure volume data on selected categories of outpatient procedures frequently performed within the outpatient department (e.g., outpatient surgery, cath lab, endoscopy). Gastrointestinal, Eye, Nervous System, Musculoskeletal, Skin, Genitourinary, Cardiovascular, Respiratory, and Other

Numerator

All-patient, all payer surgical volume data for 9 categories of outpatient surgical procedures Cardiovascular, Eye, Gastrointestinal, Genitourinary, Musculoskeletal, Nervous System, Skin, Respiratory and Other within a one-year performance period.

Numerator Exclusions

Excludes procedures performed within the emergency department (ED)

Denominator

N/A

Denominator Exclusions

N/A

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

Other: No Specialty

Measure Type

Structure

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;Electronic Clinical Data (non-EHR)

If applicable, specify the data source

N/A

Description of parts related to these sources

Structural web based. Facilities report the count of performed surgical procedures per category. Data entry will be achieved through the secure side of QualityNet.cms.gov via an online tool available to authorized users.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Hospital outpatient department (HOD)

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

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

N/A

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

Yes

CMIT ID

02753-C-HC

Alternate Measure ID

OP26

What is the endorsement status of the measure?

Never Submitted

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

99999

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, but the measure is undergoing substantial change

Range of years this measure has been used by CMS Programs

This measure was first adopted in the CY2012 OPPS/ASC final rule. It was finalized for removal in CY2018 OPPS/ASC final rule. (2012-2019)

What other federal programs are currently using this measure?

This measure was first adopted in the CY2012 OPPS/ASC final rule. It was finalized for removal in CY2018 OPPS/ASC final rule.

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

Because this is an aggregate account of procedures performed, facilities should be able to readily obtain and submit these counts through QualityNet.

Previously CMS estimated that participating hospital will spend 10 minutes per year to collect and submit the data, making the estimated annual burden associated with this measure 533 hours (3,200 hospitals x 0.167 hours per measure x 1 all-patient volume measure per hospital).(76 FR 74552).

Method of Measure Calculation

Other (enter here):: This measure strictly counts the procedures performed in each separate facility over each 1-year performance period.

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

According to a 2021 MedPAC analysis, the volume of outpatient services for Medicare FFFS beneficiaries continued to increase, in both 2018 and 2019 [1]. The increase in the volume of outpatient services is in part due to the continued shift of complex surgical procedures from the inpatient setting to the outpatient setting. Examples of these procedures include: knee replacement; endovascular procedures; and removal, replacement, or insertion of defibrillator systems or pulse generators.

[1] Medicare Payment Advisory Commission. March 2021. Report to the Congress: Medicare and the health care delivery system. Chapter 3. Washington, DC: MedPAC

Unintended Consequences

N/A

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

000000

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

As noted above, more complex procedures have been moving from the inpatient setting to the outpatient setting [1]. This underscores the need to address the potential for poorer outcome for beneficiaries who are treated at low volume providers, for procedures that have a volume-outcome relationship. In addition, better understanding the volume of procedures from an all-payer perspective will allow CMS to target and prioritize future quality measure development.

There are scores of systematic reviews that examine the volume outcome relationship for surgery. Most of the reviews that have been published support a volume-outcome relationship, but the relationship is weak for some procedures, and stronger for others. In addition, most studies have addressed procedures that are performed in the inpatient setting (although in some cases, like for knee replacement surgery those procedures are migrating to the outpatient space); studies differ in if they examined surgeon volume vs. hospital volume. A recent scoping review of the volume/outcome relationship examined 403 studies that addressed 90 types of surgery. Study authors found that most (about 87%) of the studies had a significant volume-outcome relationship; there were 61 different types of outcomes that were examined in these studies. About half of the studies addressed cancer-related surgery [2].

Below we summarize the one systematic review that has addressed outpatient surgery.

One systematic review published in 2020 [3] examined outpatient surgery using international data, analyzed data from eight retrospective studies that addressed seven procedures: anterior cruciate ligament reconstruction, cataract surgery, meniscectomy, thyroidectomy, primary hip arthroscopy, open carpal tunnel release, and rotator cuff repair. Study authors found a volume outcome relationship for all but carpal tunnel release and thyroidectomy, however the results did not allow the study authors to recommend clear volume thresholds for these procedures.

[1] Medicare Payment Advisory Commission. March 2021. Report to the Congress: Medicare and the health care delivery system. Chapter 3. Washington, DC: MedPAC

[2] Levaillant, M., Marcilly, R., Levaillant, L. et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. BMC Med Res Methodol 21, 204 (2021). https://doi.org/10.1186/s12874-021-01396-6.

[3] Stanak, M., Strohmaier, C. Minimum volume standards in day surgery: a systematic review. BMC Health Serv Res 20, 886 (2020). https://doi.org/10.1186/s12913-020-05724-2.

Source of empirical data

N/A

Summarize the empirical data

N/A

Name evidence type

Published peer-reviewed original research

Summarize the evidence

There have been many studies that have examined the relationship between surgeon/facility volume of procedures and procedural outcomes. As noted above, volume-outcome relationships are commonly found across surgeries [2] however there has been less focus on this relationship for outpatient surgery. One systematic review found a volume-outcome relationship for five of seven outpatient procedures [3].

[2] Levaillant, M., Marcilly, R., Levaillant, L. et al. Assessing the hospital volume-outcome relationship in surgery: a scoping review. BMC Med Res Methodol 21, 204 (2021). https://doi.org/10.1186/s12874-021-01396-6.

[3] Stanak, M., Strohmaier, C. Minimum volume standards in day surgery: a systematic review. BMC Health Serv Res 20, 886 (2020). https://doi.org/10.1186/s12913-020-05724-2.

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

00000

Type of Evidence to Support the Measure

Peer-Reviewed Systematic Review;Other (enter here):: Peer-reviewed original research

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

Not conceptually or empirically indicated (enter here):: This is a structural measure regarding raw counts.

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

No

Reliability: Type of Reliability Testing

N/A

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

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

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: Raw count

Measure Performance Score Interpretation

Other (enter here):: Raw count

Mean performance score

00000

Median performance score

00000

Minimum performance score

1

Maximum performance score

00000

Standard deviation of performance scores

00000

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

CMS

Measure Steward Contact Information

Janis Grady
7500 Security Boulevard
Baltimore, MD 21244
janis.grady@cms.hhs.gov
(410) 786-7217

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

N/A

Hospital Value-Based Purchasing Program

MUC2022-082 Severe Sepsis and Septic Shock: Management Bundle

Program

Hospital Value-Based Purchasing Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with Surviving Sepsis Campaign guidelines, it assesses measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, the first three interventions should occur within three hours of presentation of severe sepsis, while the remaining interventions are expected to occur within six hours of presentation of septic shock.

Numerator

Patients who received ALL of the following:

Within three hours of presentation of severe sepsis:

- Initial lactate level measurement
- Broad spectrum or other antibiotics administered
- Blood cultures drawn prior to antibiotics

AND received within six hours of presentation of severe sepsis. ONLY if the initial lactate is elevated:

- Repeat lactate level measurement

AND within three hours of initial hypotension:

- Resuscitation with 30 mL/kg crystalloid fluids

OR within three hours of septic shock:

- Resuscitation with 30 mL/kg crystalloid fluids

AND within six hours of septic shock presentation, ONLY if hypotension persists after fluid administration:

- Vasopressors are administered

AND within six hours of septic shock presentation, if hypotension persists after fluid administration or initial lactate >= 4 mmol/L:

- Repeat volume status and tissue perfusion assessment is performed

Numerator Exclusions

N/A

Denominator

Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock and not equal to U07.1 (COVID-19).

Denominator Exclusions

The following patients are excluded from the denominator:

- Patients with an ICD-10-CM Principal or Other Diagnosis Code of U07.1 (COVID-19)
- Directive for Comfort Care or Palliative Care within six hours of presentation of severe sepsis
- Directive for Comfort Care or Palliative Care within six hours of presentation of septic shock
- Administrative contraindication to care within six hours of presentation of severe sepsis
- Administrative contraindication to care within six hours of presentation of septic shock
- Length of Stay >120 days
- Transfer in from another acute care facility
- Patients enrolled in a clinical trial for sepsis, severe sepsis or septic shock treatment or intervention
- Patients with severe sepsis who are discharged within six hours of presentation
- Patients with septic shock who are discharged within six hours of presentation
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis

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

Critical care medicine (intensivists)

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?

Electronic Health Record;Paper Medical Records

If applicable, specify the data source

N/A

Description of parts related to these sources

The entire measure should be abstracted from a single inpatient hospital record, regardless of electronic or paper record.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

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

N/A

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

Yes

CMIT ID

01017-C-HIQR

Alternate Measure ID

SEP1

What is the endorsement status of the measure?

Endorsed

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

0500

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

2021

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

2025

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?

2012

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

845

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

Hospital

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

IQR, OQR, LTCHQR

What was the MAP recommendation in this year?

Support: Addresses a NQS priority not adequately addressed in the program measure set

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

N/A

MAP report page number being referenced for this year:

2012, page 125

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

2015-current

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program

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

The Hospital VBP Program is a pay-for-performance program established by Section 3001(a) of the ACA.

Section 2: Measure Evidence

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

Other: chart abstracted patient level data, same submission method as hospital IQR program

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

This measure has been a chart abstracted measure in the hospital IQR program since October 2015 and has been publicly reported on the Care Compare website since July 2018. All of the information required to report the measure is available in the medical record. Most of the questions received from abstractors are related to applying the measure guidance to patient specific situations.

Method of Measure Calculation

Manual abstraction

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

See section 2b4. "Identification of Statistically Significant & Meaningful Differences in Performance", starting on page 23 of the attached NQF Testing Attachment for full details.

First, we calculated the mean; standard deviation; median; and 5th, 10th, 25th, 75th, 90th, and 95th percentile of the performance scores for each quarter. Next, we grouped hospitals by deciles and assessed whether the difference in mean measure score between each adjacent decile was statistically significant. Our goal was to determine whether there are significant differences in performance across hospitals. Finally, we compared whether there is a statistically significant difference in mean measure score by age, gender, race, ethnicity, and payer using an analysis of variance (ANOVA) analysis.

Table 1. Distribution of measure score by quarter for Q3 2018 and Q4 2018

Quarter (number of hospitals) // Mean // Standard deviation // Min // 5th // 10th // 25th // 50th // 75th // 90th // 95th // Max

Q3 2018 (3,222) // 0.58 // 0.22 // 0.00 // 0.17 // 0.29 // 0.44 // 0.59 // 0.73 // 0.85 // 0.91 // 1.00

Q4 2018 (3,235) // 0.58 // 0.23 // 0.00 // 0.13 // 0.29 // 0.45 // 0.60 // 0.74 // 0.85 // 0.91 // 1.00

Q3 and Q4 combined (3,302) // 0.57 // 0.21 // 0.00 // 0.19 // 0.30 // 0.45 // 0.60 // 0.71 // 0.82 // 0.88 // 1.00

Table 2. Differences between adjacent deciles of performance using a two-proportion z-test for Q3 and Q4 2018 combined

Percentile comparison // Pooled sample proportion (standard error) // Test statistic // p-value

10th vs 20th percentile // 0.32 (0.0054) // 28.60 // <0.001

20th vs. 30th percentile // 0.41 (0.0044) // 19.26 // <0.001

30th vs. 40th percentile // 0.49 (0.0044) // 15.33 // <0.001

40th vs. 50th percentile // 0.55 (0.0044) // 11.46 // <0.001

50th vs. 60th percentile // 0.59 (0.0043) // 10.78 // <0.001

60th vs. 70th percentile // 0.64 (0.0041) // 11.95 // <0.001

70th vs. 80th percentile // 0.69 (0.00399) // 12.48 // <0.001

80th vs. 90th percentile // 0.75 (0.0039) // 17.13 // <0.001

90th vs. 100th percentile // 0.82 (0.0039) // 23.77 // <0.001

Table 3. Disparities analysis (using ANOVA test) for Q3 and Q4 2018 combined

Patient characteristic // Number of encounters // First quartile measure score // Median measure score // Third quartile measure score // Measure score // p-value

Age (p-value <0.001)

18-35 // 14,577 // 0.423 // 0.667 // 1.00 // 0.613

36-64 // 82,404 // 0.450 // 0.600 // 0.741 // 0.588

65+ // 136,771 // 0.444 // 0.593 // 0.720 // 0.590

Gender (p-value 0.002)

Female // 112,863 // 0.436 // 0.585 // 0.722 // 0.582

Male // 120,873 // 0.455 // 0.609 // 0.743 // 0.598

Unknown // 16 // 0.375 // 1.00 // 1.00 // 0.750

Race (p-value 0.012)

Black // 30,676 // 0.399 // 0.600 // 0.805 // 0.558

Other // 7,158 // 0.333 // 0.667 // 1.00 // 0.621

Unknown // 13,748 // 0.333 // 0.667 // 1.00 // 0.583

White // 182,170 // 0.455 // 0.595 // 0.72 // 0.595

Ethnicity (p-value <0.001)

Hispanic // 20,575 // 0.400 // 0.643 // 0.929 // 0.585

Non-Hispanic // 213,177 // 0.455 // 0.596 // 0.714 // 0.591

Payer (p-value <0.001)

Medicare // 152,784 // 0.444 // 0.591 // 0.716 // 0.589

Non-Medicare // 80,968 // 0.462 // 0.609 // 0.750 // 0.593

The measure was able to detect facilities with above- and below-average performance. The facility measure scores ranged from 0.0% to 100.0%, with a mean performance of 57% and a standard deviation of 21%. Our analysis showed a statistically significant difference in performance between each decile of hospitals, suggesting consistent performance gaps across facilities. We identified statistically significant differences in mean measure scores depending on age, payer, ethnicity, gender, and payer. The disparities across these groups highlight the importance of continuing to track sepsis care quality.

Unintended Consequences

None were reported. We have not found evidence in the published literature that clearly demonstrates unintended consequences from implementation of the measure.

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

1

Outline the clinical guidelines supporting this measure

The following presents six SEP-1 elements of care as they relate to the recommendations and quality evidence ratings in the 2016 SSC guidelines and updated in 2021. The following also notes the implications of the recommendations for patients and providers, which reflects the importance of the SEP-1 measure.

1) SEP-1 element of care: Measure lactate levels and remeasure if initial lactate is >= 2 mmol/L.

SSC guideline recommendation: Obtain initial lactate levels as a marker of tissue hypoperfusion and normalize lactate in patients with elevated lactate levels.

Strength of recommendation, quality of evidence: Weak recommendation, low quality of evidence (no change from 2016 to 2021)

Implications of recommendation: The desirable effects of adherence to this recommendation probably will outweigh the undesirable effects. Consider therapy tailored to patient circumstances.

2) SEP-1 element of care: Obtain blood cultures prior to antibiotics.

SSC guideline recommendation: Obtain blood cultures before starting antimicrobial therapy in patients with suspected sepsis or septic shock.

Strength of recommendation, quality of evidence: Best practice statement (no change from 2016 to 2021)

Implications of recommendation: The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

3) SEP-1 element of care: Administer broad-spectrum antibiotics.

SSC guideline recommendation: Administer IV antibiotics as soon as possible after recognition of sepsis.

Strength of recommendation, quality of evidence: Strong recommendation, moderate quality of evidence (overall similar from 2016 to 2021 still a strong recommendation; evidence identified as low in 2021 and expanded on in 2021)

Implications of recommendation: The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

4) SEP-1 element of care: Administer crystalloid fluids for hypotension or lactate

SSC guideline recommendation: Administer crystalloid fluid within the first three hours of sepsis-induced hypoperfusion.

Strength of recommendation, quality of evidence: Strong recommendation, low quality of evidence (strength of recommendation for a 30 mL/kg crystalloid fluid volume downgraded from strong in 2016 to weak in 2021, primarily related to shifting to a fluid volume approach based on patient response instead of a set volume)

Implications of recommendation: The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

5) SEP-1 element of care: Vasopressors for hypotension that does not respond to initial fluid resuscitation

SSC guideline recommendation: Administer vasopressors for refractory hypotension.

Strength of recommendation, quality of evidence: Strong recommendation, moderate quality of evidence (no change from 2016 to 2021)

Implications of recommendation: The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

6) SEP-1 element of care: Reassess volume status and tissue perfusion after fluid administration

SSC guideline recommendation: Frequent reassessment of hemodynamic status following initial fluid resuscitation.

Strength of recommendation, quality of evidence: Best practice statement (BPS in 2016; weak recommendation, very low-quality evidence in 2021)

Implications of recommendation: The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

Name the guideline developer/entity

Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016 (2016 SSC guidelines) // Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021 (updated 2021 SSC guideline

Publication year

2016

Full citation +/- URL

Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016, Critical Care Medicine: March 2017 - Volume 45 - Issue 3 - p 486-552 doi: 10.1097/CCM.0000000000002255

https://journals.lww.com/ccmjournal/fulltext/2017/03000/surviving_sepsis_campaign__international.15.aspx

Evans L, Rhodes A, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021, Critical Care Medicine: November 2021 - Volume 49 - Issue 11 - p e1063-e1143 doi: 10.1097/CCM.0000000000005337

https://journals.lww.com/ccmjournal/fulltext/2021/11000/surviving_sepsis_campaign__international.21.aspx

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.

Sepsis and septic shock are medical emergencies, and we recommend that treatment and resuscitation begin immediately.

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?

The GRADE methodology is based on the assessment of evidence based on six categories: 1) risk of bias, 2) inconsistency, 3) indirectness, 4) imprecision, 5) publication bias, and 6) other criteria. In addition, assessment of the balance between the benefit and harm, patient values and preferences, cost and resources, and feasibility and acceptability of the intervention are considered. The guideline panel final recommendations are based on the assessment of these factors.

GRADE methodology classifies recommendations as strong or weak.

A strong recommendation reflects that the desirable effects of adherence to a recommendation will clearly outweigh the undesirable effects.

A weak recommendation means that the desirable effects of adherence to a recommendation probably will outweigh the undesirable effects, but the trade-offs are not clear, either because some of the evidence is low quality or the benefits and potential harms are closely balanced.

Some interventions carry best practice statements (BPSs), which are ungraded strong recommendations applied under strict criteria. The SSC guidelines use BPSs when the benefit or harm is clear, but the evidence is difficult to summarize or assess using the GRADE methodology.

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

Other (enter here):: Best practice statement

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?

Determination of the Quality of Evidence

Underlying methodology

1. High: RCTs

2. Moderate: Downgraded RCTs or upgraded observational studies

3. Low: Well-done observational studies with RCTs

4. Very Low: Downgraded controlled studies or expert opinion or other evidence

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):: Evidence is from observational studies, but this recommendation is considered a best practice and there is no new evidence suggesting a change is warranted.

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

Sepsis and septic shock are medical emergencies, and we recommend that treatment and resuscitation begin immediately.

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

467,504

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

Not conceptually or empirically indicated (enter here):: Not indicated for 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

R Statistic

Signal-to-Noise: Sample size

3,302

Signal-to-Noise: Statistical result

0.88

Signal-to-Noise: Interpretation of results

See section 2a2. "Reliability Testing", starting on page 8, of the attached NQF Testing Attachment for additional details on signal-to-noise sample size and result.

Across all facilities, the mean and 25th percentile of reliability for each quarter exceeded the 0.70 threshold for acceptable reliability. We also found acceptable reliability scores (> 0.70) across all deciles of hospitals by denominator size, indicating that reliability is high for hospitals regardless of denominator size. These results indicate that the measure can identify true differences in performance between individual facilities.

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

Measure score validity analysis (facility-level): We calculated the percentage of total deaths at discharge among cases for the initial measure population, cases that were excluded from the measure, all eligible cases, cases that passed the measure, and cases that failed the measure. We hypothesized that measure performance and mortality should be associated. To test the hypothesis of whether SEP-1 is associated with mortality rates, we conducted a Chi-square of Association and Equal Proportions test between the two categorical variables: measure outcome (failed or passed) and mortality result (died or survived). To assess the direction and strength of the association between SEP-1 compliance and mortality, we calculated the risk ratio with a two-sided 95% confidence interval. If the risk ratio is above 1.0, indicating that cases that fail the measure have a higher risk of mortality compared to cases that pass the measure, and the confidence interval does not span 1.0, this would support the validity of the measure. Next, we calculated pass rates for hospitals and plotted by the facilities measure performance versus mortality rates using two approaches. In the first approach, we calculated pass rate deciles based on the distribution of hospitals pass rates and assigned each hospital into a percentile grouping based on their respective pass rates. We calculated the overall pass rates for each of the ten percentile groups along with the calculated mortality rates for each percentile group and plotted mortality for each of the ten pass rate percentile groups. The second approach was to group hospitals into ten pass rate buckets. We defined these buckets based on hard cut-offs in measure performance (e.g. hospitals with measure performance of 0%-10% would be in one bucket, hospitals with measure performance of 10.01%-20% would be in the next bucket, and so on). We calculated the overall pass rates each of the ten hard cut-off pass rate buckets along with the calculated mortality rates for each bucket and plotted mortality rates for each of the ten pass rate buckets. We also calculated two-proportion z-tests for each of the pass rate deciles and mortality deciles in order to distinguish whether there are meaningful differences across the measure population. We used a p-value cut-off of 0.05 to define statistically significant differences between deciles.

Empiric Validity: Sample size

3,302

Empiric Validity: Statistical result

1.84

Empiric Validity: Methods and findings

See section 2b1. "Validity Testing", starting on page 9, of the attached NQF Testing Attachment for full details of empiric validity.

The risk ratio of 1.84, indicates that cases that fail the measure have 1.84 times the risk of dying compared to cases that pass the measure. On average there is a 95% chance that the true mortality risk for cases that fail the measure compared to cases that pass the measure is captured in the interval 1.79 to 1.88. The graphs displaying mortality by both pass rate percentile groups and pass rate buckets show an inverse relationship between pass rates and mortality rates, suggesting that SEP-1 compliance is associated with a reduction in mortality. We also found that seven out of ten percentiles comparisons have a statistically significant difference between mortality rates at a significance level of 0.05; all adjacent percentile comparisons for measure performance have statistically significant differences.

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 other gold standard and manual reviewer

Sample Size

916

Statistic Name

Kappa

Statistical Results

0.22

Interpretation of results

See section 2b1. "Validity Testing", on pages 9 to 20 of the attached NQF Testing Attachment for full details.

According to McHugh, a kappa value of less than 0 to 0.20 indicates no agreement, 0.21 to 0.39 indicates minimal agreement, 0.40 to 0.59 indicates weak agreement, 0.60 to 0.79 indicates moderate agreement, 0.80 to 0.90 indicates strong agreement, and a value above 0.9 indicates almost perfect agreement. Fifteen of the 19 critical categorical data elements with a defined kappa had a kappa value in the moderate to high range (> 0.60). One element with lower percent agreement and kappa (Blood Culture Acceptable Delay) is based on a small number of cases. This data element pertains to only 2.0 percent of cases for the overall CDAC sample, as shown in Table 2b1.3.1 in NQF Testing Form, and thus

is not likely to affect the validity of the overall measure. In some cases, we observed the Kappa paradox, in which the kappa value is low, and the percent agreement is high. This can occur if the variable values are highly imbalanced and observations tend to fall into one particular outcome category. For example, the percent agreement is high (98 percent) for the Documentation of Septic Shock data element. However, 142 of the 146 eligible cases have a value of '2' (No); this imbalance in value distribution potentially contributed to the lower kappa value (0.39) for this variable. Likewise, the data elements Repeat Lactate Level Collection and Repeat Volume Status and Tissue Perfusion Assessment Performed had high percent agreement (over 80 percent) and lower kappa values (less than 0.60). These variables were based on a small number of eligible cases (less than 25 percent of the sample). A frequently cited reference suggests that for medical literature, correlations of 0.7 to 0.9 are high and greater than 0.9 are very high. Of the 24 continuous variables, all correlations were statistically significant and were over 0.80, indicating a high correlation between the CDAC- and CDW-abstracted data. We found acceptable percent agreement between the CDAC and CDW cases for numerator and denominator calculations (over 0.7) , along with kappa values in the moderate range (over 0.6). This supports the validity of the data elements and measure construction.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

57

Median performance score

60

Minimum performance score

0

Maximum performance score

100

Standard deviation of performance scores

21

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

Henry Ford Hospital

Measure Steward Contact Information

Emanuel Rivers
2799 W Grand Blvd
Detriot, MI 48202
erivers1@hfhs.org
(313) 207-1831

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Kimberly Smuk
P.O. Box 2393
Princeton, NJ 08543-2393
ksmuk@mathematica-mpr.com
(202) 838-3721

Secondary Submitter Contact Information

Robert Dickerson
P.O. Box 2393
Princeton, NJ 08543-2393
rdickerson@mathematica-mpr.com
(312) 585-3345

Submitter Comments

This measure is being submitted at the request of CMS.

Please reach out to the Mathematica contacts listed above for questions.

Inpatient Psychiatric Facility Quality Reporting Program

MUC2022-078 Psychiatric Inpatient Experience Measurement

Program

Inpatient Psychiatric Facility Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The measure is a 23-item five-point Likert scale (i.e., "strongly agree, agree, neutral, disagree, strongly disagree" as well as a "does not apply" option) survey to assess the experience of patients who have received inpatient psychiatric services. The survey measures four key domains of patient experience for inpatient psychiatric care settings, including Relationship with the Treatment Team, Nursing Presence, Treatment Effectiveness, and the Healing Environment.

Numerator

The YPIX survey measures are calculated using top-box scoring. The top-box score refers to the percentage of patient respondents that provide the most positive response option Strongly Agree on each given question.

For domain level scores, the top box scores are totaled for all questions within a given domain.

Numerator Exclusions

None

Denominator

The top box denominator is the number of respondents who have answered a given question. For domain level scoring, the question responses are totaled within a given domain.

Denominator Exclusions

Omitted questions or responses of Does Not Apply are excluded from the calculation.

Patients who are younger than age 13.

Patients who are unable to complete the survey due to cognitive or intellectual limitations.

Denominator Exceptions

N/A

State of development

Field (Beta) Testing

State of Development Details

Field Testing

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

Behavioral health

Measure Type

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?

Patient Reported Data and Surveys

If applicable, specify the data source

N/A

Description of parts related to these sources

Patient reported surveys were collected in multiple ways: A paper-based survey was used during early pilot stages. An electronic survey platform (Qualtrics) was used to collect surveys via an iPad. Post-discharge surveys were also sent out via Qualtrics which was integrated with the EHR system.

At what level of analysis was the measure tested?

Facility;Other: Hospital Units

In which setting was this measure tested?

Inpatient psychiatric facility

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

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

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?

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?

Hospital-Consumer Assessment of Healthcare Providers & Systems

https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS/HCAHPS1

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

The measure will be (to our knowledge) the only patient experience survey that has been specifically validated for inpatient psychiatric facilities. The existing HCAHPS measures are not validated for the inpatient psychiatric care setting.

How will this measure add value to the CMS program?

The measure was tested and validated using literature, focus groups, and clinical experts within the inpatient psychiatric care setting. The existing HCAHPS surveys specifically exclude behavioral health from their protocol.

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

Yes (enter here):: Hospital type

Unit specialty / population

Age

Race/Ethnicity

Sex / Gender

Sexuality (Orientation)

Feasibility of Data Elements

No data elements are in defined fields in electronic sources

Feasibility Assessment

Data feasibility was evaluated via field testing using both paper-based and electronic surveys. Surveys were offered and collected prior to patient discharge from an inpatient facility. An additional pilot was performed using email and paper-based surveys that were distributed post-discharge.

Data availability: In-person survey collection sites received survey responses from approximately 50% of total discharges over a two-year period. The post-discharge email and paper-based surveys returned < 3% of the total surveys distributed.

Missing data: Uncollected surveys occurred because a patient refused, or the unit was unable to offer the survey due to operational reasons. Among completed surveys, less than approximately 5% of total items were missing.

Barriers: In-person survey collection requires human resources. We estimate that 30 minutes per day by unit clerical staff were required to locate and distribute the survey prior to discharge. Unexpected discharges with a short lead-time also inhibit survey collection.

Sampling bias: There may be a systematic difference between the patients who elect to complete a survey vs. those who do not. Overall, the magnitude of this bias is arguably less in in-person surveys vs mail/email collection.

Unintended consequences: in-person distribution of surveys creates the potential for gaming or falsification of results.

Feasibility of transferring: Data transfers would occur with third-party vendors in the same way HCAHPS surveys are currently transferred.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Top box performance across all questions at the facility level of analysis was as follows:

Avg: 64%

Median: 64%

Min: 55%

Max: 72%

Std Dev: 6.9%

Top Box performance across all questions at the hospital unit level of analysis was as follows:

Avg: 63%

Median: 62%

Min: 50%

Max: 77%

Std Dev: 8.5%

- See the attached analysis of the performance gap for full question and domain scores.

Evidence of a performance gap is demonstrated by comparing these scores to the national average HCAHPS top box scores. Among the nationally reported HCAHPS questions, the average top box scores range from 54% to 80%, with an overall average of 70%. The current measure's top box results are similar to those from the HCAHPS.

Unintended Consequences

Implementation of a survey tool that is collected prior to patient discharge may have several unintended consequences:

1) There may be a small cost associated with technology or human capital resources;

2) There may be less focus on improving care processes that are not measured by the survey;

3) Pre-discharge surveys are potentially gamed or falsified;

4) Survey results may be subject to misinterpretation due to measurement error or bias.

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

1

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

This systematic review identifies the methodological quality of the psychometric properties of instruments measuring quality and satisfaction with care from the perspective of mental health patients and professionals. In the 34 studies selected, a total of 22 instruments that measure quality and satisfaction with care provided, according to patients and/or professionals, were identified. Most are instruments with sound and contemporary theoretical foundations. They vary to the extent to which they have been used in empirical studies and with respect to the evaluation of their validity and reliability, although five instruments stand out as yielding good to excellent values in quality criteria. The psychometric review found that five of the instruments met valid psychometric criteria. Future reviews

should include an analysis of the usefulness of instruments based on cost-effectiveness, acceptability, and educational impact.

Sanches-Balcells et al, Psychometric properties of instruments measuring quality and satisfaction in mental health: a systematic review, J Adv Nurs. 2018;74:2497-2510.

Source of empirical data

Published, peer-reviewed original research;Internal data analysis

Summarize the empirical data

The purpose of this study was to develop a psychometrically valid survey using rigorous measurement development and validation processes. A 23-item, four-domain survey with two additional open-ended prompts for narrative feedback was developed using exploratory and confirmatory factor analyses on a sample of 2,438 individuals receiving care in inpatient psychiatric units. Factor analyses produced four factors (treatment team relationships, nursing team presence, treatment effectiveness, and healing environment). A thematic analysis was performed based on the open-ended items to elucidate the primary themes of patient feedback. Results from both analyses were used to further inform a conceptual framework for the measurement of patient experience. The Yale Psychiatric Inpatient Experience survey integrates patient experience theory as well as aspects of patient-centered care that are important to psychiatric inpatients.

Limitations include: Opportunities to fully assess construct validity, including discriminant and convergent validity, were limited due to the burdensome nature of administering multiple surveys to an inpatient psychiatric population. Additional research may be useful to further establish internal and construct validity. Additionally, because data were collected for quality improvement purposes, a test-retest procedure was not performed, therefore limiting the conclusiveness of internal validity. Finally, results relied on self-report, which is vulnerable to social response bias. Nonetheless, this was partially mitigated by administering the survey approximately 24 hours prior to discharge.

Klemanski, D. H., Barnes, T., Bautista, C., Tancreti, B., Klink, B., Dix, E. (in press). Development and validation of the Yale Psychiatric Inpatient Experience Survey: A novel measure of patient experience quality improvement. Journal of Patient Experience.

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

1,200,000

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

N/A

Rationale for not using risk adjustment

Not conceptually or empirically indicated (enter here):: Not recommended at this stage of development and implementation process.

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

5

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

4

Meaningful to Patients: Numbers consulted

3535

Meaningful to Patients: Number indicating survey/tool is meaningful

3535

Meaningful to Clinicians: Numbers consulted

15

Meaningful to Clinicians: Number indicating survey/tool is meaningful

15

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

10

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency

Testing methodology and results

Cronbach's Alpha: Full Sample: 0.93; Yale New Haven Hospital: 0.93; St. Raphael's Hospital: 0.94; Lawrence+Memorial Hospital: 0.91; Bridgeport Hospital: 0.91. All items on the measure (full sample and psychiatric facilities) are related to each other.

Burden for Provider: Was a provider workflow analysis conducted?

Yes

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

8

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

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

Guttman Split-Half Coefficient

Random Split-Half Correlation: Sample size

1,648

Random Split-Half Correlation: Statistical result

0.93

Random Split-Half Correlation: Interpretation of results

All of the items on the measure are highly related to each other. Split-half reliability has an acceptable threshold of 0.80 (Parsons, Sam. 2020a. Split half: Robust Estimates of Split Half Reliability. https://doi.org/10.6084/m9.figshare.11956746.v4).

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

25

Face Validity: Result

25

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: Confirmatory Factor Analysis

Sample Size

2204

Statistic Name

Other (enter here):: Root Mean Square Error of Approximation [RMSEA < 0.06 (90% CI: 0.056-0.062)]

Comparative Fit Index [CFI: 0.92]

Tucker Lewis Index [TLI: 0.94]

Statistical Results

0.92

Interpretation of results

CFA results demonstrate an acceptable fit.

RMSEA <0.08 is considered acceptable.

SRMR <0.08 is considered acceptable.

CFI >= 0.9 is considered acceptable.

TLI >0.9 <0.95 is considered acceptable.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

64

Median performance score

64

Minimum performance score

55

Maximum performance score

72

Standard deviation of performance scores

7

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

Yes

Surveys or patient-reported outcome tools

The Yale Psychiatric Inpatient Experience (YPIX) survey is empirically validated with a peer-reviewed manuscript in press.

Section 5: Measure Contact Information

Measure Steward

Yale New Haven Psychiatric Hospital, New Haven, CT, 06519

Measure Steward Contact Information

David Klemanski
184 Liberty Street

New Haven, CT 06519
mailto:david.klemanski@yale.edu
(203) 214-5426

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

1. Post-discharge response rates for psychiatric hospital patients are typically low (e.g., the post-discharge response rate in our pilot sample was < 5% for both paper- and email-based survey collection modalities), we, therefore, recommend that surveys be offered prior to discharge.
2. Although the use of top-box scores is recommended, it may be useful for the distribution of total Likert-scale responses to be made available during initial implementation.

Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program

MUC2022-120 Documentation of Goals of Care Discussions Among Cancer Patients

Program

Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Measuring documentation of goals of care discussions is a critical step toward achieving the outcome of goal concordant care. Oncologists are responsible for ensuring documentation of these discussions. Documentation of goals in structured fields prompts discussions, enhances their quality and efficiency, and promotes accessibility. This measure assesses goals of care discussion documentation among patients with cancer who die while receiving care at the reporting hospital. In this process measure, reported annually, hospitals will report the percent of cancer patients who died during the reporting period and had the patient's goals of care documented prior to death.

Numerator

The numerator for this measure is the number of individual deceased patients (decedents) in the measurement period for whom a Goals of Care (GOC) conversation was documented in a structured field in the medical record. Goals of care reflect the patient's values, preferences, and wishes. The measure requires any documentation in patient goals field(s) (including that the patient opted not to have the discussion) to report a yes for the numerator.

Numerator Exclusions

There are no numerator exclusions.

Denominator

The denominator is the number of patients who died in the measurement period. This population is defined as:

1: Patients with a diagnosis of cancer, AND

2: Had a least 2 eligible contacts at the reporting hospital within the 6 months prior to death. Eligible contacts are inpatient admissions and/or hematology/oncology ambulatory visits at the reporting hospital.

Definitions and codes will be provided in the associated data dictionary (in revision pending testing results).

Denominator Exclusions

Emergency department or observation stays

Denominator Exceptions

None

State of development

Field (Beta) Testing

State of Development Details

This measure was conceptualized in 2020 and endorsed by the Steering Committee for the Improving Goal Concordant Care (IGCC) initiative of the Alliance of Dedicated Cancer Centers (ADCC). Specifications were developed in October 2021-March 2022 by the IGCC Measure Expert Panel, with oversight and endorsement of the IGCC Implementation Workgroup. Alpha testing was performed in April 2022 at 8 hospitals. Initial data collection (beta testing) is occurring in May 2022.

What is the target population of the measure?

Patients who received care at a PPS-Exempt Cancer Hospital in the measurement period

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

Hematology/oncology

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

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?

Facility

In which setting was this measure tested?

PPS-exempt cancer hospital

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

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

None

Section 2: Measure Evidence

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

Web interface

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

The measure was alpha tested in April 2022. Eight hospitals were provided the specifications and performed a feasibility analysis. The findings of this alpha testing were shared with the IGCC Measure Technical Expert Panel. The primary challenges at this phase of developed identified were:

1. For the denominator, reliably identifying patients who died external to the reporting hospital
2. For the numerator, the participating hospitals are at different stages of
    a. provider use of structured documentation to capture patient goals and
    b. ability to generate reports of the presence of documentation in the "patient goals" field of the medical record (i.e., to completely remove the need for manual medical record abstraction).

This is expected to be resolved through continued maturation of the electronic health records and education.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Goals of care discussions can be difficult for patients, families and providers, and substantial published evidence demonstrates that these discussions do not occur/are not documented (see Evidence). Among 10 cancer hospitals participating in testing, goals of care notes were documented for fewer than 40% of all decedents (range: 3% - 100%), including among those that died as inpatients (range: 0% - 100%). Goals of care are distinct from end of life planning, and should occur earlier and amongst a broader patient population. There is clear room for improvement in eliciting and documenting patient goals of care.

Unintended Consequences

None

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

1

Outline the clinical guidelines supporting this measure

Patient-Clinician Communication: American Society of Clinical Oncology Consensus Guideline.

Purpose: To provide guidance to oncology clinicians on how to use effective communication to optimize the patient-clinician relationship, patient and clinician well-being, and family well-being.

Methods: ASCO convened a multidisciplinary panel of medical oncology, psychiatry, nursing, hospice and palliative medicine, communication skills, health disparities, and advocacy experts to produce recommendations. Guideline development involved a systematic review of the literature and a formal consensus process. The systematic review focused on guidelines, systematic reviews and meta-analyses, and randomized controlled trials published from 2006 through October 1, 2016.

Results: The systematic review included 47 publications. With the exception of clinician training in communication skills, evidence for many of the clinical questions was limited. Draft recommendations underwent two rounds of consensus voting before being finalized.

Recommendations: In addition to providing guidance regarding core communication skills and tasks that apply across the continuum of cancer care, recommendations address specific topics, such as discussion of goals of care and prognosis, treatment selection, end-of-life care, facilitating family involvement in care, and clinician training in communication skills. Recommendations are accompanied by suggested strategies for implementation.

Name the guideline developer/entity

American Society of Clinical Oncology (ASCO)

Publication year

2017

Full citation +/- URL

Gilligan, T., Coyle, N., Frankel, R. M., Berry, D. L., Bohlke, K., Epstein, R. M., Finlay, E., Jackson, V.

A., Lathan, C. S., Loprinzi, C. L., Nguyen, L. H., Seigel, C., & Baile, W. F. (2017). Patient-Clinician Communication: American Society of Clinical Oncology Consensus Guideline. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 35(31), 3618-3632. https://doi.org/10.1200/JCO.2017.75.2311

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.

Recommendation 1.4

Clinicians should provide information that is timely and oriented to the patient's concerns and preferences for information. After providing information, clinicians should check for patient understanding and document important discussions in the medical record (Type of recommendation: formal consensus; Strength of recommendation: strong).

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?

Quality of Evidence

High: High confidence that the available evidence reflects the true magnitude and direction of the net effect (e.g., balance of benefits versus harms) and further research is very unlikely to change either the magnitude or direction of this net effect

Intermediate: Intermediate confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research is unlikely to alter the direction of the net effect, however it might alter the magnitude of the net effect.

Low: Low confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research may change the magnitude and/or direction

Insufficient: Evidence is insufficient to discern the true magnitude and direction of the net effect. Further research may better inform the topic. Reliance on consensus opinion of experts may be reasonable to provide guidance on the topic until better evidence

Strength of Recommendation:

Strong: There is high confidence that the recommendation reflects best practice. This is based on:

- strong evidence for a true net effect (e.g., benefits exceed harms);
- consistent results, with no or minor exceptions;
- minor or no concerns about study quality; and/or
- the extent of panelists' agreement.

Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a strong recommendation.

Moderate: There is moderate confidence that the recommendation reflects best practice. This is based on:

- good evidence for a true net effect (e.g., benefits exceed harms);
- consistent results with minor and/or few exceptions;
- minor and/or few concerns about study quality; and/or
- the extent of panelists' agreement.

Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a moderate recommendation.

Weak: There is some confidence that the recommendation offers the best current guidance for practice. This is based on:

- limited evidence for a true net effect (e.g., benefits exceed harms);
- consistent results, but with important exceptions;
- concerns about study quality; and/or
- the extent of panelists' agreement.

Other considerations (discussed in the guideline's literature review and analyses) may also warrant a weak recommendation

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?

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?

Quality of Evidence

High: High confidence that the available evidence reflects the true magnitude and direction of the net effect (e.g., balance of benefits versus harms) and further research is very unlikely to change either the magnitude or direction of this net effect

Intermediate: Intermediate confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research is unlikely to alter the direction of the net effect, however it might alter the magnitude of the net effect.

Low: Low confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research may change the magnitude and/or direction

Insufficient: Evidence is insufficient to discern the true magnitude and direction of the net effect. Further research may better inform the topic. Reliance on consensus opinion of experts may be reasonable to provide guidance on the topic until better evidence

Strength of Recommendation:

Strong: There is high confidence that the recommendation reflects best practice. This is based on:

- strong evidence for a true net effect (e.g., benefits exceed harms);
- consistent results, with no or minor exceptions;
- minor or no concerns about study quality; and/or
- the extent of panelists' agreement.

Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a strong recommendation.

Moderate: There is moderate confidence that the recommendation reflects best practice. This is based on:

- good evidence for a true net effect (e.g., benefits exceed harms);
- consistent results with minor and/or few exceptions;
- minor and/or few concerns about study quality; and/or
- the extent of panelists' agreement.

Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a moderate recommendation.

Weak: There is some confidence that the recommendation offers the best current guidance for practice. This is based on:

- limited evidence for a true net effect (e.g., benefits exceed harms);
- consistent results, but with important exceptions;
- concerns about study quality; and/or
- the extent of panelists' agreement.

Other considerations (discussed in the guideline's literature review and analyses) may also warrant a weak recommendation

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):: Formal consensus

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

Recommendation 1.4

Clinicians should provide information that is timely and oriented to the patient's concerns and preferences for information. After providing information, clinicians should check for patient understanding and document important discussions in the medical record (Type of recommendation: formal consensus; Strength of recommendation: strong).

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

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

0000

Type of Evidence to Support the Measure

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

Not conceptually or empirically indicated (enter here):: Risk adjustment is not needed as goals of care discussions should occur for all patients in the denominator population.

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

8

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

8

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.

11

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

10

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?

10

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

Yes

Section 4: Measure Testing Details

Reliability

No

Reliability: Type of Reliability Testing

N/A

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

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

45

Face Validity: Result

42

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: During alpha testing, 8 hospitals collected and submitted data for the measure denominator and numerator elements. Automated reports reduce reporting burden but may introduce validity and reliability concerns. Sites performed chart reviews to assess reliability.

Sample Size

40

Statistic Name

Percent agreement

Statistical Results

90

Interpretation of results

Identification of deceased patients requires time (for data maturity) and defined processes. During Alpha testing, we compared numbers of identified decedents by quarter to identify the most proximal time period in which hospitals could report the numerator. We selected the most recent time period in which quarterly data reflected normal variation vs artificially low numbers due to incomplete data capture. This measure should be reported with a 6 month delay before the denominator window begins.

Reports that automatically pull data from structured EHR fields are crucial to reducing reporting burden, but can introduce reliability and validity issues. Further, hospitals appropriately customize the exact words and phrases used for structured goals of care documentation. Experts ensured that hospitals that

had goals of care fields could identify them appropriately using the measure instructions. This was found to be the case. Abstractors compared automated reports with actual chart documentation using a defined audit process. Reports correctly identified that there was goals of care field documentation 90% of the time. Results of the audit allowed hospitals to begin addressing EHR report inconsistencies when those were found.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

40

Median performance score

39

Minimum performance score

0

Maximum performance score

100

Standard deviation of performance scores

25

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

Alliance of Dedicated Cancer Centers

Measure Steward Contact Information

Thomas Ross
17703 Bright Wheat Drive
Lithia, FL 33547
snookbeagle@gmail.com
(813) 431-6076

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

Note 1: Mean performance score. Field will only accept one numerical value. 32% (reported on MUC form) of patients who died as inpatients had goals of care note present. For patients who died elsewhere an average of 15% of patients had such documentation. Beta testing is beginning in May 2022.

Note 2: Median performance score: Field will only accept one numerical value. Alpha testing results: goals of care notes were documented for a median of 30% of patients who died as inpatients at the reporting hospital and documented for a median of 13% of patients who died elsewhere.

Beta testing is beginning in May 2022.

Note 3: Minimum performance score: Field will only accept one numerical value. Alpha testing results: minimum performance score of 3% of patients who died as inpatients at the reporting hospital and 6% of patients who died elsewhere. Beta testing is beginning in May 2022.

Note 4: Maximum performance score: Field will only accept one numerical value. Alpha testing results: maximum performance score of 76% of patients who died as inpatients at the reporting hospital and 35% of patients who died elsewhere. Beta testing is beginning in May 2022.

Rural Emergency Hospital Quality Reporting Program (REHQRP)

MUC2022-039 Median Time from emergency department (ED) Arrival to ED Departure for Discharged ED Patients

Program

Rural Emergency Hospital Quality Reporting Program (REHQRP)

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Median time from ED arrival to time of departure from the ED for patients discharged from the ED. The measure is calculated using chart abstracted data, on a rolling quarterly basis, and is publicly reported in aggregate for one calendar year. The measure has been publicly reported since 2013 as part of the ED Throughput measure set of the CMS Hospital Outpatient Quality Reporting (OQR) Program.

Numerator

N/A: This is a continuous variable measure.

Numerator Exclusions

N/A: This is a continuous variable measure.

Denominator

Continuous Variable measure: Time (in minutes) from ED arrival to ED departure for patients discharged from the ED.

Denominator Exclusions

Patients who expired in the ED, left against medical advice (AMA), or whose discharge was not documented or unable to be determined (UTD) are excluded from the target population.

Denominator Exceptions

N/A: This is a continuous variable measure.

State of development

Field (Beta) Testing

State of Development Details

The Median Time from ED Arrival to ED Departure for Discharged Patients measure has been in the hospital Outpatient Quality Reporting (OQR) program for at least 10 years and has been publicly reported since 2015. The most recent testing for this measure, performed in 2020, included empirical reliability testing and critical data element validity testing consistent with the NQF testing requirements at the time, which did not require validity testing for both critical data elements and empiric validity testing of the measure score. We are submitting this existing and tested measure based on a request from CMS that it be added to the new outpatient Rural Emergency Hospital (REH) program. Although the most recent level of testing does not meet the current MUC list definition for a fully developed measure, this measure has been previously tested, is currently in use, and is being publicly reported,

which reflects that it is fully developed. This measure is recommended by CCSQ leadership for adoption in the REH program based upon previous reliability and validity testing and its long-term utility in the OQR program, all of which speaks to its validity.

What is the target population of the measure?

All payer, children, dual eligible beneficiaries, elderly, individuals with multiple chronic conditions, populations at risk, veterans, women

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

Emergency medicine

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; 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?

Facility

In which setting was this measure tested?

Emergency department

Multiple Scores

No

What one healthcare domain applies to this measure?

Seamless Care Coordination

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

N/A

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

Yes

CMIT ID

00930-C-HOQR

Alternate Measure ID

OP18

What is the endorsement status of the measure?

Endorsement Removed

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

0496

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?

2011

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

MUC167-2011

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

2012, Hospital

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

2012, Hospital Quality Reporting Program

What was the MAP recommendation in this year?

2012, Hospital Quality Reporting Program, Support

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

N/A

MAP report page number being referenced for this year:

2012, page 89

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

2011-present

What other federal programs are currently using this measure?

Hospital Outpatient Quality Reporting Program

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?

Other: manual abstraction

Stratification

Yes (enter here):: Overall rate: The overall rate includes all eligible patients.

Reporting rate: The reporting rate includes cases from the overall rate that are not included in the psychiatric/mental health rate or transfer patient rate.

Psychiatric/mental health rate: The psychiatric/mental health rate includes cases from the overall rate for which the principal diagnosis is captured in the psychiatric/mental health code set.

Transfer patient rate: The transfer patient rate includes cases from the overall rate for which the discharge code indicates that the patient was transferred to a facility that is an acute care facility for inpatient care of the general population or a facility operated by the Department of Defense or the Department of Veteran's Affairs.

Feasibility of Data Elements

ALL data elements are in defined fields in electronic sources

Feasibility Assessment

Median time from ED arrival to time of departure is a chart abstracted measure that has been in the Hospital Outpatient Quality Reporting Program (OQR) Program since 2011 and is publicly reported on Care Compare. Based on the volume and content of question received from abstractors, the information is available in medical records and can be consistently located. Most questions are related to the presence of conflicting documentation and unique facility-specific or patient-specific workflows.

Nine expert work group (EWG) members, with backgrounds in healthcare administration, management, and clinical expertise in emergency medicine, pediatric emergency medicine, and clinical pharmacy, provided feedback on the feasibility of this measure through an online survey. Most respondents agreed or strongly agreed that the practical aspects of reporting median time from ED arrival to time of departure as a chart abstracted measure do not place undue burden on hospitals for its data. However, one respondent commented that the degree of burden may vary depending on the programming structure of different electronic health records (EHRs). Most respondents also indicated that the data elements are currently available in an electronic health record EHR structured field. Overall, the

respondents generally supported the feasibility of measuring median time from ED arrival to time of departure.

Method of Measure Calculation

Claims;Manual abstraction

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Analysis of facility-level data from the Hospital Compare downloadable files indicates that there is variation in the median time from ED arrival to time of departure. During the January 2014 to December 2014 data collection periods, median facility-level throughput times ranged from 46 minutes to 424 minutes, with a median of 140 minutes. During the January 2016 to December 2016 data collection periods, median facility-level throughput times ranged from 45 minutes to 440 minutes, with a median of 136 minutes.

During the January 2014 to December 2016 data collection periods, there is documentation of substantial variation in facility performance. The interquartile range is consistently wide, ranging from 51 minutes in 2014 to 53 minutes in 2016. Additionally, the maximum time for ED discharge increased between 2014 and 2016 from 424 to 440 minutes. While median performance is improving, decreasing from 140 minutes in 2014 to 136 minutes in 2016, there is ongoing opportunity for improvement in performance at the facility level.

Unintended Consequences

Measure testing did not identify any unintended consequences. Similarly, no evidence of unintended consequences to individuals or populations has been reported by external stakeholders since its 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

Summarize the empirical data

see Empirical_data_Summary.docx

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

2495927

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

Not conceptually or empirically indicated (enter here):: As a process-of-care measure, timely discharge from the ED should not be influenced by sociodemographic factors, doing so would potentially mask important inequities in care delivery. Variation across patient populations is reflective of differences in the quality of care provided to the disparate patient population included in the effective sample.

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

Other (enter here): hierarchical linear model (HLM)

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

Intraclass Correlation Coefficient (ICC)

Other: Sample size

3,749

Other: Statistical result

0.869 - 0.872

Other: Interpretation of results

Calculated using an HLM model, the ICC of all samples and measure strata indicate that variance due to error does not contribute significantly to variation in performance scores, demonstrating strong measure reliability. The results of this test indicate that the measure is able to identify true differences in performance between facilities.

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

9

Face Validity: Result

7

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between other gold standard and manual reviewer

Sample Size

13,187

Statistic Name

Pearson correlation coefficient

Statistical Results

1.0

Interpretation of results

Results of the quantitative and qualitative analysis are positive and support the conclusion that the measure and its calculation are valid representations of facility performance. There was almost perfect agreement between facility and auditor abstraction of data elements. For data elements arrival time, ED departure date, and ED departure time, all estimated Pearson correlation coefficient values were equal to 1.0 and were statistically significant ($p<0.001$). This suggests strong validity for the critical data elements of the measure, as currently specified.

Measure performance – Type of Score

Continuous Variable – Median

Measure Performance Score Interpretation

Lower score is better

Mean performance score

141.7

Median performance score

136

Minimum performance score

45

Maximum performance score

440

Standard deviation of performance scores

42.1

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

Melissa Hager
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244
Melissa.Hager@cms.hhs.gov
(410) 786-6278

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Kingsley Weaver
Mathematica
111 E Wacker Dr
Chicago, IL 60601
kweaver@mathematica-mpr.com
617-583-1956

Secondary Submitter Contact Information

Robert Dickerson
P.O. Box 2393
Princeton, NJ 08543-2393
rdickerson@mathematica-mpr.com
(312) 585-3345

Submitter Comments

There should be two attachments, "OP-18 MIF.pdf" (5/16/2022) and "Empirical_data_Summary.docx" (5/17/2022). I am unable to see them, but Battelle confirmed that they are uploaded.

MUC2022-066 Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy

Program

Rural Emergency Hospital Quality Reporting Program (REHQRP)

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Facility-level risk-standardized rate of acute, unplanned hospital visits within 7 days of a colonoscopy procedure performed at a Rural Emergency Hospital among Medicare Fee-For-Service (FFS) patients aged 65 years and older. An unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission.

Numerator

The outcome for this measure is all-cause, unplanned hospital visits within 7 days of an outpatient colonoscopy performed at a Rural Emergency Hospital. We define a hospital visit as any emergency department (ED) visit, observation stay, or unplanned inpatient admission.

Numerator Exclusions

N/A

Denominator

Outpatient colonoscopies performed at Rural Emergency Hospitals for Medicare FFS patients aged 65 years and older. Specifically: The measure includes patients undergoing routine (not high-risk) colonoscopies, identified using HCPCS codes and CPT codes. Qualifying colonoscopy procedures are not included in the measure if they are concurrently billed with a high-risk colonoscopy procedure code. The measure includes patients with continuous enrollment in Medicare FFS Parts A and B in the 12 months prior to the procedure.

Denominator Exclusions

1) Colonoscopies for patients who lack continuous enrollment in Medicare FFS Parts A and B in the 1 month after the procedure. 2) Colonoscopies that occur concurrently with high-risk upper gastrointestinal (GI) endoscopy procedures. 3) Colonoscopies for patients with a history of inflammatory bowel disease (IBD) or diagnosis of IBD at time of index colonoscopy or on a subsequent hospital visit outcome claim. 4) Colonoscopies for patients with a history of diverticulitis or diagnosis of diverticulitis at time of index colonoscopy or on a subsequent hospital visit outcome claim. 5) Colonoscopies followed by a subsequent outpatient colonoscopy procedure within 7 days. 6) Colonoscopies that occur on the same day and at the same hospital as an ED visit that is billed on a different claim than the index colonoscopy, unless the ED visit has a diagnosis indicative of a complication of care.6 7) Colonoscopies that are billed on the same hospital claim as an ED visit and that occur on the same calendar day, unless the ED visit has a diagnosis indicative of a complication of care. 8) Colonoscopies that are billed on the same hospital outpatient claim and that occur after the ED visit. 9) Colonoscopies that are billed on the same hospital outpatient claim as an observation stay.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Elderly, Populations at risk

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

Gastroenterology

Measure Type

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?

Administrative Data (non-claims);Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

We use Medicare FFS claims to identify colonoscopies performed in the outpatient setting and subsequent hospital visits, as well as CMS enrollment and demographic data. Patient history is also assessed using claims data collected in the 12 months prior to the colonoscopy procedure.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Hospital outpatient department (HOD)

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

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

N/A

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

Yes

CMIT ID

02086-C-HOQR

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Endorsed

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

2539

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

2020

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

2024

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?

2013

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

XDEMA

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

Hospital, 2014

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

2014, HOQR

What was the MAP recommendation in this year?

Conditional Support

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

N/A

MAP report page number being referenced for this year:

Page 184, MAP 2014 Recommendations on Measures for More Than 20 Federal Programs, January 2014

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

In the Calendar Year (CY) 2017 Hospital Outpatient Prospective Payment System (OPPS) final rule, CMS added the colonoscopy measure for implementation in the Outpatient Quality Reporting (OQR) program.

What other federal programs are currently using this measure?

Ambulatory Surgical Center Quality Reporting Program;Hospital Outpatient Quality Reporting Program

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?

ASC-12, OP-32

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

This measure would capture outpatient colonoscopies performed in REHs, whereas the OP 32 and ASC 12 measures capture outpatient colonoscopies performed in HOPDs and ASCs.

How will this measure add value to the CMS program?

This measure is currently in use in HOQR and ASCQR to measure unplanned hospital visits for colonoscopies performed in HOPDs and ASCs. The inclusion of this measure in the REH program will allow CMS to report quality information on colonoscopies performed in facilities that acquire the newly established REH designation.

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

Stratification

No

Feasibility of Data Elements

ALL data elements are in defined fields in electronic sources

Feasibility Assessment

This measure uses beneficiary enrollment and claims data. Administrative claims data used in this measure are routinely captured as part of the billing process and there are no fees associated with collecting the data.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

We include performance data for all HOPDs captured by the colonoscopy measure, which includes but is not limited to entities that will convert to REHs.

For the performance period between January 1, 2016-December 31, 2018, risk-standardized hospital visit rates (RSHVRs) per 1000 colonoscopies, for all facilities (n = 4034) were as follows: minimum, 11.67,

10th percentile, 14.92, 25th percentile, 15.76, 50th percentile (median), 16.38; 75th percentile, 17.10, 90th percentile, 18.10, maximum, 24.27; mean (SD), 16.47 (1.32).

The distribution of measure scores indicates that there is substantial variation in performance among HOPDs.

Among HOPDs, the median RSHVR is 16.4 hospital visits per 1,000 colonoscopies, which indicates that patients undergoing colonoscopy at a facility performing at the median are expected to have an ED visit, observation stay, or admission to the hospital within 7 days 1.64% of the time.

The 10th and 90th percentiles (14.9 and 18.1 hospital visits per 1,000 colonoscopies, respectively) represent meaningful deviations from the median: a facility performing at the 10th percentile is performing about 9% better than an average performer, and a facility performing at the 90th percentile is performing about 11% worse than an average performer.

Furthermore, the best performing facilities (11.7 hospital visits per 1,000 colonoscopies) are performing 29% better than the median performer, while the worst (24.3 hospital visits per 1,000 colonoscopies) are performing 48% worse than the median performer.

Unintended Consequences

We have encountered no unexpected findings during implementation, including unintended impacts on patients.

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

Internal data analysis

Summarize the empirical data

Please refer to attachment 2 for details.

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?

No

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?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment;Severity of Illness

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

01/2018- 12/2019 and 07/2020- 12/2020 Predictive ability, % (lowest decile, highest decile): 0.69-4.65 c-statistic: 0.681

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

Facility-level signal to noise reliability. Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling, reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

Signal-to-Noise: Sample size

3583

Signal-to-Noise: Statistical result

.782

Signal-to-Noise: Interpretation of results

High, based on similar measures that have been NQF endorsed.

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

Noting that our validity testing was not a specific statistical test and cannot be accommodated by the fields that we are restricted to. Please see the appendix for the information.

Empiric Validity: Sample size

000000

Empiric Validity: Statistical result

00000

Empiric Validity: Methods and findings

Please see Appendix for details.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

14

Face Validity: Result

12

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

Lower score is better

Mean performance score

0000

Median performance score

0000

Minimum performance score

0000

Maximum performance score

0000

Standard deviation of performance scores

0000

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

Janis Grady
7500 Security Boulevard
Baltimore, MD 21244
janis.grady@cms.hhs.gov
(410) 786-7217

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

While a TEP was not convened for the use of this measure in the REH program, stakeholders were involved in the development of the HOPD version of this measure. The OP-32 (Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy) measure was developed consistent with CMS's quality measure development guidance. The CORE project team, a multidisciplinary team of clinicians, health services researchers and statisticians, was supported and informed by surgical consultants and a national technical expert panel (TEP) consisting of patients, surgeons, methodologists, researchers, and providers. CMS also held a public comment period, soliciting stakeholder input on the measure methodology, and publicly posted a summary of the comments received as well as the responses.

MUC2022-067 Risk-standardized hospital visits within 7 days after hospital outpatient surgery

Program

Rural Emergency Hospital Quality Reporting Program (REHQRP)

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Facility-level risk-standardized rate of acute, unplanned hospital visits within 7 days of an outpatient surgical procedure performed at a Rural Emergency Hospital among Medicare Fee-For-Service (FFS) patients aged 65 years and older. An unplanned hospital visit is defined as an emergency department (ED) visit, observation stay, or unplanned inpatient admission.

Numerator

The outcome is all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the surgery or 2) an unplanned hospital visit post discharge (emergency department [ED] visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the outpatient surgical procedure performed at a Rural Emergency Hospital. If more than one unplanned hospital visit occurs, only the first hospital visit within the outcome timeframe is counted in the outcome.

Numerator Exclusions

N/A

Denominator

Outpatient same-day surgeries performed at Rural Emergency Hospital for Medicare FFS patients aged 65 years and older.

Denominator Exclusions

1. Surgeries performed at Rural Emergency Hospitals, for patients without continuous enrollment in Medicare FFS Parts A and B in the 7 days after the surgery.
2. Surgeries performed at Rural Emergency Hospitals, for patients who have an ED visit on the same day but billed on a separate claim unless the ED visit has a diagnosis indicative of a complication of care.
3. Surgeries performed at Rural Emergency Hospitals, that are billed on the same hospital claim as an emergency department (ED) visit and that occur on the same calendar day unless the ED visit has a diagnosis indicative of a complication of care.
4. Surgeries performed at Rural Emergency Hospitals, that are billed on the same hospital outpatient claim and that occur after the ED visit.
5. Surgeries performed at Rural Emergency Hospitals, that are billed on the same outpatient claim as an observation stay.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

The target population is Medicare FFS patients aged 65 years and older undergoing same-day surgery (those that do not typically require an overnight stay) at Rural Emergency Hospitals.

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

General surgery

Measure Type

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?

Claims Data

If applicable, specify the data source

N/A

Description of parts related to these sources

We use Medicare FFS claims to identify surgeries performed in the outpatient setting and subsequent hospital visits, as well as CMS enrollment and demographic data. Patient history is also assessed using claims data collected in the 12 months prior to the eligible same-day surgery.

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Hospital outpatient department (HOD)

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

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

N/A

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

Yes

CMIT ID

02930-C-HOQR

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Endorsed

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

2687

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

No

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

Other: The HOQR version of the measure (OP 36) is NQF-endorsed. To implement the measure in the REH program, the measure outcome remains the same, but the cohort is outpatient surgical procedures performed in REHs as opposed to in all HOPDs.

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

The HOQR version of the measure (OP 36) is NQF-endorsed. To implement the measure in the REH program, the measure outcome remains the same, but the cohort is outpatient surgical procedures performed in REHs as opposed to in all HOPDs.

If endorsed: Year of most recent CDP endorsement

2020

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

2023

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?

2015

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

MUC15-982

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

Hospital 2016

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

HOQR 2016

What was the MAP recommendation in this year?

2016; HOQR, SUPPORT.

MAP supported the measure, citing the vital importance of measures that help facilities reduce unnecessary hospital visits.

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

N/A

MAP report page number being referenced for this year:

MAP 2016 Considerations for Implementing Measures in Federal Programs: Hospitals, PAGE 10.

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

In the Calendar Year (CY) 2017 Hospital Outpatient Prospective Payment System (OPPS) final rule, CMS added this surgery measure for CY 2020 payment determination in the Outpatient Quality Reporting (OQR) program.

What other federal programs are currently using this measure?

Hospital Outpatient Quality Reporting Program

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?

OP-36

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

This measure would capture risk-standardized rate of acute, unplanned hospital visits within 7 days of an outpatient surgical procedure performed in REHs, whereas the OP-36 measure captures risk-standardized rate of acute, unplanned hospital visits within 7 days of a procedure performed in HOPDs.

How will this measure add value to the CMS program?

This measure is currently in use in HOQR (OP-36) to measure risk-standardized rate of acute, unplanned hospital visits within 7 days of a procedure performed in HOPDs. The inclusion of this measure in the REH program will allow CMS to report quality information on risk-standardized rate of acute, unplanned hospital visits within 7 days of an outpatient surgical procedure performed in facilities that acquire the newly established REH designation.

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

Stratification

No

Feasibility of Data Elements

ALL data elements are in defined fields in electronic sources

Feasibility Assessment

This measure uses beneficiary enrollment and claims data. Administrative claims data used in this measure are routinely captured as part of the billing process and there are no fees associated with collecting the data.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

We have performance information showing a quality gap for all HOPDs with qualifying procedures; this includes facilities but is not limited to facilities that may opt to become Rural Emergency Hospitals. Overall, our results suggest that there is substantial need to both reduce the expected rate and the variation in rates across HOPDs, and that this improvement goal is achievable. We characterize the degree of variation by: 1) Providing the median odds ratio (MOR) [1]. The MOR represents the median increase in odds of a hospital visit if a procedure on a single patient was performed at a higher risk HOPD compared to a lower risk HOPD. It is calculated by taking all possible combinations of HOPDs, always comparing the higher risk HOPD to the lower risk HOPD. The MOR is interpreted as a traditional odds ratio would be. The median odds ratio is 1.28 which indicates that a patient has a 28% increase in the odds of a hospital visit if the same procedure was performed at higher risk HOPD compared to a lower risk HOPD indicating the impact of quality on the outcome rate is substantial. 2) Reporting the distribution of the RSHVR. Of the 3,974 facilities (representing data from January 1, 2018-December 31, 2018), the range of RSHVRs was 0.54-2.39 (IQR 0.93-1.07). The range of performance on the HOPD Surgery measure demonstrates that there is a significant quality gap. Specifically, the best-performing HOPD (RSHVR of 0.54) is performing 46% better than average, whereas the worst-performing HOPD (RSHVR of 2.39) is performing 139% worse than the average. Furthermore, our outlier analysis identified about 300 or about 8 percent of HOPDs as outliers (3.77% significantly better and 3.98% significantly worse than expected). Note that the that average performer refers to an HOPD with the same case and service-line mix, performing at the average. 3) Presenting performance categories. Because the measure score is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate to determine if a HOPD is performing better than, worse than, or no different than expected. A HOPD is considered as better than expected if their entire confidence interval falls below 1, and considered worse if the entire confidence interval falls above 1. They are considered no different if the confidence interval overlaps 1. A total of 150 facilities (3.77%) performed Better than Expected, 2,671 facilities (67.21%) performed No Different than Expected, and the remaining 158 facilities (3.98%) performed Worse than Expected.

Unintended Consequences

During public comment period when the measure was first proposed in 2017 OPPS rule, concern was raised about the potential unintended consequence of providers avoiding certain patients and procedures depending on the inclusion criteria and robustness of the risk adjustment. But the measure was adopted, implemented, and CMS also routinely monitors for unintended consequences of quality measures.

We have encountered no unexpected findings during implementation, including unintended impacts on patients.

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

Internal data analysis

Summarize the empirical data

Please see attachment 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?

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?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions

Patient-level demographics: please select all that apply:

Age

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

Case-Mix Adjustment

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

Predictive ability, % (lowest decile, highest decile): 1.85-14.12 c-statistic: 0.677

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

Facility-level signal to noise reliability. Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling, reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

Signal-to-Noise: Sample size

3974

Signal-to-Noise: Statistical result

.759

Signal-to-Noise: Interpretation of results

The median signal-to-noise reliability score is sufficiently high for both all facilities, and facilities with at least 30 procedures (the public reporting cutoff). See appendix for additional details.

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

Pleases see Appendix for details. Noting that our validity testing was not a specific statistical test and cannot be accommodated by the fields that we are restricted to. Please see the appendix for the information.

Empiric Validity: Sample size

00000

Empiric Validity: Statistical result

00000

Empiric Validity: Methods and findings

Please see appendix for details

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

13

Face Validity: Result

13

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

Lower score is better

Mean performance score

0000

Median performance score

0000

Minimum performance score

0000

Maximum performance score

0000

Standard deviation of performance scores

0000

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

CMS

Measure Steward Contact Information

Janis Grady
7500 Security Boulevard
Baltimore, MD 21244
janis.grady@cms.hhs.gov
(410) 786-7217

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

While a TEP was not convened for the use of this measure in the REH program, stakeholders were involved in the development of the HOPD version of this measure. The OP-36 (Risk-standardized hospital visits within 7 days after hospital outpatient surgery) measure was developed consistent with CMS quality measure development guidance. The CORE project team, a multidisciplinary team of clinicians, health services researchers and statisticians, was supported and informed by surgical consultants and a national technical expert panel (TEP) consisting of patients, surgeons, methodologists, researchers, and providers. CMS also held a public comment period, soliciting stakeholder input on the measure methodology, and publicly posted a summary of the comments received as well as the responses.

MUC2022-081 Abdomen Computed Tomography (CT) Use of Contrast Material

Program

Rural Emergency Hospital Quality Reporting Program (REHQRP)

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

This measure calculates the percentage of abdomen studies that are performed with and without contrast out of all abdomen studies performed (those with contrast, those without contrast, and those with both).

Numerator

The number of abdomen and abdominopelvic CT studies with and without contrast (combined studies).

Numerator Exclusions

N/A

Denominator

The number of abdomen CT studies performed (with contrast, without contrast, or both with and without contrast).

Denominator Exclusions

Cases are excluded from the denominator if the CT scan was performed for an adrenal mass, bladder cancer, hematuria, infection of the kidney, jaundice, a liver lesion, pancreatic cancer, non-traumatic aortic disease, urinary system disease, or other unspecified disorder of the kidney or ureter.

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

Radiation oncology

Measure Type

Efficiency

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

If applicable, specify the data source

N/A

Description of parts related to these sources

The data are calculated only for facilities paid through the OPPS for abdomen CT studies performed in the hospital outpatient setting. Data from the hospital outpatient file is used to determine beneficiary inclusion (for example, a CT abdomen study performed at the hospital outpatient department) and exclusion (that is, diagnoses of adrenal mass, hematuria, infections of the kidney, jaundice, liver lesions, malignancies of the pancreas, diseases of urinary system, pancreatic disorders, and unspecified disorder of kidney or ureter diagnosis codes).

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Hospital outpatient department (HOD)

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)?

Yes

CMIT ID

02599

Alternate Measure ID

OP10

What is the endorsement status of the measure?

Failed Endorsement

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

0000

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

2011-present

What other federal programs are currently using this measure?

Hospital Outpatient Quality Reporting Program

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 identical to the Abdomen CT Use of Contrast Material implemented in CMS's Hospital Outpatient Quality Reporting program since 2011.

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

The measure specifications will remain similar to those for Abdomen CT Use of Contrast Material implemented in CMS's Hospital Outpatient Quality Reporting program. However, the application of this measure to the Rural Emergency Health program will allow CMS to measure a new care setting where these imaging studies are performed.

How will this measure add value to the CMS program?

A CT abdomen study is a very common imaging procedure in the Medicare population. Implementing the measure for Rural Emergency Hospitals promotes the use of CT abdomen imaging aligned with current clinical guidance, while avoiding the potentially harmful effects of unnecessary radiation and contrast exposure. A CT study performed with and without contrast doubles the radiation dose to the beneficiary and exposes the beneficiary to the potential harmful side effects of the contrast material itself. Reducing the unnecessary use of combined CT abdomen studies-defined as those that are performed both with and without contrast agents for the evaluation of solid organs and body cavities-represents an important opportunity to improve practice and patient safety.

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

Stratification

No

Feasibility of Data Elements

ALL data elements are in defined fields in electronic sources

Feasibility Assessment

The measure is calculated using data from final claims that facilities submit for Medicare beneficiaries enrolled in fee for service Medicare. The codes included in the measure specifications appear on the claims used to calculate the measure.

Method of Measure Calculation

Claims

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Since the first public reporting of the measure in 2011 for the HOQR program, there have been large improvements in facility performance at Hospital Outpatient Departmentsâ?the measure median dropped from 9.5% in 2011 to 1.4% in 2021, with a much larger downward trend in performance for facilities falling in or above the 75th percentile originally. Overall, the downward trend suggests that the quality of the performance of abdomen CT studies improved nationally across each score stratum during public reporting. Though we cannot calculate the performance gap for REHs, we anticipate the measure would similarly help facilities address any variation in performance.

In public reporting year 2020, rural, small (0-50 beds), and government-owned facilities account for a disproportionally high percentage of outlier facilities (45.1%, 37.6%, and 0.9%, respectively), indicating any opportunity for performance improvement.

Unintended Consequences

None have been identified by the measure developer, Technical Expert Panel, or other stakeholders.

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

21

Outline the clinical guidelines supporting this measure

See attachment, section title "Evidence - Field 102 Response"

Name the guideline developer/entity

American College of Radiology (ACR)

Publication year

2021

Full citation +/- URL

0000

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.

N/A

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

Other (enter here):: 0000

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?

Other (enter here):: 0000

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

Other (enter here):: 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?

Other (enter here):: 0000

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

N/A

Number of systematic reviews that inform this measure concept

0000

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

See attachment, section title "Evidence - Field 102 Response"

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies)

Summarize the empirical data

0000 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;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 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?

Yes

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

8

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

8

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

No

Reliability: Type of Reliability Testing

N/A

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

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

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: Proportion reported as a percentage

Measure Performance Score Interpretation

Lower score is better

Mean performance score

0000

Median performance score

0000

Minimum performance score

0000

Maximum performance score

0000

Standard deviation of performance scores

0000

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

CMS

Measure Steward Contact Information

Janis Grady
7500 Security Boulevard
Baltimore, MD 21244
janis.grady@cms.hhs.gov
(410) 786-7217

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor

New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

Validity, reliability, and measure performance results, as well as the measure concept, and evidence, are based on the measure as implemented in the Hospital Outpatient Quality Reporting program. There are no results currently for Rural Emergency Hospitals.

There were no patients on the Technical Expert Panel for this measure during development prior to its implementation in 2011. However, the current Technical Expert Panel convened to support measure reevaluation includes three patients.

Cross-Program Measures

These measures were submitted to multiple federal programs.

MUC2022-024 Hospital Harm - Acute Kidney Injury

Program

Hospital Inpatient Quality Reporting Program; Medicare Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (CAHs)

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The proportion of inpatient hospitalizations for patients 18 years of age or older who have an acute kidney injury (stage 2 or greater) that occurred during the encounter as evidenced by a substantial increase in serum creatinine value, or by the initiation of kidney dialysis (continuous renal replacement therapy [CRRT], hemodialysis or peritoneal dialysis).

Numerator

Inpatient hospitalizations for patients who develop AKI (stage 2 or greater) during the encounter, as evidenced by:

- A subsequent increase in the serum creatinine value at least 2 times higher than the lowest serum creatinine value, and the increased value is greater than the highest sex-specific normal value for serum creatinine; or
- Kidney dialysis (hemodialysis or peritoneal dialysis) initiated 48 hours or more after the start of the encounter.

Numerator Exclusions

None

Denominator

Inpatient hospitalizations for patients 18 years of age or older at the start of the encounter without a diagnosis of obstetrics, with a length of stay of 48 hours or longer who had at least one serum creatinine value after 48 hours from the start of the encounter.

Denominator Exclusions

Inpatient hospitalizations for patients with an increase in serum creatinine value of at least 0.3 mg/dL between the index serum creatinine and a subsequent serum creatinine taken within 48 hours of the encounter start.

Inpatient hospitalizations for patients with an eGFR value of <60 mL/min within 48 hours of the encounter start.

Inpatient hospitalizations for patients who have less than two serum creatinine results within 48 hours of the encounter start.

Inpatient hospitalizations for patients who have kidney dialysis (CRRT, hemodialysis or peritoneal dialysis) initiated within 48 hours of the encounter start.

Inpatient hospitalizations for patients with at least one specified diagnosis present on admission that puts them at extremely high risk for AKI.

Inpatient hospitalizations for patients with at least one specified procedure during the encounter that puts them at extremely high risk for AKI.

The "index" serum creatinine is defined as the lowest serum creatinine within the first 24 hours of encounter start. If there are no serum creatinine values within the first 24 hours, then the index is the first serum creatinine within the first 48 hours of the start of the encounter.

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

Internal medicine

Measure Type

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 Health Record

If applicable, specify the data source

N/A

Description of parts related to these sources

The measure uses structured fields within the EHR to calculate the initial population, denominator exclusion, denominator, numerator, and variables used in the risk adjustment such as:

- Admission, Discharge, Transfer
- Encounter Information
- Procedures
- Laboratory Test Results

- Assessments
- Diagnosis and Present on Admission Indication

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

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

N/A

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

Yes

CMIT ID

05904-E-HIQR

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

2022

Digital Measure Information

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

Yes

If eCQM, enter Measure Authoring Tool (MAT) number

832

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?

No

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?

AKI is encompassed in the Agency for Healthcare Research and Quality's (AHRQ) Patient Safety Indicator (PSI) 90 composite, under the provider-level PSI 10: Postoperative Acute Kidney Injury Requiring Dialysis Rate (CMIT Ref No. 05021) measure. PSI 10 is used to measure Postoperative physiologic and metabolic derangements (secondary diagnosis) or acute renal failures (secondary diagnosis) with dialysis per 1,000 elective surgical discharges for patients ages 18 years and older. PSI 10 is available on Hospital Compare.

Reference:

AHRQ. (2020). Patient Safety Indicator 10 (PSI 10) Postoperative Acute Kidney Injury Requiring Dialysis Rate ICD-10-CM/PCS Specification v2020

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

PSI 10 measures how often hospitalized patients had renal failure requiring dialysis after having an operation. Additionally, PSI 10 utilizes claims data and is not National Quality Forum (NQF) endorsed (though the composite PSI 90 (CMIT Ref No.03282/05537), of which it is a component, is endorsed).

In comparison, this measure measures how often AKI occurs in the inpatient hospital setting and will be developed as an eCQM.

Reference:

AHRQ. (2020). Patient Safety Indicator 10 (PSI 10) Postoperative Acute Kidney Injury Requiring Dialysis Rate ICD-10-CM/PCS Specification v2020

How will this measure add value to the CMS program?

Although there are many occurrences of AKI in hospital settings, many of which are preventable, there is currently no measure in a Centers for Medicare & Medicaid Services (CMS) quality reporting program or public reporting that quantifies how AKI occurs in hospitalized patients.

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

To better understand if critical data elements used in the measure are available in a structured format and if the form in which they exist aligns with measure intent, we designed a web-based questionnaire and distributed the survey to 34 hospitals (17 Meditech and 17 Cerner). The survey began with an inquiry into the measure's critical data elements (concepts) and ended with questions on the overall measure. The goal of the survey was to determine, within each hospital's EHR system, if critical data elements are:

- readily available in a structured format,
- from an authoritative source and/or highly likely to be correct,
- coded in a nationally accepted terminology standard or can be mapped to that terminology standard, and
- routinely collected as part of clinical care and require no or limited additional data entry from a clinician or other providers, and no EHR interface changes needed.

All 34 hospitals confirmed that the following data elements are captured in the EHR in a structured and codified manner: Birthdate, Ethnicity, Payer, Race, Sex. Encounter Inpatient, Emergency Department Visit, Observation Services, Serum Creatinine Lab Test., and Procedure Performed (all denominator exclusion procedures).

For the data element, Procedure, Performed: Hospital based Dialysis Services, we encountered some inconsistencies in workflow and ability to capture the data across the 34 sites. Only 29 of the sites offer dialysis services onsite, and most of these sites utilize a contracted service provider to complete the procedure. Roughly 24% of the sites offering dialysis services only maintain dialysis performed documentation in unstructured fields, which prompted us to look at the feasibility of using a union of dialysis performed and dialysis ordered. All 29 sites reported the ability to structurally capture Procedure, Ordered: Hospital based Dialysis Services with some variability in provider workflows; some hospitals extract orders through the contracted dialysis provider while others extract orders through the EHR system.

For the data element Diagnosis, used in the measure exclusions, there were challenges in retrieving the attribute "present on admission" (POA) from 15 sites. The documentation was available in the EHR; however, the location of that data was not accessible for extract at the time of initial testing, although it became available later. Since POA status is a mandated element in hospital billing, we do not have concerns about the availability, accuracy and use of standards. Technical glitches and/or workflow modifications may be required for some organizations to ensure that all POA indicators pushed into the EHR from other systems can be extracted.

Upon collecting responses, we held debrief meetings with participants to resolve ambiguities. We then translated final responses to numeric values used in the NQF scorecard.

Method of Measure Calculation

eCQM

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

This safety eCQM captures the number of patients who had acute kidney injury (AKI), as evidenced by a substantial increase in serum creatinine, or the initiation of renal dialysis (hemodialysis or peritoneal dialysis) during the hospitalization. Published literature suggests that the incidence of AKI in general hospitalized patients is 10% - 20%, in ICU patients ranges from 10-20%, and in cardiac surgery patients ranges from 30%-50% (Thongprayoon, 2020).

Using the EHR data from 20 hospitals and in year 2020, we found that hospital-level measure performance rates ranged from 0.76% (for every 1,000 qualified hospital admissions there are 7.6 inpatient encounters where patients suffered AKI) to 4.43% (for every 1,000 qualified hospital admissions there are 44 inpatient encounters where patients suffered AKI), with a system-wide, weighted average rate equal to 1.52%.

While AKI may be due to natural progression of underlying illness or a complication of a necessary treatment such as chemotherapy, a proportion of AKI cases are preventable and treatable (KDIGO,

2012). The Kidney Disease: Improving Global Outcomes (KDIGO) guidelines suggest careful management of hemodynamic status, fluids, and vasoactive medications for the prevention of AKI (Wilson et al., 2015). Both worsening renal function and injury requiring dialysis have lasting negative impact including loss of kidney function, uremic complications, and symptoms associated with drug toxicity and volume overload (Hoste & De Corte., 2011; Levey & James., 2017; Liborio et. al., 2015). Literature also suggests early AKI treatment such as nephrotoxic avoidance, drug dose adjustment, and attention to fluid balance are also effective preventive measures (Perazella, 2012; Onuigbo et al., 2017). The KDIGO guidelines also suggest careful management of hemodynamic status, fluids, and vasoactive medications for the prevention of AKI (KDIGO, 2012). Clinical consensus continues to support the KDIGO recommendations following a recent conference review (Ostermann et al., 2019).

This measure will fill a gap in measurement and provide incentives for hospital quality improvement, as there is no current inpatient AKI measure in a CMS program. Systematically measuring the rates of AKI in the hospital setting will provide hospitals with a reliable and timely assessment and will allow for hospitals to improve quality and reduce AKI harm rates.

References:

Hoste, E., & De Corte, W. (2011). Clinical consequences of acute kidney injury. Contributions to nephrology, 174, 56-64.

Kidney Disease: Improving Global Outcomes (KDIGO). (2012). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney international, Suppl. 2, 1-138.

Levey, A. S., & James, M. T. (2017). Acute Kidney Injury. Annals of internal medicine, 167(9), ITC66-ITC80.

Liborio, A. B., Leite, T. T., Neves, F. M., Teles, F., & Bezerra, C. T. (2015). AKI complications in critically ill patients: association with mortality rates and RRT. Clinical journal of the American Society of Nephrology: CJASN, 10(1), 21-28.

Onuigbo, M. A., Samuel, E., & Agbasi, N. (2017). Hospital-acquired nephrotoxic exposures in the precipitation of acute kidney injury - A case series analysis and a call for more preventative nephrology practices. J Nephropharmacol, 6(2), 90-97.

Ostermann, M., Bellomo, R., Burdmann, E. A., Doi, K., Endre, Z. H., Goldstein, S. L., Kane-Gill, S. L., Liu, K. D., Prowle, J. R., Shaw, A. D., Srisawat, N., Cheung, M., Jadoul, M., Winkelmayer, W. C., Kellum, J. A., & Conference Participants (2020). Controversies in acute kidney injury: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Conference. Kidney international, 98(2), 294-309.

Perazella M. A. (2012). Drug use and nephrotoxicity in the intensive care unit. Kidney international, 81(12), 1172-1178.

Thongprayoon, C., Hansrivijit, P., Kovvuru, K., Kanduri, S. R., Torres-Ortiz, A., Acharya, P., Gonzalez-Suarez, M. L., Kaewput, W., Bathini, T., & Cheungpasitporn, W. (2020). Diagnostics, Risk Factors, Treatment and Outcomes of Acute Kidney Injury in a New Paradigm. Journal of clinical medicine, 9(4), 1104. https://doi.org/10.3390/jcm9041104

Unintended Consequences

We did not identify any unintended consequences during eCQM development or testing. However, CMS is committed to monitoring this eCQM's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, and other negative unintended consequences for patients.

It is possible that by measuring AKI in the hospital setting, some hospital clinicians may be less likely to provide aminoglycoside antibiotics, nonsteroidal anti-inflammatory drugs (NSAIDs), or other medications that are thought to contribute to the occurrence of AKI in some patients. Increased incentives to avoid these medications could lead to higher pain burden in some patients, although alternative medications are always available.

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

1

Outline the clinical guidelines supporting this measure

One evidence-based guideline directly supports the measure as follows:

Kidney Disease: Improving Global Outcomes (KDIGO). (2012). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney international, Suppl. 2, 1-138.

The Kidney Disease: Improving Global Outcomes (KDIGO) Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease provides the most modern and accepted definition and staging system for AKI (KDIGO, 2012). This guideline is evidence based and recommends that serum creatinine is an accepted proxy for acute kidney disease (KDIGO, 2012).

Realizing that there is an increasing prevalence of acute (and chronic) kidney disease worldwide and that the complications and problems of patients with kidney disease are universal, KDIGO, a nonprofit foundation, was established in 2003 "to improve the care and outcomes of kidney disease patients worldwide through promoting coordination, collaboration, and integration of initiatives to develop and implement clinical practice guidelines," (Ekinoyan et al., 2004). The Board of Directors of KDIGO quickly realized that there is room for improving international cooperation in the development, dissemination, and implementation of clinical practice guidelines in the field of AKI. At its meeting in December of 2006, the KDIGO Board of Directors determined that the topic of AKI meets the criteria for developing clinical practice guidelines. These criteria were formulated as follows:

- AKI is common.
- AKI imposes a heavy burden of illness (morbidity and

mortality).

- The cost per person of managing AKI is high.
- AKI is amenable to early detection and potential prevention.
- There is considerable variability in practice to prevent,

diagnose, treat, and achieve outcomes of AKI.

- Clinical practice guidelines in the field have the potential to reduce variations, improve outcomes, and reduce costs.
- Formal guidelines do not exist on this topic.

This guideline was published in 2012 and uses the GRADE system to rate the strength of evidence and the strength of recommendations. In all, there were only 11 (18%) recommendations in this guideline for which the overall quality of evidence was graded 'A,' whereas 20 (32.8%) were graded 'B,' 23 (37.7%) were graded 'C,' and 7 (11.5%) were graded 'D.' Although there are reasons other than quality of evidence to make a grade 1 or 2 recommendation, in general, there is a correlation between the quality of overall evidence and the strength of the recommendation. Thus, there were 22 (36.1%) recommendations graded '1' and 39 (63.9%) graded '2.' There were 9 (14.8%) recommendations graded '1A,' 10 (16.4%) were '1B,' 3 (4.9%) were '1C,' and 0 (0%) were '1D.' There were 2 (3.3%) graded '2A,' 10 (16.4%) were '2B,' 20 (32.8%) were '2C,' and 7 (11.5%) were '2D.' There were 26 (29.9%) statements that were not graded.

These recommendations have been assessed for quality and the guidelines have been deemed appropriate to meet our criteria.

KDIGO defines AKI Stage 1 as the sudden decrease (in 48 h) of renal function, defined by an increase in absolute serum creatinine (SCr) of at least 26.5 Âµmol/L (0.3 mg/dL) from baseline or by a percentage increase in SCr greater or equal to greater than 1.5-1.9 times the baseline value, or by a decrease in the urinary output (UO) (documented oliguria less than 0.5 mL/kg/h for more than 6 h). Stage 2 is defined as an increase of absolute serum creatinine SCr greater than 2-2.9 times the baseline value, or by a decrease in UO less than 0.5 mL/kg/h for more than 12 h.

References (alphabetical order):

Eknoyan, G., Lameire, N., Barsoum, R., Eckardt, K. U., Levin, A., Levin, N., Locatelli, F., MacLeod, A., Vanholder, R., Walker, R., & Wang, H. (2004). The burden of kidney disease: improving global outcomes. Kidney international, 66(4), 1310-1314.

Kidney Disease: Improving Global Outcomes (KDIGO). (2012). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney international, Suppl. 2, 1-138.

Name the guideline developer/entity

Kidney Disease Improving Global Outcomes (KDIGO)

Publication year

2012

Full citation +/- URL

Kidney Disease: Improving Global Outcomes (KDIGO). (2012). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney international, Suppl. 2, 1-138.

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.

KDIGO offers clinical practice guidelines for the prevention and treatment of AKI. The guideline statement that most closely aligns with the measure concept is:

3.1.1: In the absence of hemorrhagic shock, we suggest using isotonic crystalloids rather than colloids (albumin or starches) as initial management for expansion of intravascular volume in patients at risk for acute kidney injury or with acute kidney injury.

Strength of recommendation: Level 2

Strength of evidence: Grade B.

For information on additional KDIGO clinical practice guideline statements which closely align with this measure concept, see Table 3 (Section 1) of the 2022 MUC List Attachment - Acute Kidney Injury.

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?

The strength of each KDIGO guideline recommendation is indicated as Level 1, Level 2, or Not Graded using the GRADE system:

Level 1 - "We recommend." This recommendation can be evaluated as a candidate for developing a policy or a performance measure.

Level 2 - "We suggest." This recommendation is likely to require substantial debate and involvement of stakeholders before policy can be determined.

Not graded- "Not Graded" was used, typically, to provide guidance based on common sense or where the topic does not allow adequate application of evidence. The most common examples include recommendations regarding monitoring intervals, counseling, and referral to other clinical specialists. The ungraded recommendations are generally written as simple declarative statements but are not meant to be interpreted as being stronger recommendations than Level 1 or 2 recommendations.

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?

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?

Within each recommendation, the quality of the supporting evidence is shown as A, B, C, or D. The KDIGO guidelines use the GRADE system.

Grade A- High quality of evidence. We are confident that the true effect lies close to that of the estimate of the effect.

Grade B- Moderate quality of evidence. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Grade C- Low quality of evidence. The true effect may be substantially different from the estimate of the effect.

Grade D- Very low quality of evidence. The estimate of effect is very uncertain, and often will be far from the truth.

GRADE, Grades of Recommendation Assessment, Development, and Evaluation. GRADE System for grading quality of evidence. Modified from Uhlig 2006, GRADE Working Group 2004, and Kunz and Farquhar 2004.

References:

Uhlig, K., Macleod, A., Craig, J., Lau, J., Levey, A. S., Levin, A., Moist, L., Steinberg, E., Walker, R., Wanner, C., Lameire, N., & Eknoyan, G. (2006). Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney international, 70(12), 2058-2065. https://doi.org/10.1038/sj.ki.5001875

GRADE Working Group (2004). Grading quality of evidence and strength of recommendations. BMJ (Clinical research ed.), 328(7454), 1490. https://doi.org/10.1136/bmj.328.7454.1490

Kunz K, Farquhar C (2004) . Grading and the GRADE instrument. Second Guidelines International Network Conference: Evidence in Action, Wellington, NZ.

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?

Moderate or similar

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

KDIGO offers clinical practice guidelines for the prevention and treatment of AKI. The guideline statement that most closely aligns with the measure concept is:

3.1.1: In the absence of hemorrhagic shock, we suggest using isotonic crystalloids rather than colloids (albumin or starches) as initial management for expansion of intravascular volume in patients at risk for acute kidney injury or with acute kidney injury.

Strength of recommendation: Level 2

Strength of evidence: Grade B.

For information on additional KDIGO clinical practice guideline statements which closely align with this measure concept, see Table 3 (Section 1) of the 2022 MUC List Attachment - Acute Kidney Injury.

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

On average, AKI affects up to 10% of hospitalized patients, which is comparable to the rates of severe sepsis and acute lung injury (McCoy et al., 2010; Hoste & Schurgers., 2008; Chertow et al., 2005; Perzazella, 2012). AKI requiring dialyses and less severe AKI affects approximately 200-300 and 2,000-3,000 per million population per year, respectively (Chertow et al., 2005). Accurately monitoring the rate at which AKI occurs in the hospital setting allows for hospitals to improve quality and reduce AKI harm rates.

This AKI eCQM uses a seven-day rolling window to examine a rise in serum creatinine to 2.0 times or greater is based on the KDIGO definition established in the 2012 KDIGO AKI clinical practice guidelines (KDIGO, 2012). A rise in serum creatinine is associated with an increased risk of mortality, even if values are within a 'normal' range. (KDIGO, 2012).

One enhancement that was made to this measure specification was based on recent evidence on the calculation of the glomerular filtration rate (GFR). The GFR is estimated from the serum creatine levels from serum concentrations of endogenous filtration markers such as creatinine or cystatin C (the equation is known as eGFR) (Inker et al., 2021). Based on a study from the Chronic Kidney Disease and Epidemiology Collaboration (CKD-EPI), the study found that eGFR equations that incorporate creatinine and cystatin C but omit race are more accurate and led to smaller differences between Black participants and non-Black participants than new equations without race with either creatinine or cystatin C alone (Inker et al., 2021). As a result, a new race-neutral eGFR equation that that measures serum creatinine or cystatin C incorporate age, sex, and race to estimate measured GFR has been developed and is recommended by the Task Force from the National Kidney Foundation and American Society of Nephrology (Inker et al., 2021; Diao et al., 2021; Delgado et al., 2021; Delgado et al., 2022). It was recommended by the task force to use within the measure a CKD-EPI creatinine equation refit without the race variable. This functionality has been available to all laboratories in the United States (Delgado et al., 2021; Delgado et al.,2022), and has acceptable performance characteristics and potential consequences that do not disproportionately affect any one group of individuals. This eCQM will utilize this change in algorithm.

References (Alphabetical):

Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V., & Bates, D. W. (2005). Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. Journal of the American Society of Nephrology : JASN, 16(11), 3365-3370.

Delgado, C., Baweja, M., Crews, D. C., Eneanya, N. D., Gadegbeku, C. A., Inker, L. A., Mendu, M. L., Miller, W. G., Moxey-Mims, M. M., Roberts, G. V., St Peter, W. L., Warfield, C., & Powe, N. R. (2021). A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. Journal of the American Society of Nephrology : JASN, 32(12), 2994-3015. Advance online publication.

Delgado, C., Baweja, M., Crews, D. C., Eneanya, N. D., Gadegbeku, C. A., Inker, L. A., Mendu, M. L., Miller, W. G., Moxey-Mims, M. M., Roberts, G. V., St Peter, W. L., Warfield, C., & Powe, N. R. (2022). A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. American journal of kidney diseases : the official journal of the National Kidney Foundation, 79(2), 268-288.e1.

Diao, J. A., Inker, L. A., Levey, A. S., Tighiouart, H., Powe, N. R., & Manrai, A. K. (2021). In Search of a Better Equation - Performance and Equity in Estimates of Kidney Function. The New England journal of medicine, 384(5), 396-399.

Inker, L. A., Eneanya, N. D., Coresh, J., Tighiouart, H., Wang, D., Sang, Y., Crews, D. C., Doria, A., Estrella, M. M., Froissart, M., Grams, M. E., Greene, T., Grubb, A., Gudnason, V., Gutierrez, O. M., Kalil, R., Karger, A. B., Mauer, M., Navis, G., Nelson, R. G., ... Chronic Kidney Disease Epidemiology Collaboration (2021). New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. The New England journal of medicine, 385(19), 1737-1749.

Hoste, E. A., & Schurgers, M. (2008). Epidemiology of acute kidney injury: how big is the problem?. Critical care medicine, 36(4 Suppl), S146-S151.

Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group (2012). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney inter., Suppl. 2013; 3: 1-150. McCoy, A. B., Waitman, L. R., Gadd, C. S., Danciu, I., Smith, J. P., Lewis, J. B., Schildcrout, J. S., & Peterson, J. F. (2010). A computerized provider order entry intervention for medication safety during acute kidney injury: a quality improvement report. American journal of kidney diseases : the official journal of the National Kidney Foundation, 56(5), 832-841.

Perazella M. A. (2012). Drug use and nephrotoxicity in the intensive care unit. Kidney international, 81(12), 1172-1178.

For information on clinical practice guidelines that support the empirical evidence, see Table 1-3 (Section 1) in the 2022 MUC List Attachment - Acute Kidney Injury.

Name evidence type

Conclusions from the KDIGO AKI conference (2019)

Summarize the evidence

KDIGO held a conference in 2019 to determine best practices for prevention of AKI and areas of uncertainty in treating AKI; review key relevant literature published since the 2012 KDIGO AKI guideline;; identify new topics or issues to be revisited for the next iteration of the KDIGO AKI guideline; and outline research needed to improve AKI management (Ostermann et al., 2020). The effectiveness of the 2012 KDIGO recommendations in preventing AKI was also noted to have been confirmed in small single-

center randomized controlled trials (RCTs), such as the Prevention of AKI (PrevAKI) and the Biomarker Guided Intervention for Prevention of AKI (BigpAK) studies (Meersh 2017, Gocze 2018). In addition, results of RCTs have provided new data relevant to several facets of preventing and managing AKI, including early resuscitation, fluid therapy, prevention of contrast-associated AKI, and timing of acute renal replacement therapy (RRT) (Kellum 2016, Nijssen 2017, Self 2018, Semler 2018, Zarbock 2016, Gaudry 2016, Barbar 2018). Much of the focus of the conference focused on prevention and early treatment to prevent progression of AKI. The overall findings from the conference suggest that since publication of the KDIGO guidelines in 2012, consensus of clinical opinion and several recent trials support fluid management and drug stewardship to reduce the occurrence of AKI.

Ostermann, M., Bellomo, R., Burdmann, E. A., Doi, K., Endre, Z. H., Goldstein, S. L., Kane-Gill, S. L., Liu, K. D., Prowle, J. R., Shaw, A. D., Srisawat, N., Cheung, M., Jadoul, M., Winkelmayer, W. C., Kellum, J. A., & Conference Participants (2020). Controversies in acute kidney injury: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Conference. Kidney international, 98(2), 294-309.

Meersch, M., Schmidt, C., Hoffmeier, A., Van Aken, H., Wempe, C., Gerss, J., & Zarbock, A. (2017). Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. Intensive care medicine, 43(11), 1551-1561.

Gocze, I., Jauch, D., Gotz, M., Kennedy, P., Jung, B., Zeman, F., Gnewuch, C., Graf, B. M., Gnann, W., Banas, B., Bein, T., Schlitt, H. J., & Bergler, T. (2018). Biomarker-guided Intervention to Prevent Acute Kidney Injury After Major Surgery: The Prospective Randomized BigpAK Study. Annals of surgery, 267(6), 1013-1020.

Kellum, J. A., Chawla, L. S., Keener, C., Singbartl, K., Palevsky, P. M., Pike, F. L., Yealy, D. M., Huang, D. T., Angus, D. C., & ProCESS and ProGReSS-AKI Investigators (2016). The Effects of Alternative Resuscitation Strategies on Acute Kidney Injury in Patients with Septic Shock. American journal of respiratory and critical care medicine, 193(3), 281-287.

Nijssen, E. C., Rennenberg, R. J., Nelemans, P. J., Essers, B. A., Janssen, M. M., Vermeeren, M. A., Ommen, V. V., & Wildberger, J. E. (2017). Prophylactic hydration to protect renal function from intravascular iodinated contrast material in patients at high risk of contrast-induced nephropathy (AMACING): a prospective, randomised, phase 3, controlled, open-label, non-inferiority trial. Lancet (London, England), 389(10076), 1312-1322.

Self, W. H., Semler, M. W., Wanderer, J. P., Wang, L., Byrne, D. W., Collins, S. P., Slovis, C. M., Lindsell, C. J., Ehrenfeld, J. M., Siew, E. D., Shaw, A. D., Bernard, G. R., Rice, T. W., & SALT-ED Investigators (2018). Balanced Crystalloids versus Saline in Noncritically Ill Adults. The New England journal of medicine, 378(9), 819-828.

Zarbock, A., Kellum, J. A., Schmidt, C., Van Aken, H., Wempe, C., Pavenstdt, H., Boanta, A., Ger, J., & Meersch, M. (2016). Effect of Early vs Delayed Initiation of Renal Replacement Therapy on Mortality in Critically Ill Patients With Acute Kidney Injury: The ELAIN Randomized Clinical Trial. JAMA, 315(20), 2190-2199.

Gaudry, S., Hajage, D., Schortgen, F., Martin-Lefevre, L., Pons, B., Boulet, E., Boyer, A., Chevrel, G., Lerolle, N., Carpentier, D., de Prost, N., Lautrette, A., Bretagnol, A., Mayaux, J., Nseir, S., Megarbane, B., Thirion, M., Forel, J., Maizel, J., Yonis, H., Markowicz, P., Thiery, G., Tubach, F., Ricard, J. and Dreyfuss, D., 2016. Initiation Strategies for Renal-Replacement Therapy in the Intensive Care Unit. New England Journal of Medicine, 375(2), pp.122-133.

Barbar, S. D., Clere-Jehl, R., Bourredjem, A., Hernu, R., Montini, F., Bruyre, R., Lebert, C., Boh, J., Badie, J., Eraldi, J. P., Rigaud, J. P., Levy, B., Siami, S., Louis, G., Bouadma, L., Constantin, J. M., Mercier, E., Klouche, K., du Cheyron, D., Piton, G., ... IDEAL-ICU Trial Investigators and the CRICS TRIGGERSEP Network (2018). Timing of Renal-Replacement Therapy in Patients with Acute Kidney Injury and Sepsis. The New England journal of medicine, 379(15), 1431-1442.

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

15,910,905

Type of Evidence to Support the Measure

Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines;Empirical data;Other (enter here):: Consensus conference & recent randomized controlled trials (RCTs) published since the most recent guideline

Is the measure risk adjusted?

Yes

Risk adjustment variables

Patient-level demographics ;Patient-level health status & clinical conditions

Patient-level demographics: please select all that apply:

Age;Sex

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

Case-Mix Adjustment; Severity of Illness

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 two metrics to evaluate the risk model's performance: 1) receiver operating characteristics (ROC) curve and the area under the curve (AUROC or C-statistic) and 2) precision-recall (PR) curve and the area under the curve (AUPRC). (Section 2) of the 2022 MUC List Attachment - Acute Kidney Injury shows ROC and PR curves for the baseline risk model on the holdout set using coefficient estimates derived from the training set. We note that data points in the holdout set are not used for model training or parameter tuning and hence, they provide an unbiased assessment of the model's performance. AUROC and AUPRC are shown on the bottom of the graph, and the dashed lines (reference lines) serve as a benchmark classifier, implying what the model performance would be for a completely uninformative model. Testing results suggest strong model performance. In particular, C-statistic is larger than 0.8, which is a benchmark frequently cited for demonstrating excellent model performance. Similarly, AUPRC indicates that the baseline risk model is close to 15 times better than a random prediction. We caution against the interpretation of AUPRC based on its absolute value, as by construction AUPRC is affected by the base rate and inversely related to data imbalance. That is, as data become more balanced and as base rate rises, the absolute value of AUPRC rises. To alleviate the concern that our testing sample may not be large enough to yield a reliable performance metric, we conducted a simulation exercise. Specifically, we generated 100 random holdout sets and calculated 100 AUROC and AUPRC using coefficient estimates previously derived from the training set, and the Exhibit 3 (Section 2) of the 2022 MUC List Attachment - Acute Kidney Injury shows that the odds of strong model performance occurring by chance are very small. We provide further testing results about the risk adjustment model in Section 2 of the 2022 MUC List Attachment - Acute Kidney Injury.

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

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.

10

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

10

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?

20

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;Random Split-Half Correlation

Signal-to-Noise: Name of statistic

Adams' signal-to-noise ratio (SNR)

Signal-to-Noise: Sample size

20

Signal-to-Noise: Statistical result

0.91

Signal-to-Noise: Interpretation of results

Evaluating against the conventional standards, test statistics showed strong measure score-level reliability and suggested that the measure, as currently specified, can distinguish performance of one hospital from another. A sizable portion of the variability across hospitals is attributable to the real difference in quality of care.

Random Split-Half Correlation: Name of statistic

Intra-class correlation coefficient (ICC) via the split-half sample approach

Random Split-Half Correlation: Sample size

20

Random Split-Half Correlation: Statistical result

0.79

Random Split-Half Correlation: Interpretation of results

Evaluating against the conventional standards, test statistics showed strong measure score-level reliability and suggested that the measure, as currently specified, can distinguish performance of one hospital from another. A sizable portion of the variability across hospitals is attributable to the real difference in quality of care.

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

Known groups validity, which focuses on the measure's ability to differentiate between groups of measured entities that are known to differ on their underlying latent construct. Known groups validity can be viewed as measuring the relative scores between groups - in this case, comparing teaching hospitals with non-teaching hospitals. This relative difference is expressed as a risk ratio, or the risk of AKI at teaching hospitals divided by the risk of AKI at non-teaching hospitals.

Empiric Validity: Sample size

20

Empiric Validity: Statistical result

0.73

Empiric Validity: Methods and findings

Known groups validity focuses on the measure's ability to differentiate between groups of measured entities that are known to differ on their underlying latent construct. In terms of hospital quality and patient safety, prior research has shown several known groups that are identifiable using information collected:

- Hospital teaching/academic status
- Hospital bed size (<25, 25-99, 100-199, 200-499, and >499)
- Hospital urban/rural location

Prior to analyses, we hypothesized that the rate of AKI will be lower at teaching, large-sized, and urban hospitals (in general, better resourced hospitals) than non-teaching, small-sized, and rural hospitals, respectively.

Testing results showed that teaching hospitals performed (27%) better than non-teaching hospitals, with the average risk-adjusted measure performance rate equal to 1.37 and 1.87 per 100 qualified admissions, respectively (risk ratio = 0.73). Urban hospitals performed (22%) better than rural hospitals, with the average risk-adjusted measure performance rate equal to 1.63 and 2.08 per 100 qualified admissions, respectively (risk ratio = 0.78). Large-sized hospitals performed better than small-sized hospitals. Note that lower measure rate denotes better quality of care, as the measured outcome is a harm event.

We provide detailed testing results in Table 23 (Section 4) of the 2022 MUC List Attachment - Acute Kidney Injury.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

11

Face Validity: Result

11

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between eCQM and manual reviewer

Sample Size

727

Statistic Name

Other (enter here):: Percent Agreement, Kappa, Positive Predictive Value, and Sensitivity (results listed in order below)

Statistical Results

0.95,0.92,0.91,0.79

Interpretation of results

First, test results indicated strong measure data element-level reliability and validity. The relatively lower sensitivity value was due to cases meeting denominator exclusion per clinical abstractors that weren't correctly classified by the EHR. We investigated the root cause and found that the discrepancy was caused by the EHR-exported data showing blank POA information, even though valid values (e.g., Yes) were visible upon inspection of the EHR. The POA information in the affected sites exists in an external data environment and only partial POA information (e.g., ICD-10-CM diagnoses) flows into the EHR in a structured format. As a result, the technical difficulty led to disagreements between the EHR and clinical abstractors. However, we note that POA status is a mandatory data element in hospital billing and hence, we have no concerns about this issue in the longer term. As a follow-up, we conducted further analyses with the affected sites and confirmed the electronic retrievability of POA status for all their inpatient discharges in year 2020.

Across the 20 test sites, measure denominator ranged from 151 to 7,948 qualified inpatient encounters and measure rate ranged from a low of 0.76 to a high of 4.43 per 100 qualified inpatient encounters. The wide variability indicates ample room for quality improvement in hospital inpatient setting. Exhibit 4 (Section 4) of the 2022 MUC List Attachment - Acute Kidney Injury translates the distribution of observed measure rate and its 95% confidence interval into a histogram and further displays the system-wide, weighted average measure rate (green dashed horizontal line). Testing data showed that several hospitals' performance rates are consistently below the overall mean while a few others are above that mean.

Comparing to the variability in measure performance rates, variation in Kappa, PPV, and sensitivity across test sites is more limited. The minimum Kappa equals 0.92, but all other kappas were 1.0, which denotes perfect concordance. The minimum PPV equals 0.91 and the second smallest PPV equals 0.96. The median and mode of PPV are both 1.0. The minimum sensitivity equals 0.79 and the second smallest sensitivity equals 0.96. The median and mode of sensitivity are both 1.0. Limited variation

does not impact interpretation above; on the contrary, it substantiates the conclusion that the measure has strong data element-level reliability and validity.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Lower score is better

Mean performance score

1.52

Median performance score

1.36

Minimum performance score

0.76

Maximum performance score

4.43

Standard deviation of performance scores

1.01

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 (CMS)

Measure Steward Contact Information

Donta Henson
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

Hannah Klein
1400 Crystal Drive
10th floor
Arlington, VA 22202
hklein@air.org
(206) 939-4978

Secondary Submitter Contact Information

Anna Michie
1400 Crystal Drive
10th floor
Arlington, VA 22202
amichie@air.org
(443) 259-5180

Submitter Comments

N/A

MUC2022-026 Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA PRO-PM) in the HOPD or ASC Setting

Program

Ambulatory Surgical Center Quality Reporting Program; Hospital Outpatient Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The measure will estimate a facility-level risk-standardized improvement rate for patient-reported outcomes (PROs) following elective primary THA/TKA for Medicare fee-for-service (FFS) patients 65 years of age or older. Substantial clinical benefit (SCB) improvement will be measured by the change in score on the joint-specific patient-reported outcome measure (PROM) instruments, measuring hip or knee pain and functioning, from the preoperative assessment (data collected 90 to 0 days before surgery) to the postoperative assessment (data collected 275 to 425 days following surgery).

Numerator

The numerator is the risk-adjusted proportion of patients undergoing an elective primary THA/TKA who meet or exceed a SCB threshold of improvement between preoperative and postoperative assessments on joint-specific PROMs as follows:

- For THA patients, meeting or exceeding a 22-point increase in score on the Hip dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR)1,) and
- For TKA patients, meeting or exceeding a 20-point increase in score on the Knee injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR)2.).

Numerator Exclusions

None

Denominator

The cohort (target population) includes Medicare FFS patients 65 years of age and older undergoing elective primary THA/TKA procedures in an HOPD or ASC setting. The measure includes patients who are enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the procedure, who are alive 10 months after the procedure and who do not have more than two THA or TKA procedure codes on the same claim. The cohort does not include patients with revision THAs/TKAs, and bone metastases. The rationale for each is outlined below: A concurrent partial hip or knee arthroplasty procedure Rationale: Partial arthroplasty procedures are primarily done for hip and knee fractures and are typically performed on patients who are older, frailer, and have more comorbid conditions. A concurrent revision, resurfacing, or implanted device/prosthesis removal procedure Rationale: Revision procedures may be performed at a disproportionately small number of hospitals and are associated with higher mortality, complication, and readmission rates. Resurfacing procedures are a different type of procedure involving only the joints articular surface and are typically performed on younger, healthier

patients. Elective procedures performed on patients undergoing removal of implanted device/prostheses procedures may be more complicated. Malignant neoplasm of the pelvis, sacrum, coccyx, lower limbs, or bone/bone marrow or a disseminated malignant neoplasm coded in the principal discharge diagnosis field on the index admission claim Rationale: Patients with these malignant neoplasms are at increased risk for complication, and the procedure may not be elective.

Denominator Exclusions

The measure has two denominator exclusions, listed below.

1. Staged Procedures

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

2. Patients who die within 10 months of the procedure

This exclusion is to remove patients who die prior to the postoperative window and are not "available" to provide PROs during postop assessment. With the updated postoperative window, this should align.

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

Orthopedic surgery

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?

Administrative Data (non-claims);Claims Data;Electronic Health Record;Patient Reported Data and Surveys

If applicable, specify the data source

Other:

Description of parts related to these sources

Patient Reported Data: Numerator: HOOS, JR; KOOS, JR

Risk adjustment: Mental health from VR-12 or Global PROMIS, Health literacy (SILS), Total painful joint count from Pain in Non-Operative Joint, Back Pain at preoperative assessment

Claims: Identifying the cohort and exclusions

Risk adjustment Electronic health record data (noting that this data is submitted by facilities with the patient-reported data), BMI, Narcotic use (>=90 days)

Administrative: Death data, Gender

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

Hospital outpatient department (HOD);Hospital inpatient acute care facility

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

N/A

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

Yes

CMIT ID

03198

Alternate Measure ID

NQF ID 3559

What is the endorsement status of the measure?

Endorsed

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

3559

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

No

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

Other: Other

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

The hospital level measure is NQF endorsed (NQF 3559), and the clinician level measure (NQF 3639) has been recommended for endorsement by NQF's Surgery Standing Committee (the measure is currently in public comment). This current MUC submission is for a

If endorsed: Year of most recent CDP endorsement

2020

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

2023

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?

MUC20-0003

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

Hospital, 2021

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

HIQR Hospital Inpatient Quality Reporting Program

What was the MAP recommendation in this year?

Supported the measure

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

N/A

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, page 18

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

Inpatient Quality Reporting (proposed for 2025 voluntary reporting)

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program

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?

Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary

Total Hip and/or Total Knee Arthroplasty (THA/TKA) (NQF 3559)

Clinician-Level and Clinician Group-Level Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA and TKA) Patient-Reported Outcome-Based Performance Measure (PRO-PM) (NQF 3639)

Average change in functional status following total knee replacement surgery (NQF 2653)

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

NQF 3559 and NQF 3639: Outcomes from procedures performed in the outpatient setting are currently not captured by these measures for Medicare Fee For Service patients. Other than attribution, these measures are exactly the same as the measure at the HOPD/ASC level.

NQF 2653: This measure does capture outpatient procedures but is reported at the clinician/clinician group level and therefore does not hold facilities accountable for their performance on this measure.

The measure proposed in this submission differs from NQF 2653 in important ways out lined below:

1. This PRO-PM reflects outcomes for both THA and TKA recipients (rather than TKA recipients only), allowing for measurement of a greater number of patients and providers to provide CMS with broader influence on quality improvement. This approach aligns with the typical provision of orthopedic care, delivered to patients undergoing THA/TKA procedures by the same providers and staff.
2. This PRO-PM assesses improvement in patient-reported pain and function using a binary outcome that elucidates for providers and patients the risk-adjusted proportion of patients with and without improvement (a clear, understandable metric that patients support); this is preferable to measuring an average change score, as NQF #2653 does, which cannot distinguish between providers with mostly average outcomes from providers whose patients either did very well or very poorly. In addition, using a SCB to define the measure outcome ensures that the measure does not penalize clinicians who operate on those patients with the worst baseline pain and function (often those with higher social risk or non-white race).
3. NQF Measure #2653 uses an average change score adjusted for the baseline PROM score this fundamentally equates to measuring post-operative PROM scores, which may incentivize surgeons to operate on those with the least severe symptoms at baseline and potentially avoid patients with the most severe pain and functional limitations at baseline. This would likely result in worsening disparities over time.
4. This PRO-PM uses a more robust and stakeholder-driven risk model and methodology to address non-response bias, anticipated to produce a measure with greater face validity with stakeholders. Specifically, this measure includes key clinical risk variables for a PRO-PM identified by clinical experts and supported by orthopedic professional societies, such as health literacy, back pain, and contralateral leg pain. These ensure accurate assessment of the index THA/TKA procedure and account for concomitant comorbidities such as chronic back or contralateral joint disease that can interfere with PROM interpretation. In addition, this measure accounts for non-response bias. We have seen no evidence of NQF #2653 analytically addressing non-response bias. Non-response bias is a critical potential threat to the validity of PRO-PMs and failure to account for it may lead to worsening disparities.

How will this measure add value to the CMS program?

CMS has recently allowed elective primary THA and TKA procedures to be performed in both the HOPD and ASC setting; CMS removed both procedures from the inpatient-only (IPO) list and added them to the ASC Covered Procedures List (CPL): TKA was removed from the IPO for CY 2018 and added to the ASC CPL in CY2020; THA was removed from the IPO for CY 2020 and added to the ASC CPL in CY2021.

Between April 1, 2018 and March 30, 2021, there were about 330,000 THA and TKA procedures performed in the outpatient (HOPD setting). In addition, after the onset of the COVID-19 pandemic in 2020, outpatient procedures for both THA and TKA procedures outnumbered inpatient procedures; as of September 2021, inpatient THA/TKA volume was about 30 percent lower than in April of 2020. These trends are expected to continue. Given the proportion of THA/TKA procedures that are moving the

outpatient setting, this measure fills a gap by measuring performance at the HOPD/ASC facility level (in the HOQR and ASCQR programs, respectively.

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: Hospital Quality Reporting (HQR) tool

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

The primary data source for development and testing of this measure was patient-reported outcome data collected with PROM instruments and additional patient and provider-reported risk variable data collected through the Center for Medicare and Medicaid Innovation (CMMI) Comprehensive Care for Joint Replacement (CJR) payment model. This model is an ongoing proof of concept among participating hospitals for broad, prospective collection of PRO data, implementing real-world data collection and data submission for centralization, risk adjustment and measure calculation. Feasibility is an NQF criteria; there are two other versions of this measure (both capture inpatient procedures) one at the hospital level which is NQF endorsed (NQF 3559), and one at the clinician level (NQF 3639) which has been recommended for endorsement by NQFs Surgery Standing Committee (the measure is currently in public comment). NQF 3559 received a moderate rating for feasibility; NQF 3639 received a high rating. In addition, during data collection for CJR, some hospitals inadvertently submitted data on procedures for HOPDs, therefore demonstrating the feasibility of implementing this measure in HOQR. The same data fields would be collected from ASC facilities.

Method of Measure Calculation

Claims;Other (enter here):: Patient-reported outcome (PRO) data, other health data (BMI, narcotic use)

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

Below we provide the range of performance for the hospital-level measure. We also provide preliminary results for HOPDs in the form of pre-procedural PROM results which are predictive of meeting the threshold of improvement. We note that the hospital-level measure is NQF endorsed, and passed the evaluation criteria of performance gap this is also true of the clinician group/clinician level measure. The distribution of performance for the clinician level measure is provided in Attachment 1. Inpatient

Procedure Results (hospital-level): In 123 hospitals with at differences least 25 THA/TKA patients with complete PRO data in the measurement period, we found variation in RSIRs suggesting meaningful in performance measure scores across hospitals. The mean risk-standardized improvement rate (representing the risk-standardized percentage of patients achieving substantial clinical benefit improvement) across hospitals was 60.16% with a standard deviation of 19.58. The minimum hospital RSIR was 6.65% and the maximum hospital RSIR was 86.84%. The interquartile range (54.36, 72.51%) represents a difference of 18 percentage points, and the difference between the 10th and 90th percentiles (20.94% and 78.85%, respectively) is just shy of 58 percentage points. This variation indicates an important quality gap among hospitals measured. Variation in hospital performance was also evaluated by calculating the median odds ratio (OR) for all hospitals in the dataset (n=238). The median OR represents the median increase in odds of the patient outcome (substantial clinical benefit improvement in PROM score from preoperative to postoperative assessment) if a procedure on a single patient was performed by a higher performing hospital compared to a lower performing hospital. It is calculated by taking all possible combinations of hospitals always comparing the higher performing hospitals to the lower performing hospitals. The median OR is interpreted as a traditional odds ratio would be. Results suggest significant and substantial increases in the likelihood of substantial clinical benefit improvement by higher performing hospitals compared to lower performing hospitals. At the hospital level, the median OR value indicates that a patient is 3.44 times more likely to achieve substantial clinical benefit improvement if their elective primary THA/TKA procedure was performed by a higher performing hospital than by a lower performing hospital. Pre-procedure results (HOPD-performed procedures): In the Performance Year 5 reporting for CMMIs Comprehensive Joint Replacement (CJR) model, 82 facilities reported 1,351 procedures (matched to claims) performed in the HOPD setting; 21 of the 82 HOPDs have >=25 cases. At the patient level, mean pre-procedure PROM scores were 47.9 (SD, 14.1); the median was 47.5 (IQR: 39.9-57.1). This suggests that about 50%-60% of the patients in this dataset will experience an improvement in PROM scores that meet the current criteria for substantial clinical benefit improvement. Among the 82 facilities, the range of pre-procedure PROM scores was 28.1-86.7 with a mean of 49.0 (SD 9.34); the median was 49.0 (IQR: 44.1-53.5). We note that this is preliminary (pre-procedure) data on a limited set of facilities. We expect a wider distribution when data is collected more broadly.

Unintended Consequences

N/A

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

Internal data analysis

Summarize the empirical data

Addressing quality of care for common and costly procedures such as THAs and TKAs is essential. THAs and TKAs are common surgeries among Medicare beneficiaries, with Medicare direct payments to hospitals for THA/TKA exceeding $15 billion annually (Miller et al., 2011). Between April 1, 2017 to October 2, 2019, there were 786,830 THA and TKA procedures performed in the inpatient setting for Medicare Fee-for-Service (FFS) beneficiaries 65 years and older (DeBuhr et al., 2021). For the US population as a whole, some project that annual THA and TKA procedures performed will reach nearly 2 million by 2030 (Lopez et al., 2020). CMS has recently allowed elective primary THA and TKA procedures to be performed in both the HOPD and ASC setting; CMS removed both procedures from the inpatient-only (IPO) list and added them to the ASC Covered Procedures List (CPL): TKA was removed from the IPO for CY 2018 and added to the ASC CPL in CY2020; THA was removed from the IPO for CY 2020 and added to the ASC CPL in CY2021.

Based on claims analyses performed by the developer (Yale/CORE) on Medicare Fee For Service claims, between April 1, 2018 and March 30, 2021 , there were about 330,000 THA and TKA procedures performed in the outpatient (HOPD setting). In addition, after the onset of the COVID-19 pandemic in 2020, outpatient procedures for both THA and TKA procedures outnumbered inpatient procedures; as of September 2021, inpatient THA/TKA volume was about 30 percent lower than in April of 2020. These trends are expected to continue. Given the proportion of THA/TKA procedures that are moving the outpatient setting, this measure fills a gap by measuring performance at the HOPD/ASC facility level.

THA/TKAs are important, effective procedures performed on a broad population, and the patient-reported outcomes for these procedures (for example, pain, mobility, and quality of life) can be measured in a scientifically sound way (Alviar et al., 2011 [a]; Alviar et al., 2011 [b]; Bauman et al., 2007; Collins & Roos, 2012; Jones et al., 2007; Jones & Pohar, 2012; Lau et al., 2012; Liebs, 2016; Montin et al., 2008; Papalia et al., 2012; Rolfson et al., 2011; Thorborg et al., 2010; White & Master, 2016) and are influenced by a range of improvements across the full spectrum of care.

THA/TKA provides a suitable environment for optimizing care, as there are many studies indicating how providers can improve outcomes of the patients by addressing aspects of pre-, peri-, and postoperative care (Brown et al., 2012; Choong et al., 2009; Galea et al., 2008; Kim, 2019; McGregor et al., 2004; Moffet et al., 2004; Monticone et al., 2013; Walters, 2016).

Optimal clinical outcomes depend not just on the surgeon performing the procedure, but also on: the entirety of the team's efforts in the care of the patient; care coordination across provider groups and specialties; and the patients' engagement in their recovery (Feng et al, 2018; Saufl et al, 2007). Even the best surgeon will not get outstanding results if there are gaps in the quality of care provided by others caring for the patient before, during, and/or after surgery. The goal of facility-level outcome measurement is to capture the full spectrum of care to incentivize collaboration and shared responsibility for improving patients' health and reducing the burden of their disease.

Name evidence type

Patient Workgroup feedback

Summarize the evidence

Patients who have undergone a THA or TKA have been engaged for input on measure development through participation on the Technical Expert Panel (TEP) and through a Patient Working Group assembled with assistance from the National Partnership for Women and Families in 2018. Overall, five patients (two males and three females) have provided input through TEP participation: two patients participated in four TEP meetings in 2013 and 2014; they were unavailable to continue participation when the TEP was reconvened in 2018, and two new patients participated in two TEP meetings in 2018 and 2019; and a fifth patient participated in the final TEP meeting in 2020 when one of two prior patients could not continue. The Patient Working Group consisted of five females and one male who have undergone at least one hip and/or knee replacement and were distinct from those who participated in the TEP. These patients were convened for three meetings, one in July 2018, one in February 2019, and one in February 2020. Additional input was sought from both the TEP and the Patient Working Group through online surveys following some of their meetings.

Feedback from patients on both the TEP and the Patient Working Group indicate strong support for a patient reported outcomes-based performance measure following primary elective THA and TKA. Patients stated that they expect a significant amount of improvement in both pain level and functional status following a THA/TKA procedure and felt this was an extremely important aspect of care to be captured in this measure. Patients also noted that their surgical experience positively impacted not only their physical health, but their quality of life as well. Patients in the Patient Working Group supported a measure cohort that combined THA and TKA patients, while two patients on the TEP expressed some concern about differing postoperative recovery for hips and knees. All patients supported the risk model and accounting for social risk factors in an analytic approach to non-response bias. Patients expressed a desire to see measure results that reflect physician-level performance but agreed that a hospital-level measure is a good way to encourage communication across providers to improve coordination of care at a facility overall.

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;Other (enter here):: Feedback from patients

Is the measure risk adjusted?

Yes

Risk adjustment variables

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

Patient-level demographics: please select all that apply:

Age;Gender

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

Case-Mix Adjustment;Severity of Illness;Other (enter here):: Index admission of THA; number of procedures

Patient functional status: please select all that apply:

N/A

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

Other (enter here):: Health literacy

Proxy social risk factors: please select all that apply

N/A

Patient community characteristic: please select all that apply:

N/A

Risk model performance

Model performance is reported using data collected for CMMI's Comprehensive Joint Replacement (CJR) model. Because the risk model was developed based on clinical input (and not empirically) and because patients undergoing joint replacement in the outpatient settings are known to have the same risk factors, the model is expected to perform similarly on patients who undergo procedures in the outpatient setting. We note that CMS will have the opportunity to re-evaluate the model during a likely period of voluntary reporting, however we expect the model to work as robustly for procedures performed in the outpatient setting.

Discrimination statistics:

The calculated c-statistic was 0.68 using the Development Dataset and 0.69 using the Validation Dataset and indicates adequate model discrimination across the cohort models. With both the Development and Validation Datasets, the model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration statistics (Î³0, Î³1):

The calibration values which are consistently close to 0 at one end and close to 1 at the other end

indicates good calibration of the model. If the Î³0 in the model performance using Validation data is

substantially far from zero and the Î³1 is substantially far from 1, there is potential evidence of overfitting. The calibration values of close to zero at one end and close to 1 on the other end indicates

good calibration of the model between the Development and Validation Datasets.

Risk Decile Plots:

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which

show a good calibration of the model. This plot indicates good discrimination of the model and

good predictive ability

Response Bias Analysis:

Potential non-response bias due to non-response of PROs was addressed using stabilized inverse probability weighting, created with a multinomial logistic regression to calculate stabilized inverse probability weights. Due to the voluntary nature of PRO data and because PRO data are unlikely to be missing at random, we understand that accounting for potential non-response bias is important for this measure.

All eligible THA/TKA procedures performed during the measurement period at the 238 hospitals submitting complete PRO and risk variable data during the measurement period among 1,254 clinicians and 526 clinician groups with at least one of these procedures were identified via CMS claims data. These were categorized into one of three PRO response groups (complete PRO submission, incomplete PRO submission, and no response). Variables associated with unit non-response were identified in the data and through a literature review. Propensity scores were calculated using a multinomial logistic regression where the outcome was 1) complete PRO submission, 2) incomplete PRO submission, and 3) no response. Stabilized Inverse Probability Weights (IPW) were calculated for each of the three groups and incorporated into the hierarchical risk-adjustment model for SCB improvement following elective primary THA/TKA and used in calculation of the risk-adjusted and bias-adjusted RSIRs. Incorporating the stabilized weights in the calculation of the RSIRs helps to reduce bias due to non-response by giving higher weight to patients who were less likely to respond and deflating the weight of patients who were more likely to respond based on patient characteristics. Weighting the responders based on their likelihood of response, given their patient characteristics, helps reduce non-response bias in our RSIR measure.

The comparison of RSIRs for risk-adjusted model of SCB improvement with stabilized inverse probability weighting and without stabilized inverse probability weighting revealed only a small impact on the measure results of adjusting for potential non-response. However, we expect that non-response bias will be a factor for the THA/TKA PRO-PM, due to associations with non-response including socioeconomic status and health status. We therefore retained response bias adjustment within the measure specifications.

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

11

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

11

Meaningful to Patients: Numbers consulted

66

Meaningful to Patients: Number indicating survey/tool is meaningful

66

Meaningful to Clinicians: Numbers consulted

10

Meaningful to Clinicians: Number indicating survey/tool is meaningful

8

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.

00000

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

000000

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Construct Validity;Other (enter here):: Pearson Separation Index; test-retest reliability

Testing methodology and results

As noted in the section above, the instruments used to capture the outcome have been tested extensivity to establish their reliability and validity. The reliability results from the literature demonstrate that the HOOS, JR and the KOOS, JR PROM instrument

Burden for Provider: Was a provider workflow analysis conducted?

Yes

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

14

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

Facility-level signal-to-noise reliability

Signal-to-Noise: Sample size

123

Signal-to-Noise: Statistical result

.959

Signal-to-Noise: Interpretation of results

The facility-level signal to noise reliability provided above is for the hospital-level version however the THA/TKA PRO-PM measure under consideration for the HOPD and ASC settings has the same measure specifications as the NQF endorsed hospital measure.

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

Comparison of means within performance categories of the comparator measure.

Empiric Validity: Sample size

123

Empiric Validity: Statistical result

00000

Empiric Validity: Methods and findings

To assess empirical measure score validity, we compared the THA/TKA PRO-PM risk-standardized improvement rates (RSIRs) to the NQF endorsed Hip/Knee Complication Measure (NQF #1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary THA/TKA.) The THA/TKA Complications measure estimates the risk-adjusted rate that patients who have experienced an elective primary THA/TKA experience at least one of eight complications within 90 days of the procedure. The RSCR is categorized into 3 groups: worse than national average, same as national average, and better than national average. We hypothesized that hospitals in the worse than national average category would have lower mean performance on the THA/TKA PRO-PM measure compared with hospitals in the average or better than national average performance categories. Data for the hospital RSCRs from April 1, 2015 to March 31, 2018 were compared to RSIRs for procedures performed July 1, 2016 to June 30, 2017. We examined the distribution of THA/TKA PRO-PM RSIRs by THA/TKA RSCR national categories within hospitals submitting complete PRO data for at least 25 THA/TKA procedures: Hospitals worse than national average (those with higher complication rates), Hospitals the same as national average, and Hospitals better than national average (those with lower complication rates). Comparison of THA/TKA PRO-PM RSIRs to RSCR categories indicated an increasing monotonic trend. Those hospitals in the RSCR Worse than National Average category had lower median RSIRs (51.87%) than the median RSIR (66.49%) of hospitals in the "RSCR Same as National Average" category, which is lower than that of hospitals in the "RSCR Better than National Average" category (71.13%). The hospitals with lower risk-

adjusted complication rates had higher risk-adjusted THA/TKA improvement rates. As these outcomes are not clinically expected to be perfectly correlated but do reflect hospital-level care and processes impacting quality of care for patients experiencing elective primary THA/TKA surgery, we interpret the increasing monotonic trend between RSIRs and RSCR national categories as reflective of empiric measure validity.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

21

Face Validity: Result

19

Patient/Encounter Level Testing

Yes

Type of Analysis

Other (enter here):: How efficiently a set of items is able to separate the persons being measured.

Sample Size

2291

Statistic Name

Other (enter here):: Pearson separation reliability

Statistical Results

.84

Interpretation of results

The reliability results from the literature demonstrate that the HOOS, JR and the KOOS, JR PROM instruments are sufficiently reliable and exceed accepted norms for reliability testing. The results assessing internal consistency indicated person separation reliability of 0.86 - 0.87 for the HOOS, JR (Lyman et al., 2016a) and 0.84-0.85 for the KOOS, JR (Lyman et al., 2016b). Values above 0.7 indicate the ability of the instruments to differentiate patients with varying levels of pain and functioning, which in turn provides evidence of good internal consistency. Test-retest reliability was not tested by developers of the HOOS, JR as it had already been tested in the Hip dysfunction and Osteoarthritis Outcome Score (HOOS) in several validation studies (Klassbo et al, 2003; de Groot et al, 2007; Ornetti et al, 2010; Nilsdotter & Bremander, 2011). Intra-class correlation coefficients (ICCs) were used to determine test-retest reproducibility and ranged from 0.75 to 0.97 in the validation studies. Specifically, the Pain and Activity of Daily Living domains, from which HOOS, JR pain and functioning questions are drawn, had ICCs of 0.83 - 0.89 (Pain sub-scale) and 0.86 - 0.94 (Activity of Daily Living sub-scale). Test-retest

reliability was also not tested by developers of the KOOS, JR as it had already been tested in the Knee injury and Osteoarthritis Outcome Score (KOOS) (Roos et al, 1998). Intra-class correlation coefficients (ICCs) were used to determine test-retest reproducibility and ranged from 0.75 to 0.93. Specifically, the Pain, Activity of Daily Living and Symptom domains, from which KOOS, JR pain, functioning and stiffness questions are drawn, had ICCs of 0.85 (Pain sub- scale), 0.75 (Activity of Daily Living sub-scale), and 0.93 (Symptoms). The validity results from the literature demonstrate that the HOOS, JR and the KOOS, JR PROM instruments are valid and meaningful measures for assessing PROs following THA/TKA procedures. The HOOS, JR and the KOOS, JR showed very high responsiveness, well beyond the 0.8 standardized response mean value considered very large (Steiner et al., 2003). Spearman correlation values between the HOOS, JR and the HOOS domains from which the HOOS, JR questions were drawn (Pain and Activity of Daily Living domains) were high; likewise, Spearman correlation values between the KOOS, JR and the KOOS Pain and Activity of Daily Living domains were high, and were moderate between the KOOS, JR and the Symptom domain. Floor effects were small; ceiling effects for the HOOS, JR were 37% 46%, but were comparable to or better than HOOS domains and the WOMAC (Lyman et al., 2016a).

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

00000

Median performance score

00000

Minimum performance score

000000

Maximum performance score

000000

Standard deviation of performance scores

000000

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

Yes

Surveys or patient-reported outcome tools

All of the instruments/patient reported data elements are listed below. All have been validated (references shown). None of the surveys require licenses or fees for use. Mode of administration include: paper, telephone and electronic. Please see Attachmen

Section 5: Measure Contact Information

Measure Steward

CMS

Measure Steward Contact Information

Janis Grady
7500 Security Boulevard
Baltimore, MD 21244
janis.grady@cms.hhs.gov
(410) 786-7217

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

N/A

MUC2022-027 Facility Commitment to Health Equity

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program; Inpatient Psychiatric Facility Quality Reporting Program; Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

This structural measure assesses facility commitment to health equity using a suite of equity-focused organizational competencies aimed at achieving health equity for racial and ethnic minority groups, people with disabilities, members of the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) community, individuals with limited English proficiency, rural populations, religious minorities, and people living near or below poverty level. Facilities will receive one point each for attesting to five different domains of commitment to advancing health equity for a total of five points.

Numerator

"This structural measure assesses facility commitment to health equity using a suite of equity-focused organizational competencies aimed at achieving health equity for racial and ethnic minority groups, people with disabilities, members of the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) community, individuals with limited English proficiency, rural populations, religious minorities, and people living near or below poverty level. Facilities will receive one point each for attesting to five different domains of commitment to advancing health equity for a total of five points.

Facilities participating in the specific Quality Reporting Programs must answer the questions during the CMS specified time period. The five domains for facility attestation and key questions for each domain are the following:

Domain 1: Equity is a Strategic Priority

Facility commitment to reducing healthcare disparities is strengthened when equity is a key organizational priority. Please attest that your facility has a strategic plan for advancing healthcare equity and that it includes all of the following elements. Select all that apply (note: attestation of all elements is required in order to qualify for the measure numerator):

A. Our facility strategic plan identifies priority populations who currently experience health disparities.
B. Our facility strategic plan identifies healthcare equity goals and discrete action steps to achieving these goals.
C. Our facility strategic plan outlines specific resources which have been dedicated to achieving our equity goals.
D. Our facility strategic plan describes our approach for engaging key stakeholders, such as community-based organizations.

Domain 2: Data Collection

Collecting valid and reliable demographic and social determinant of health data on patients served in a facility is an important step in identifying and eliminating health disparities. Please attest that your facility engages in the following activities. Select all that apply (note: attestation of all elements is required in order to qualify for the measure numerator):

A. Our facility collects demographic information, including self-reported race and ethnicity, and/or social determinant of health information on the majority of our patients.
B. Our facility has training for staff in culturally sensitive collection of demographic and/or social determinant of health information.
C. Our facility inputs demographic and/or social determinant of health information collected from patients into structured, interoperable data elements using a certified EHR technology.

Domain 3: Data Analysis

Effective data analysis can provide insights into which factors contribute to health disparities and how to respond. Please attest that your facility engages in the following activities. Select all that apply (note: attestation of all elements is required in order to qualify for the measure numerator):

A. Our facility stratifies key performance indicators by demographic and/or social determinants of health variables to identify equity gaps and includes this information on facility performance dashboards.

Domain 4: Quality Improvement

Health disparities are evidence that high quality care has not been delivered equally to all patients. Engagement in quality improvement activities can improve quality of care for all patients. Select all that apply (note: attestation of all elements is required in order to qualify for the measure numerator):

A. Our facility participates in local, regional, or national quality improvement activities focused on reducing health disparities.

Domain 5: Leadership Engagement

Leaders and staff can improve their capacity to address disparities by demonstrating routine and thorough attention to equity and setting an organizational culture of equity. Please attest that your facility engages in the following activities. Select all that apply (note: attestation of all elements is required in order to qualify for the measure numerator):

A. Our facility senior leadership, including chief executives and the entire hospital board of trustees, annually reviews our strategic plan for achieving health equity.
B. Our facility senior leadership, including chief executives and the entire facility board of trustees, annually reviews key performance indicators stratified by demographic and/or social factors."

Numerator Exclusions

There are no numerator exclusions

Denominator

The denominator for each facility is 5 which represents the total number of questions.

The measure is calculated as the number of complete attestations / total number of questions. There is no partial credit for any question. Attestation of all elements is required in order to qualify for the measure numerator

Denominator Exclusions

There are no denominator exclusions

Denominator Exceptions

There are no denominator exceptions.

State of development

Fully Developed

State of Development Details

N/A

What is the target population of the measure?

Facilities serving Medicare Fee for Service beneficiaries

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

Other: N/A

Measure Type

Structure

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?

Other: Provider data entry (attestation-based statements)

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?

Facility

In which setting was this measure tested?

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

N/A

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

Yes

CMIT ID

12759

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Never Submitted

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

00000

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?

2021

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

MUC2021-106

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

Rural Health Advisory, 2021 Health Equity Advisory, 2021 Clinician, 2021 Hospital, 2021 Post-Acute Care/Long-Term Care, 2021

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

Hospital IQR Program, 2021

What was the MAP recommendation in this year?

Hospital IQR Program, 2021, Conditionally Support

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

For the Hospital IQR Program, MAP conditionally supported the measure for rulemaking pending CBE endorsement.

MAP report page number being referenced for this year:

2021, pages 20-21

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

Currently proposed for CY 2023 reporting period/FY 2025 payment determination

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program

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?

Hospital Commitment to Health Equity

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

This measure will address different care settings from the IQR version of the measure

How will this measure add value to the CMS program?

This measure will address additional care settings

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

000000

Section 2: Measure Evidence

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

Web interface

Stratification

No

Feasibility of Data Elements

No data elements are in defined fields in electronic sources

Feasibility Assessment

The data source for the structural measure is self-attestation by hospitals participating in the Hospital Inpatient Quality Reporting Program.

CMS has previously collected attestation-based measures, such as the proposed measure under consideration. Attestation may be provided to CMS using existing electronic data submission portals with minimal administrative burdens.

Method of Measure Calculation

Other digital method

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

N/A

Unintended Consequences

For facilities that do not meet the five areas emphasized in this measure, this could create burden to address the measurement area and move resources from other areas of focus. Because this is a structural measure, there is no direct assessment of improvement in quality on the basis of these actions. However, the intent of measurement is to support facilities making needed investments in leadership, data and culture to advance equity. We believe the activities outlined in the attestation questions are foundational best practices for advancing health equity for patients and communities.

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

99999

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

Significant and persistent inequities in health care outcomes exist in the United States. Belonging to a racial or ethnic minority group, living with a disability, being a member of the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) community, living in a rural area, or being near or below the poverty level, is often associated with worse health outcomes.[1],[2],[3],[4],[5],[6],[7],[8] Numerous studies have shown that among Medicare beneficiaries, racial and ethnic minority individuals often receive lower quality of hospital care, report lower experiences of care, and experience more frequent hospital readmissions and procedural complications. [9],[10],[11],[12],[13],[14] Readmission rates for the most common conditions in the Hospital Readmissions Reduction Program are higher for black Medicare beneficiaries and higher for Hispanic Medicare beneficiaries with Congestive Heart Failure and Acute Myocardial Infarction.[15],[16],[17],[18],[19] To ensure that all patients receive excellent care when hospitalized regardless of their individual characteristics, strong and committed leadership from hospital executives and board members is essential. Publications from the Agency for Healthcare Research and Quality and The Joint Commission identify the important role of hospital leadership in promoting a culture of quality and safety.[20],[21] Studies have shown that interventions taken by hospital leadership can positively influence culture[22] and that health care organizational culture can translate into better quality outcomes and experience of care. [23] ,[24],[25] A 2013 systematic review of 122 published studies found an association between hospital board composition and processes and high-performance.[26] Health disparities are evidence that high quality care has not been delivered equally to all patients. Studies from the Institute for Healthcare Improvement identified five core features of health care organizations that make health equity a core strategy, including making health equity a leader-driven priority and developing structures and processes that support equity.[27] This measure aligns with the National Quality Forum strategic goal of advancing health equity and addressing disparities.[28] The five questions of the structural measures are adapted from the CMS Office of Minority Health, Building an Organizational Response to Health Disparities[29] framework for helping health care organizations build a response to health disparities through focus on data collection, data analysis, culture of equity, and quality improvement.

1. Joynt KE, Orav E, Jha AK. Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care. JAMA. 2011;305(7):675-681.

2. Lindenauer PK, Lagu T, Rothberg MB, et al. Income Inequality and 30 Day Outcomes After Acute Myocardial Infarction, Heart Failure, and Pneumonia: Retrospective Cohort Study. British Medical Journal. 2013;346.

3. Trivedi AN, Nsa W, Hausmann LRM, et al. Quality and Equity of Care in U.S. Hospitals. New England Journal of Medicine. 2014;371(24):2298-2308.

4. Polyakova, M., et al. Racial Disparities In Excess All-Cause Mortality During The Early COVID-19 Pandemic Varied Substantially Across States. Health Affairs. 2021; 40(2): 307-316.

5. Rural Health Research Gateway. Rural Communities: Age, Income, and Health Status. Rural Health Research Recap. November 2018.

6. https://www.minorityhealth.hhs.gov/assets/PDF/Update_HHS_Disparities_Dept-FY2020.pdf

7. www.cdc.gov/mmwr/volumes/70/wr/mm7005a1.htm

8. Poteat TC, Reisner SL, Miller M, Wirtz AL. COVID-19 Vulnerability of Transgender Women With and Without HIV Infection in the Eastern and Southern U.S. Preprint. medRxiv. 2020;2020.07.21.20159327. Published 2020 Jul 24. doi:10.1101/2020.07.21.20159327

9. Martino, SC, Elliott, MN, Dembosky, JW, Hambarsoomian, K, Burkhart, Q, Klein, DJ, Gildner, J, and Haviland, AM. Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage. Baltimore, MD: CMS Office of Minority Health. 2020.

10. Guide to Reducing Disparities in Readmissions. CMS Office of Minority Health. Revised August 2018. Available at: https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/OMH_Readmissions_Guide.pdf

11. Singh JA, Lu X, Rosenthal GE, Ibrahim S, Cram P. Racial disparities in knee and hip total joint arthroplasty: an 18-year analysis of national Medicare data. Ann Rheum Dis. 2014 Dec;73(12):2107-15.

12. Rivera-Hernandez M, Rahman M, Mor V, Trivedi AN. Racial Disparities in Readmission Rates among Patients Discharged to Skilled Nursing Facilities. J Am Geriatr Soc. 2019 Aug;67(8):1672-1679.

13. Joynt KE, Orav E, Jha AK. Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care. JAMA. 2011;305(7):675-681

14. Tsai TC, Orav EJ, Joynt KE. Disparities in surgical 30-day readmission rates for Medicare beneficiaries by race and site of care. Ann Surg. Jun 2014;259(6):1086-1090.

15. Rodriguez F, Joynt KE, Lopez L, Saldana F, Jha AK. Readmission rates for Hispanic Medicare beneficiaries with heart failure and acute myocardial infarction. Am Heart J. Aug 2011;162(2):254-261 e253.

16. Centers for Medicare and Medicaid Services. Medicare Hospital Quality Chartbook: Performance Report on Outcome Measures; 2014.

17. Guide to Reducing Disparities in Readmissions. CMS Office of Minority Health. Revised August 2018. Available at: https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/OMH_Readmissions_Guide.pdf

18. Prieto-Centurion V, Gussin HA, Rolle AJ, Krishnan JA. Chronic obstructive pulmonary disease readmissions at minority-serving institutions. Ann Am Thorac Soc. Dec 2013;10(6):680-684.

19. Joynt KE, Orav E, Jha AK. Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care. JAMA. 2011;305(7):675-681

20. Leadership Role in Improving Patient Safety. Agency for Health Care Research and Quality. Patient Safety Primer, September 2019: Available at: https://psnet.ahrq.gov/primer/leadership-role-improving-safety

21. Joint Commission on Accreditation of Healthcare Organizations, USA. Leadership committed to safety. Sentinel Event Alert. 2009 Aug 27;(43):1-3. PMID: 19757544

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

00000

Type of Evidence to Support the Measure

Peer-Reviewed Systematic Review

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):: 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

000000

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

000000

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

No

Reliability: Type of Reliability Testing

N/A

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

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

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: 5 Point Score

Measure Performance Score Interpretation

Higher score is better

Mean performance score

00000

Median performance score

00000

Minimum performance score

00000

Maximum performance score

5

Standard deviation of performance scores

00000

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

CMS

Measure Steward Contact Information

Jennifer Robinson
7500 Security Blvd
Baltimore, MD 21244
jennifer.robinson@cms.hhs.gov
(443) 729-6368

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Samantha Mancuso
195 Church St, 5th Floor
New Haven, CT 06510
samantha.mancuso@yale.edu
(203) 497-1239

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

N/A

MUC2022-050 Screen Positive Rate for Social Drivers of Health

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program; Inpatient Psychiatric Facility Quality Reporting Program; Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The Screen Positive Rate for Social Drivers of Health is a structural measure that provides information on the percent of patients admitted for an inpatient facility stay or that have received established care in the case of dialysis facilities, and who are 18 years or older on the date of admission or date of established care in the case of dialysis facilities, were screened for all five HSRNs, and who screen positive for one or more of the following five HRSNs: Food insecurity, housing instability, transportation problems, utility difficulties, or interpersonal safety.

Numerator

The numerator consists of the number of patients admitted for an inpatient facility stay or that have received established care in the case of dialysis facilities, who are 18 years or older on the date of admission, who were screened for all five HSRNs, and who screen positive for having a need in one or more of the following five HRSNs (calculated separately): Food insecurity, housing instability, transportation needs, utility difficulties or interpersonal safety.

Numerator Exclusions

N/A

Denominator

The denominator consists of the number of patients admitted for an inpatient facility stay or that have received established care in the case of dialysis facilities who are 18 years or older on the date of admission and are screened for all five HSRNs (food insecurity, housing instability, transportation needs, utility difficulties and interpersonal safety).

Denominator Exclusions

The following patients would be excluded from the denominator: (1) Patients who opt-out of screening; and (2) patients who are themselves unable to complete the screening during their inpatient stay or during established care in the case of dialysis facilities and have no legal guardian or caregiver able to do so on the patient's behalf during their inpatient stay or during established care in the case of dialysis facilities.

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

Public and/or population health

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 Clinical Data (non-EHR);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?

Facility

In which setting was this measure tested?

Ambulatory/office-based care;Behavioral health clinic;Inpatient psychiatric facility;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

N/A

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

Yes

CMIT ID

12763-C-TBD

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?

2021

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

MUC2021-134

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

Rural Health Advisory, 2021 Health Equity Advisory, 2021 Clinician, 2021 Hospital, 2021 Post-Acute Care/Long-Term Care, 2021

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

Hospital IQR Program, 2021, MIPS, 2021

What was the MAP recommendation in this year?

Hospital IQR Program, 2021, Conditionally Support MIPS, 2021, Conditionally Support

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

For the MIPS Program, MAP conditionally supported the measure for rulemaking pending CBE endorsement. An additional suggested condition was the results of MUC2021-134 not being used to penalize or criticize healthcare providers under the MIPS or IQR programs. For the Hospital IQR Program, MAP conditionally supported this MUC. Conditions for support are contingent upon CBE endorsement to address reliability and validity concerns.

MAP report page number being referenced for this year:

2021, pages 43-44

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), HIQR: Proposed for voluntary reporting in the CY 2023 reporting period and mandatory reporting in the CY 2024 reporting period/FY 2026 payment determination and for subsequent years.

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program;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?

Web interface

Stratification

Yes (enter here):: The result of this measure would be calculated as five separate rates. Each rate is derived from the number of patients admitted for an inpatient facility stay or that received established care in the case of dialysis facilities and are

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

To report the measure, providers must collect the total number of patients and the number of patients who were screened for all five elements; the only demographic information needed is patient age. The screening tool data can be electronically collected and recorded; therefore all of these data points should be available to providers for reporting. The screening tool has been in use in 21 states across the US, with nearly one million patients screened.

Some data elements are in defined fields in electronic sources; Patient/family-reported information: electronic; Patient/family-reported information: paper

SDOH screening and data collection is already occurring at scale throughout the sector:

Pre-COVID JAMA study found that 24% of hospitals and 16% of physician practices are already screening for all 5 SDOH domains and 92% of hospitals and 66% of physician practices are screening for one or more of the 5 SDOH domains specified in the measures.

Source:

Fraze, Taressa K., et al. "Prevalence of screening for food insecurity, housing instability, utility needs, transportation needs, and interpersonal violence by US physician practices and hospitals." JAMA network open 2.9 (2019): e1911514-e1911514.

Using a standard, validated screening tool, CMS' Accountable Health Community (AHC) model has screened nearly 1 million beneficiaries for Health-Related Social Needs (HRSN) across 21 states, with 33% of beneficiaries screened having at least one HRSN.

Sources:

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

CMMIs 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.

Sources:

https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2

Method of Measure Calculation

Claims;Other digital method;Hybrid

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

CMS has already identified social and economic determinants as both a measurement priority and gap in Meaningful Measures 2.0. 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.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

A potential unintended consequence of the measure is that health systems and facilities 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

Health outcomes are ~80 percent driven by socioeconomic factors, health behaviors, and the physical environment (1). Reviews have collected numerous studies identifying a causal relationship between poor health outcomes and homelessness (2) food insecurity (3), and other needs screened for by the tool cited in this measure (4). The process of screening itself is consistent with guidance promulgated by the American Academy of Pediatrics (5), The American Academy of Family Practitioners (6), and guidance by the U.S. Preventative Services Task Force (USPSTF) (7) as well as a recommendation/clinical guideline from USPSTF. USPSTF concludes that screening for intimate partner violence (IPV) in women of reproductive age and providing or referring women who screen positive to ongoing support services has a moderate net benefit. The USPSTF notes there is evidence that available tools accurately identify interpersonal violence experienced by women. This recommendation is based on a systematic review of the evidence, including 3 randomized controlled trials (RCT) comparing IPV screening with brief intervention and information on referral options with no screening and 2 RCTs that reported no harm in screening. The review also examined 15 studies assessing the accuracy of screening tools. (8) This clinical guideline supports screening for interpersonal safety, which is one of five social domains included in this measure. (9)

Sources:

(1)(2) Hood, Carlyn M et al., County Health Rankings: Relationships Between Determinant Factors and Health Outcomes, American journal of preventive medicine vol. 50,2 (2016): 129-35. doi:10.1016/j.amepre.2015.08.024

(3) Stafford, Amanda, and Lisa Wood.Tackling Health Disparities for People Who Are Homeless Start with Social Determinants., International journal of environmental research and public health vol. 14,12 1535. 8 Dec. 2017, doi:10.3390/ijerph14121535

(4)https://www.healthcarevaluehub.org/download_file/1489/0

(5) https://www.aap.org/en-us/advocacy-and-policy/aap-%20health-initiatives/Screening/Pages/Social-Determinants-%20of-Health.aspx

(6)https://www.aafp.org/dam/AAFP/documents/patient_care/everyone_project/sdoh-guide.pdf

(7)Davidson, Karina W., et al. "Developing primary care based recommendations for social determinants of health: methods of the US Preventive Services Task Force." Annals of internal medicine 173.6 (2020): 461-467.

(8)Feltner C, Wallace I, Berkman N, et al. Screening for Intimate Partner Violence, Elder Abuse, and Abuse of Vulnerable Adults: An Evidence Review for the US Preventive Services Task Force: Evidence Synthesis No. 169. Rockville, MD: Agency for Healthcare Research and Quality; 2018. AHRQ publication 18-05240-EF-1.

(9)Screening for Intimate Partner Violence, Elder Abuse, and Abuse of Vulnerable Adults: US Preventive Services Task Force Final Recommendation Statement | Geriatrics | JAMA | JAMA Network)

Name the guideline developer/entity

United States Preventive Services Task Force (USPSTF)

Publication year

2018

Full citation +/- URL

US Preventive Services Task Force. Screening for Intimate Partner Violence, Elder Abuse, and Abuse of Vulnerable Adults: US Preventive Services Task Force Final Recommendation Statement. JAMA. 2018;320(16):16781687. doi:10.1001/jama.2018.14741
https://jamanetwork.com/journals/jama/fullarticle/2708121

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 interpersonal 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. B: The USPSTF recommends the service. There is high certainty that the net benefit is moderate and there is moderate certainty that the net benefit is moderate to substantial. 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. D:The USPSTF recommends against the service. There is moderate to high certainty that the service has no net benefit or that the harms outweigh the benefits. 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 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?

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 prevention service on health outcomes. This conclusion is therefore likely 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, and 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 the effects on health outcomes. Evidence is insufficient because of: - The limited number of 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?

Moderate or similar

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

The USPSTF recommends that clinicians screen for interpersonal 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

1

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

Please refer to the attached document "Screen Positive Rate for SDOH Table of Peer Reviewed Evidence and Related Research.pdf".

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies);Internal data analysis

Summarize the empirical data

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

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

Estimate of Average Cost Savings Per Event: 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

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

3162

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

2441

Meaningful to Patients: Numbers consulted

24413162

Meaningful to Patients: Number indicating survey/tool is meaningful

24413162

Meaningful to Clinicians: Numbers consulted

10078

Meaningful to Clinicians: Number indicating survey/tool is meaningful

8800

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.

10078

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

8800

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Construct Validity

Testing methodology and results

Through AHC, the measures have been tested for 5 years with 1M+beneficiaries in 644 clinical sites with 40% of the DOH screenings in hospital inpatient or ED settings and 54% in primary care practices. 6 of 30 AHC awardees served either mostly rural counties or served exclusively rural counties. Refer to Reliability and Empiric Validity sections for more details on testing results.

Burden for Provider: Was a provider workflow analysis conducted?

Yes

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

3224

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): Measure Score Reliability; Data Element Reliability, IRR (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

IRR (Inter-rater reliability)

Other: Sample size

1008

Other: Statistical result

0.60,0.52,0.75

Other: Interpretation of results

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.

Empiric Validity

Yes

Empiric Validity: Statistic name

Data Element Validity Internal Consistency; Predictive Validity; Other: 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/)

Empiric Validity: Sample size

60,000

Empiric Validity: Statistical result

99.8

Empiric Validity: Methods and findings

Validity Testing Statistical Result:

Study 1: Sample Size: 1,008. A reported social risk on the AHC and YCLS measures was strongly associated with having fair or poor self-rated health

Source:

https://www.jfmpc.com/viewimage.asp img=JFamMedPrimaryCare_2020_9_9_5026_296311_t6.jpg

Study 2: Sample Size: 30,098. 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: Sample Size: 60,000. 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-foodinsecurity-accurately-identifies-highrisk-us-adults/81A4F5E162241E289A5181A10C056125

Validity Testing Interpretation of Results:

Study 1: These results are the first to suggest that both the AHC and YCLS have concurrent and predictive validity, supporting their use in healthcare settings, including by primary care physicians to engage in social risk-informed care.

Source:

PAGE 67 Top of Document Screening for Social Drivers of Health
https://www.jfmpc.com/viewimage.asp?img=JFamMedPrimaryCare_2020_9_9_5026_296311_t6.jpg

Study 2: A 2-item FI screen was sensitive, specific, and valid among low-income families with young children. The FI screen rapidly identifies households at risk for FI, enabling providers to target services that ameliorate the health and developmental consequences associated with FI.

Source:

https://pubmed.ncbi.nlm.nih.gov/20595453/

Study 3: The test characteristics of multiple two-item combinations of questions assessing food insecurity had adequate sensitivity (>97 %) and specificity (>70 %) for widespread adoption as clinical screening measures.

Source:

https://www.cambridge.org/core/journals/public-health-nutrition/article/brief-assessment-of-foodinsecurity-accurately-identifies-highrisk-us-adults/81A4F5E162241E289A5181A10C056125

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;Agreement between other gold standard and manual reviewer

Sample Size

1,000,000

Statistic Name

Sensitivity

Statistical Results

97

Interpretation of results

The AHC screening tool used to generate the measures has been psychometrically evaluated at both the item/domain (F/H/T) and tool level and has demonstrated evidence of both reliability and validity, including predictive and concurrent validity, in healthcare settings. This includes comparison with other screening tools (e.g., Your Current Life Situation and We Care instruments) producing high kappa statistics (generally > 0.6) as well as adequate sensitivity and specificity (up to 97% sensitivity) Through AHC, the measures have been tested for 5 years with 1M+ beneficiaries in 644 clinical sites with 40% of the DOH screenings in hospital inpatient or ED settings and 54% in primary care practices. 6 of 30 AHC awardees served either mostly rural counties or served exclusively rural counties. Refer to Reliability and Empiric Validity sections for more details on testing results.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Lower score is better

Mean performance score

33

Median performance score

0000

Minimum performance score

0000

Maximum performance score

0000

Standard deviation of performance scores

0000

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

Yes

Surveys or patient-reported outcome tools

Accountable Health Communities (AHC) Health-Related Social Needs (HRSN) Screening Tool. The AHC screening tool used to generate the measures has been psychometrically evaluated at both the item/domain (F/H/T) and tool level and has demonstrated evidence of both reliability and validity, including predictive and concurrent validity, in healthcare settings.

This includes comparison with other screening tools (e.g., Your Current Life Situation and We Care instruments) producing high kappa statistics (generally > 0.6) as well as adequate sensitivity and specificity (up to 97% sensitivity). See https://innovation.cms.gov/media/document/ahcm-screening-tool-citation.

Due to variability across facility settings and the populations they serve, we are proposing to allow facilities flexibility with selection of tools to screen patients for food insecurity, housing instability, transportation needs, utility difficulties, and interpersonal safety. Potential sources of these data could include, for example, administrative claims data, electronic clinical data, standardized patient assessments, or patient-reported data and surveys. Multiple screening tools exist, and many facilities already have screening tools integrated into their electronic health records (EHRs). We suggest facilities refer to the Social Interventions Research and Evaluation Network (SIREN) website, for example, for comprehensive information about the most widely used HRSN screening tools.1,2 SIREN contains descriptions of the content and characteristics of various tools, including information about intended populations, completion time, and number of questions.

Sources:

1. Social Interventions Research & Evaluation Network. (2019). Social Needs Screening Tool Comparison Table. Available at: https://sirenetwork.ucsf.edu/tools-resources/resources/screening-tools-comparison.

2. The Social Interventions Research and Evaluation Network (SIREN) at University of California San Francisco was launched in the spring of 2016 to synthesize, disseminate, and catalyze research on the social determinants of health and healthcare delivery.

Section 5: Measure Contact Information

Measure Steward

CMS

Measure Steward Contact Information

Jennifer Robinson
7500 Security Blvd
Baltimore, MD 21244

jennifer.robinson@cms.hhs.gov
(410) 786-2066

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Mariel Thottam
195 Church St, 5th Floor
New Haven, CT 06510
mariel.thottam@yale.edu
(330) 417-7434

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

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 are projected to experience food insecurity 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. 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, he 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. -

Sources:

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-economys-effects-on-food-housing-and

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/2021-roadshow-deck/

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/

MUC2022-053 Screening for Social Drivers of Health

Program

End-Stage Renal Disease (ESRD) Quality Incentive Program; Inpatient Psychiatric Facility Quality Reporting Program; Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The Screening for Social Drivers of Health measure assesses the total number of patients, aged 18 years and older, screened for social risk factors (specifically, food insecurity, housing instability, transportation needs, utility difficulties, and interpersonal safety) during an inpatient facility stay, or during established care in the case of dialysis facilities. The measure cohort includes patients who are admitted to an inpatient facility or who have established care in the case of dialysis facilities and are 18 years or older on the date of admission or on the date of established care in the case of dialysis facilities.

Numerator

Number of patients admitted to an inpatient facility stay or who have established care in the case of dialysis facilities, who are 18 years or older on the date of admission or date of established care in the case of dialysis facilities and are screened for all of the following five HRSNs: Food insecurity, housing instability, transportation needs, utility difficulties, and interpersonal safety during their facility stay or during established care in the case of dialysis facilities.

Numerator Exclusions

N/A

Denominator

Number of patients who are admitted to a facility inpatient stay or have established care in the case of dialysis facilities, and who are 18 years or older.

Denominator Exclusions

The following patients would be excluded from the denominator: (1) Patients who opt-out of screening; and (2) patients who are themselves unable to complete the screening during their inpatient stay or during established care in the case of dialysis facilities and have no legal guardian or caregiver able to do so on the patient's behalf during their inpatient stay or during established care in the case of dialysis facilities.

Denominator Exceptions

N/A

State of development

Fully Developed

State of Development Details

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

Public and/or population health

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 Clinical Data (non-EHR);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?

Facility

In which setting was this measure tested?

Ambulatory/office-based care;Behavioral health clinic;Inpatient psychiatric facility;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

N/A

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

Yes

CMIT ID

12765-C-TBD

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?

2021

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

MUC2021-136

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

Rural Health Advisory, 2021 Health Equity Advisory, 2021 Clinician, 2021 Hospital, 2021 Post-Acute Care/Long-Term Care, 2021

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

Hospital IQR Program, 2021, MIPS, 2021

What was the MAP recommendation in this year?

Hospital IQR Program, 2021, Conditionally Support MIPS, 2021, Conditionally Support

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

For the MIPS Program, MAP conditionally supported this measure for rulemaking, pending CBE endorsement and successful testing of the measure's reliability and validity. For the Hospital IQR Program, MAP conditionally supported the measure for rulemaking pending CBE endorsement.

MAP report page number being referenced for this year:

2021, pages 41-43

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), HIQR: Proposed for voluntary reporting in the CY 2023 reporting period and mandatory reporting in the CY 2024 reporting period/FY 2026 payment determination and for subsequent years.

What other federal programs are currently using this measure?

Hospital Inpatient Quality Reporting Program;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?

Web interface

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

To report the measure, providers must collect the total number of patients and the number of patients who were screened for all five elements; the only demographic information needed is patient age. The screening tool data can be electronically collected and recorded; therefore all of these data points should be available to providers for reporting. The screening tool has been in use in 21 states across the US, with nearly one million patients screened.

Some data elements are in defined fields in electronic sources; Patient/family-reported information: electronic; Patient/family-reported information: paper

SDOH screening and data collection is already occurring at scale throughout the sector:

Pre-COVID JAMA study found that 24% of hospitals and 16% of physician practices are already screening for all 5 SDOH domains and 92% of hospitals and 66% of physician practices are screening for one or more of the 5 SDOH domains specified in the measures.

Source:

Fraze, Taressa K., et al. "Prevalence of screening for food insecurity, housing instability, utility needs, transportation needs, and interpersonal violence by US physician practices and hospitals." JAMA network open 2.9 (2019): e1911514-e1911514.

Using a standard, validated screening tool, CMS' Accountable Health Community (AHC) model has screened nearly 1 million beneficiaries for Health-Related Social Needs (HRSN) across 21 states, with 33% of beneficiaries screened having at least one HRSN.

Sources:

https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

https://innovation.cms.gov/media/document/ahc-fact-sheet-2020-prelim-findings

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.

Sources:

https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2

Method of Measure Calculation

Claims;Other digital method;Hybrid

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

CMS has already identified social and economic determinants as both a measurement priority and gap in Meaningful Measures 2.0. 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.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

A potential unintended consequence of the measure is that health systems and facilities 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

Health outcomes are ~80 percent driven by socioeconomic factors, health behaviors, and the physical environment (1). Reviews have collected numerous studies identifying a causal relationship between poor health outcomes and homelessness (2) food insecurity (3), and other needs screened for by the tool cited in this measure (4). The process of screening itself is consistent with guidance promulgated by the American Academy of Pediatrics (5), The American Academy of Family Practitioners (6), and guidance by the U.S. Preventative Services Task Force (USPSTF) (7) as well as a recommendation/clinical guideline from USPSTF. USPSTF concludes that screening for intimate partner violence (IPV) in women of reproductive age and providing or referring women who screen positive to ongoing support services has a moderate net benefit. The USPSTF notes there is evidence that available tools accurately identify interpersonal violence experienced by women. This recommendation is based on a systematic review of the evidence, including 3 randomized controlled trials (RCT) comparing IPV screening with brief intervention and information on referral options with no screening and 2 RCTs that reported no harm in screening. The review also examined 15 studies assessing the accuracy of screening tools. (8) This clinical guideline supports screening for interpersonal safety, which is one of five social domains included in this measure.(9)

Sources:

(1)(2) Hood, Carlyn M et al., County Health Rankings: Relationships Between Determinant Factors and Health Outcomes. American journal of preventive medicine vol. 50,2 (2016): 129-35. doi:10.1016/j.amepre.2015.08.024

(3) Stafford, Amanda, and Lisa Wood. Tackling Health Disparities for People Who Are Homeless? Start with Social Determinants. International journal of environmental research and public health vol. 14,12 1535. 8 Dec. 2017, doi:10.3390/ijerph14121535

(4)https://www.healthcarevaluehub.org/download_file/1489/0

(5) https://www.aap.org/en-us/advocacy-and-policy/aap-%20health-initiatives/Screening/Pages/Social-Determinants-%20of-Health.aspx

(6)https://www.aafp.org/dam/AAFP/documents/patient_care/everyone_project/sdoh-guide.pdf

(7)Davidson, Karina W., et al. "Developing primary care based recommendations for social determinants of health: methods of the US Preventive Services Task Force." Annals of internal medicine 173.6 (2020): 461-467.

(8) Feltner C, Wallace I, Berkman N, et al. Screening for Intimate Partner Violence, Elder Abuse, and Abuse of Vulnerable Adults: An Evidence Review for the US Preventive Services Task Force: Evidence Synthesis No. 169. Rockville, MD: Agency for Healthcare Research and Quality; 2018. AHRQ publication 18-05240-EF-1.

(9)Screening for Intimate Partner Violence, Elder Abuse, and Abuse of Vulnerable Adults: US Preventive Services Task Force Final Recommendation Statement | Geriatrics | JAMA | JAMA Network)

Name the guideline developer/entity

United States Preventive Services Task Force (USPSTF)

Publication year

2018

Full citation +/- URL

US Preventive Services Task Force. Screening for Intimate Partner Violence, Elder Abuse, and Abuse of Vulnerable Adults: US Preventive Services Task Force Final Recommendation Statement. JAMA. 2018;320(16):16781687. doi:10.1001/jama.2018.14741
https://jamanetwork.com/journals/jama/fullarticle/2708121

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 interpersonal 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. B:The USPSTF recommends the service. There is high certainty that the net benefit is moderate and there is moderate certainty that the net benefit is moderate to substantial. 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. D:The USPSTF recommends against the service. There is moderate to high certainty that the service has no net benefit or that the harms outweigh the benefits. 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 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?

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 prevention service on health outcomes. This conclusion is therefore likely 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, and 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 the effects on health outcomes. Evidence is insufficient because of: - The limited number of 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?

Moderate or similar

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

The USPSTF recommends that clinicians screen for interpersonal 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

1

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

Please refer to the attached document "Screening SDOH Table of Peer Reviewed Evidence and Research.pdf".

Source of empirical data

Published, peer-reviewed original research;Published and publicly available reports (e.g., from agencies);Internal data analysis

Summarize the empirical data

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 "many 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

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

Estimate of Average Cost Savings Per Event:

Extensive research exists demonstrating increased healthcare expenditures to patients including

Medicare beneficiaries associated with DOH. The example below provides the annualized increase in annual healthc

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

3162

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

2441

Meaningful to Patients: Numbers consulted

24413162

Meaningful to Patients: Number indicating survey/tool is meaningful

24413162

Meaningful to Clinicians: Numbers consulted

10078

Meaningful to Clinicians: Number indicating survey/tool is meaningful

8800

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.

10078

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

8800

Survey level testing

Yes

Type of Testing Analysis

Internal Consistency;Construct Validity

Testing methodology and results

Through AHC, the measures have been tested for 5 years with 1M+beneficiaries in 644 clinical sites with 40% of the DOH screenings in hospital inpatient or ED settings and 54% in primary care practices.

6 of 30 AHC awardees served either mostly rural counties or served exclusively rural counties.

Refer to Reliability and Empiric Validity sections for more details on testing results.

Burden for Provider: Was a provider workflow analysis conducted?

Yes

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

3224

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): Measure Score Reliability; Data Element Reliability, IRR (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

IRR (Inter-rater reliability)

Other: Sample size

1008

Other: Statistical result

0.60,0.52,0.75

Other: Interpretation of results

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.

Empiric Validity

Yes

Empiric Validity: Statistic name

Data Element Validity, Internal Consistency; Predictive Validity; Other: 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)

Empiric Validity: Sample size

60,000

Empiric Validity: Statistical result

99.8

Empiric Validity: Methods and findings

Validity Testing Statistical Result: Study 1: Sample size: 1,008. A reported social risk on the AHC and YCLS measures was strongly associated with having fair or poor self-rated health -

Source:
https://www.jfmpc.com/viewimage.asp?img=JFamMedPrimaryCare_2020_9_9_5026_296311_t6.jpg
Study 2:

Sample size:30,098. 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: Sample size: 60,000. 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-foodinsecurity-accurately-identifies-highrisk-us-adults/81A4F5E162241E289A5181A10C056125

Validity Testing Interpretation of Results: Study 1: These results are the first to suggest that both the AHC and YCLS have concurrent and predictive validity, supporting their use in healthcare settings, including by primary care physicians to engage in social risk-informed care.

Source:

PAGE 67 Top of Document Screening for Social Drivers of Health

https://www.jfmpc.com/viewimage.asp?img=JFamMedPrimaryCare_2020_9_9_5026_296311_t6.jpg

Study 2: A 2-item FI screen was sensitive, specific, and valid among low-income families with young children. The FI screen rapidly identifies households at risk for FI, enabling providers to target services that ameliorate the health and developmental consequences associated with FI.

Source:

https://pubmed.ncbi.nlm.nih.gov/20595453/

Study 3: The test characteristics of multiple two-item combinations of questions assessing food insecurity had adequate sensitivity (>97 %) and specificity (>70 %) for widespread adoption as clinical screening measures.

Source:

https://www.cambridge.org/core/journals/public-health-nutrition/article/brief-assessment-of-foodinsecurity-accurately-identifies-highrisk-us-adults/81A4F5E162241E289A5181A10C056125

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;Agreement between other gold standard and manual reviewer

Sample Size

1,000,000

Statistic Name

Sensitivity

Statistical Results

.97

Interpretation of results

The AHC screening tool used to generate the measures has been psychometrically evaluated at both the item/domain (F/H/T) and tool level and has demonstrated evidence of both reliability and validity, including predictive and concurrent validity, in healthcare settings. This includes comparison with other screening tools (e.g., Your Current Life Situation and We Care instruments) producing high kappa statistics (generally > 0.6) as well as adequate sensitivity and specificity (up to 97% sensitivity). Through AHC, the measures have been tested for 5 years with 1M+ beneficiaries in 644 clinical sites with 40% of the DOH screenings in hospital inpatient or ED settings and 54% in primary care practices.

6 of 30 AHC awardees served either mostly rural counties or served exclusively rural counties.

Refer to Reliability and Empiric Validity sections for more details on testing results.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

.33

Median performance score

0000

Minimum performance score

0000

Maximum performance score

0000

Standard deviation of performance scores

0000

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

Yes

Surveys or patient-reported outcome tools

Accountable Health Communities (AHC) Health-Related Social Needs (HRSN) Screening Tool. The AHC screening tool used to generate the measures has been psychometrically evaluated at both the item/domain (F/H/T) and tool level and has demonstrated evidence o

Section 5: Measure Contact Information

Measure Steward

CMS

Measure Steward Contact Information

Jennifer Robinson
7500 Security Blvd
Baltimore, MD 21244
jennifer.robinson@cms.hhs.gov
(410) 786-2066

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Mariel Thottam
195 Church St, 5th Floor
New Haven, CT 06510
mariel.thottam@yale.edu
(330) 417-7434

Secondary Submitter Contact Information

Rachel Johnson-DeRycke
195 Church St, 5th Floor
New Haven, CT 06510
rachel.johnson-derycke@yale.edu
(203) 497-1239

Submitter Comments

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 are projected to experience food insecurity 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.

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, he 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, 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 measure, 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.

Sources:

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/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/2021-raodwhow-deck/

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/

MUC2022-064 Hospital Harm - Pressure Injury

Program

Hospital Inpatient Quality Reporting Program; Medicare Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (CAHs)

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

The proportion of inpatient hospitalizations for patients 18 years of age or older at the start of the encounter, who suffer the harm of developing a new stage 2, stage 3, stage 4, deep tissue, or unstageable pressure injury.

Numerator

Inpatient hospitalizations for patients with a new deep tissue pressure injury (DTI) or stage 2, 3, 4, or unstageable pressure injury, as evidenced by any of the following:

A diagnosis of DTI with the DTI not present on admission.

A diagnosis of stage 2, 3, 4 or unstageable pressure injury with the pressure injury diagnosis not present on admission.

A DTI found on exam greater than 72 hours after the start of the encounter.

A stage 2, 3, 4 or unstageable pressure injury found on exam greater than 24 hours after the start of the encounter.

Numerator Exclusions

None

Denominator

Inpatient hospitalizations where the patient is 18 years of age or older at the start of the encounter.

Denominator Exclusions

Inpatient hospitalizations for patients with a DTI or stage 2, 3, 4 or unstageable pressure injury diagnosis present on admission.

Inpatient hospitalizations for patients with a DTI found on exam within 72 hours of the encounter start.

Inpatient hospitalizations for patients with a stage 2, 3, 4, or unstageable pressure injury found on exam within 24 hours of the encounter start.

Inpatient hospitalizations for patients with diagnosis of a COVID-19 infection during the encounter.

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

Other: Nursing

Measure Type

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 Health Record

If applicable, specify the data source

N/A

Description of parts related to these sources

The measure uses several structured fields within the EHR to calculate the initial population, The measure uses several structured fields within the EHR to calculate the initial population, denominator exclusion, denominator, and numerator such as:

- Admission, Discharge, Transfer
- Encounter Information
- Procedures
- Assessments
- Diagnosis and Present on Admission Indication

At what level of analysis was the measure tested?

Facility

In which setting was this measure tested?

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

N/A

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

Yes

CMIT ID

06085-E-HIQR

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Submitted

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

3498e

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

826

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?

No

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?

MUC18-107

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

Hospital, 2018-2019

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

2018-2019; HIQR 2018-2019; EHR Incentive/EH/CAH

What was the MAP recommendation in this year?

2018-2019; HIQR; Conditionally Support (Pending NQF review) 2018-2019; EHR Incentive/EH/CAH; Conditionally Support (Pending NQF review)

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

N/A; MAP recommended conditional support, pending NQF review.

MAP report page number being referenced for this year:

p. 3- 4: Considerations for Specific Program: Hospital Inpatient Quality Reporting (IQR) Program and Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (CAHs)

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?

Patient Safety Indicator 03: Pressure Ulcer Rate (PSI 03) (CMS # 00885) is used in the Hospital-Acquired Condition Reduction Program (HACRP) as a component of the NQF #0531: Patient Safety Indicator composite (PSI 90) (CMS # 03282/05537). A recalibrated version of this measure is also used in the Hospital Compare program.

NQF #0679 Percent of High-Risk Residents with Pressure Ulcers (Long Stay) (CMS # 04057) and NQF #0678 Percent of Residents or Patients with Pressure Ulcers that are New or Worsened (Short-Stay) (CMS # 04056) are used in the Nursing Home Compare and Nursing Home quality initiative programs.

Changes in Skin Integrity Post-Acute Care: Pressure Ulcer/Injury (CMS # 05852/05740/05737/05741) is additionally used in the Nursing Home Compare, Long-Term Care Hospital Compare, Long-Term Care Hospital Quality Reporting, Skilled Nursing Facility Quality Reporting, Inpatient Rehabilitation Facility Quality Reporting, Inpatient Rehabilitation Facility Compare, Home Health Services Compare, and Home Health Quality Reporting programs.

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

PSI-03 does not include stage 2 pressure injuries in the outcome, has additional exclusions to the cohort, and uses ICD-10-CM codes via claims as a data source. Hospital Harm-Pressure Injury measure is an eCQM (EHR data-only), which stakeholders and TEP have noted as a more desirable data source with more face validity for measuring pressure injuries. While there are several measures that target the reduction of hospital-acquired pressure injuries in use in various patient populations, there are no eCQMs intended for use to compare quality across acute care hospitals.

The measures NQF# 0679 and #0678 (of the same name) target different post-acute care patient populations and use chart review data from the following sources: Minimum Data Set (MDS)(SNF); Long Term Care Hospitals Continuity Assessment Record and Evaluation (LTCH-CARE) Data set (LTCH); and the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) Data set (IRF). Additionally, NQF# 0678 measure includes worsening pressure injuries and NQF#0679 population consists of only high-risk patients defined as those who are impaired in bed mobility, comatose, or suffering malnutrition.

The new Hospital Harm -Pressure Injury eCQM identifies pressure injuries using direct extraction of structured data from the EHR and will provide hospitals with reliable and timely measurement of their pressure injury rates. As these measures do not apply to the same measured entities, it should not impact data collection burden.

Changes in Skin Integrity Post-Acute Care: Pressure Ulcer/Injury only includes specific populations in the denominator (SNF residents/LTCH patients/nursing home residents/home health patients). The method for identifying patients/residents into the numerator is also different than for the other measures and our eCQM. This measure follows the same patients longitudinally to track pressure ulcer development over time, while the proposed measure would use EHR data to retrospectively track pressure injury during hospitalization.

How will this measure add value to the CMS program?

The Hospital Harm Pressure Injury measure is an eCQM. Therefore, unlike similar measures in use in CMS programs, this measure can be applied to a broader population than Medicare beneficiaries. This Hospital Harm - Pressure Injury measure is an eCQM (EHR data-only), which stakeholders and TEP have noted as a more desirable data source with more face validity for measuring pressure injuries.

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

To better understand if critical data elements used in the measure are available in a structured format and if the form in which they exist align with measure intent, we designed a web-based questionnaire and distributed the survey to 20 hospitals (3 Epic and 17 Cerner). The survey began with an inquiry into the measure's critical data elements (concepts) and ended with questions on the measure overall. The goal of the survey was to determine, within each hospital's EHR system, if critical data elements are:

- readily available in a structured format,
- from an authoritative source and/or highly likely to be correct,
- coded in a nationally accepted terminology standard or can be mapped to that terminology standard, and
- routinely collected as part of clinical care and require no or limited additional data entry from a clinician or other providers, and no EHR interface changes are needed.

Upon collecting responses, we held debrief meetings with participants to resolve ambiguities. We then conducted an EHR walkthrough with the hospitals to identify the most authoritative location within the record for each data element used within the measure to ensure that their programmed query extracted accordingly. We then translated final responses to numeric values used in the NQF scorecard.

Feasibility results were favorable with 18 of 20 hospitals reporting the ability to capture the structured data for all 10 critical measure elements and 5 supplemental data elements. For the data elements Physical Exam, Performed: Pressure Injury Deep Tissue; Physical Exam, Performed: Pressure Injury Stage 2, 3, 4 or Unstageable, we encountered some inconsistencies in workflow and ability to capture the data across the 20 hospitals. 18 of 20 hospitals captured the necessary data in structured fields and required no workflow modifications to obtain staging information. The remaining two hospitals (Epic) captured

documentation of PIs in a structured field; however, the staging information required for the measure was only captured in unstructured notes, which precluded those sites from participating in reliability and validity testing. Given that one of 3 Epic hospitals was able to capture PI staging documentation in structured fields, there are no concerns with the vendor system's ability to capture the required data. Modifications to clinical workflow and documentation practices would enable capture of data required to report the eCQM.

Method of Measure Calculation

eCQM

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

This safety eCQM captures the number of patients who experience harm in the form of a pressure injury, during their inpatient hospitalization. The incidence of pressure injuries in hospitalized patients has been estimated at 5.4 per 10,000 patient-days and the rate of hospital-acquired pressure injuries has been estimated at 8.4% (Li et al., 2020). Over 50% of reported pressure injuries in hospitals were Stage 2 or higher (Li et al., 2020). Pressure injuries commonly cause local infection, osteomyelitis, anemia, and sepsis (Brem et al., 2010), in addition to causing significant depression, pain, and discomfort to patients (Gunningberg et al., 2011). Pressure injury (defined as any Stage 3, Stage 4, and unstageable pressure ulcers acquired after admission/presentation to a healthcare setting) is considered a serious reportable event by the National Quality Forum (NQF) (Centers for Medicare and Medicaid Services, 2015). Using the EHR data from 18 hospitals and in year 2020, we found that hospital-level measure performance rates ranged from 0.0% to 2.02% (for every 1,000 qualified hospital admissions there are 20 inpatient encounters where patients suffered Pressure Injury), with a system-wide, weighted average rate equal to 1.06%. Prior studies confirm that significant variation in rates of HAPI exists between hospitals (Rondinelli et al., 2018). It is widely accepted that the risk of developing a pressure injury can be reduced through best practices. Hospital controlled factors that have been found to be significantly associated with an increase in pressure ulcer risk include infrequent repositioning (p=0.005) and number of days to bed change (OR, 2.89 [95% CI, 1.26-6.63]) (Tayyib, Coyer, and Lewis, 2016; Bly et al., 2016). High nursing workload has additionally been found to reduce risk of pressure ulcers (OR, 0.916 [95% CI, 0.855-0.980]; p=0.011) (Cremasco et al., 2013). Systematically measuring patients who develop new pressure injuries while in the hospital setting will provide hospitals with a reliable and timely measurement, to more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. This eCQM will fill a gap in measurement and provide incentives for hospital quality improvement. Although several pressure injury measures are currently in use, there are no electronic health record (EHR)-based measures intended for use in acute care hospitals. In addition, the intent of this measure is to incentivize greater achievements in reducing harms and enhance hospital performance on patient safety outcomes.

References:

Bly, D., Schallom, M., Sona, C., & Klinkenberg, D. (2016). A model of pressure, oxygenation, and perfusion risk factors for pressure ulcers in the intensive care unit. American Journal of Critical Care, 25(2), 156-154. https://doi.org/10.4037/ajcc2016840

Brem, H., Maggi, J., Nierman, D., Rolnitzky, L., Bell, D., Rennert, R., Vladeck, B. (2010). High cost of stage IV pressure ulcers. American Journal of Surgery, 200(4), 473-477. https://doi.org/10.1016/j.amjsurg.2009.12.021 Centers for Medicare & Medicaid Services. (2015). Hospital-Acquired Conditions. Retrieved January 13, 2017, from

https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html'

Cremasco, M. F., Wenzel, F., Zanei, S. S. V., & Whitaker, I. Y. (2013). Pressure ulcers in the intensive care unit: The relationship between nursing workload, illness severity and pressure ulcer risk. Journal of Clinical Nursing, 22(15-16), 2183-2191. https://doi.org/10.1111/j.1365-2702.2012.04216.x

Gunningberg, L., Donaldson, N., Aydin, C., & Idvall, E. (2012). Exploring variation in pressure ulcer prevalence in Sweden and the USA: Benchmarking in action. Journal of Evaluation in Clinical Practice, 18(4), 904-910. https://doi.org/10.1111/j.1365-2753.2011.01702.x Li, Z., Lin, F., Thalib, L., & Chaboyer, W. (2020). Global prevalence and incidence of pressure injuries in hospitalised adult patients: A systematic review and meta-analysis. International Journal of Nursing Studies, Vol. 105.

https://doi.org/10.1016/j.ijnurstu.2020.103546

Rondinelli, J., Zuniga, S., Kipnis, P., Kawar, L. N., Liu, V., & Escobar, G. J. (2018). Hospital-Acquired Pressure Injury: Risk-Adjusted Comparisons in an Integrated Healthcare Delivery System. Nurs Res, 67(1), 16-25. https://doi.org/10.1097/NNR.0000000000000258Tayyib, N., Coyer, F., & Lewis, P. (2016). Saudi Arabian adult intensive care unit pressure ulcer incidence and risk factors: A prospective cohort study. International Wound Journal, 13(5), 912-919. https://doi.org/10.1111/iwj.12406

Unintended Consequences

We did not identify any unintended consequences during eCQM development or testing. However, CMS is committed to monitoring this eCQM's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, and other negative unintended consequences for patients.

Potential unintended consequences of the Hospital Harm Pressure Injury measure include efforts that hospitals may undertake to improve performance on the measure that may create other adverse outcomes for patients. One potential unintended consequence of the measure is increased turning of certain high-risk patients in order to reduce risk of pressure injury. Increasing incentives to increase turning of patients who are at a greater risk for complications (e.g., respiratory complications, ventilator management, dislodged lines) could increase the risk of these complications.

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

2

Outline the clinical guidelines supporting this measure

Two recent evidence-based guidelines directly support the measure as follows:

- The European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel and Pan Pacific Pressure Injury Alliance (Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline. The International Guideline. EPUAP/NPIAP/PPPIA:2019)
- The American College of Physicians (ACP) (Risk Assessment and Prevention of Pressure Ulcers: A Clinical Practice Guideline From the American College of Physicians. ACP: 2015)

The EPUAP/NPIAP/PPPIA Clinical Practice Guideline (2019 edition) was developed as a collaboration between the Partner Organizations European Pressure Ulcer Advisory Panel (EPUAP), National Pressure Injury Advisory Panel (NPIAP) and the Pan Pacific Pressure Injury Alliance (PPPIA). It provides 114 evidence-based recommendations supported by a thorough review of the relevant research. The guideline is intended to apply to all clinical settings, including acute care, rehabilitation care, long term care, assisted living at home, and unless specifically stated, can be considered appropriate for all individuals with or at risk of pressure injuries. It focuses on specific evidence-based recommendations for prevention of pressure injuries through risk assessment, assessment of skin and tissue, preventive skin care, reducing progression through treatment of pressure injuries including nutrition, repositioning and early mobilization, heel pressure injuries, support surfaces and device related pressure injuries, pressure injury classification and treatment modalities.

In order to develop this guideline, a comprehensive literature review was conducted on pressure injury prevention and treatment and a rigorous methodology was used to appraise the research and make evidence-based recommendations. The research evidence was summarized and evaluated using evidence-to-decision frameworks. Where sufficient research evidence was available, recommendations to guide clinical practice were developed. In areas without sufficient research, good practice statements were developed to promote comprehensive care. There were 699 health professionals, industry representatives, peak body organizations, researchers, policy makers, patient consumers, and informal caregivers who reviewed and/or commented on the document.

The guideline includes discussion of the science, followed by 114 recommendations and 62 good practice statements to guide practice in risk assessment, pressure injury prevention and treatment, and issues in implementing best practice.

This is the most recent set of clinical practice guidelines applicable to the measure topic and developed by American national and international healthcare professional organizations. They were graded as described in sections 110 through 112.

The primary clinical guideline described in section 108 below was based on 21 cited studies. Two studies provided evidence to support a recommendation to implement structured skin care regimen that includes regular cleansing (particularly after episodes of incontinence). A low quality Level 2 study found that a structured hygiene program was associated with a lower incidence of pressure injuries than standard care. A low quality level 4 observational study noted that skin was assessed as being healed or healing when a structured skin care regimen was implemented. A moderate quality Level 1 study reported significant reductions in erythema and broken skin when a pH-balanced (pH 5.5) foam cleanser was used, as compared to standard hospital soap. The structured skin care regimen reported in the low quality level 2 study also included replacing soap with a pH balanced (pH not reported) foam cleanser.

For a detailed description of the ACP guidelines, please see Section 1.2 in the 2022 MUC List Attachment: Pressure Injury.

Name the guideline developer/entity

The European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel, and Pan Pacific Pressure Injury Alliance (EPUAP/NPIAP/PPPIA)

Publication year

2019

Full citation +/- URL

European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel and Pan Pacific Pressure Injury Alliance. Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline. The International Guideline. Emily Haesler (Ed.). EPUAP/NPIAP/PPPIA: 2019.

https://www.internationalguideline.com/

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.

3.1 Implement a skin care regimen that includes:

- Keeping the skin clean and appropriately hydrated
- Cleansing the skin promptly after episodes of incontinence
- Avoiding use of alkaline soaps and cleansers
- Protecting the skin from moisture with a barrier product.

Strength of Recommendation: Strong Positive

Level of Evidence: B2

We have identified this recommendation statement due to its high level of evidence, relevancy to prevention of pressure injuries, and strong recommendation. For a list of additional EPUAP/NPIAP/PPPIA: (2019) and ACP: (2015) clinical practice guidelines that are relevant to this measure concept, see Section 1 in the 2022 MUC List Attachment: Pressure Injury.

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?

Strengths of Recommendation

- To achieve a strong positive (do it) or strong negative (don't do it) recommendation, 100% of votes must be cast in the same direction (positive or negative), with at least 70% voting for a strong recommendation, and 0% voting in the opposite direction.
- To achieve a weak positive (probably do it) or weak negative (probably don't do it) recommendation, at least 70% of votes must be cast in the same direction (positive or negative), and less than 20% voting in the opposite direction.
- Any other combination of voting results is 'no specific recommendation.'

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?

Other (enter here):: Levels 1-5

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

The levels of evidence for individual intervention studies were assigned to each study containing direct evidence, using the Level 1-5 classification system adapted from The Joanna Briggs Institute and provided in the following section. Strength of evidence ratings (A, B1, B2, C) for each recommendation was also adapted from the National Health and Medical Research Council (NHMRC) as outlined in the section below.

References:

Joanna Briggs Institute. (2014). Reviewers' manual 2014. Adelaide: Joanna Briggs Institute.

NHMRC GAR consultants, NHMRC additional levels of evidence and grades for recommendations for developers of guidelines 2009, National Health and Medical Research Council Canberra.

Level of Evidence for Intervention Studies

Level 1: Experimental Designs

- Randomized trial

Level 2: Quasi-experimental Design

- Prospectively controlled study design
- Pre-test post-test or historic/retrospective control group study

Level 3: Observational-analytical Designs

- Cohort study with or without control group
- Case-controlled study

Level 4: Observational-descriptive Studies (no control)

- Observational study with no control group
- Cross-sectional study
- Case series (n=10+)

Level 5: Indirect Evidence

- Studies in normal human subjects, human subjects with other types of chronic wounds, laboratory studies using animals, or computational models

Level of Evidence for Diagnostic Studies

Level 1

- Individual high quality (cross sectional) studies according to the quality assessment tools with consistently applied reference standard and blinding among consecutive persons

Level 2

- Non-consecutive studies or studies without consistently applied reference standards

Level 3

- Case-control studies or poor or non-independent reference standard

Level 4

- Mechanism-based reasoning, study of diagnostic yield (no reference standard). Low and moderate quality cross sectional studies

Level of Evidence for Prognostic Studies

Level 1

- A prospective cohort study

Level 2

- Analysis of prognostic factors amongst persons in a single arm of a randomized controlled trial

Level 3

- Case-series or case-control studies, or low quality prognostic cohort study, or retrospective cohort study

Strengths of Evidence

A

- More than one high quality Level I study providing direct evidence

- Consistent body of evidence

B1

- Level 1 studies of moderate or low quality providing direct evidence
- Level 2 studies of high or moderate quality providing direct evidence
- Most studies have consistent outcomes and inconsistencies can be explained

B2

- Level 2 studies of low quality providing direct evidence
- Level 3 or 4 studies (regardless of quality) providing direct evidence
- Most studies have consistent outcomes and inconsistencies can be explained

C

- Level 5 studies (indirect evidence) e.g., studies in normal human subjects, humans with other types of chronic wounds, animal models
- A body of evidence with inconsistencies that cannot be explained, reflecting genuine uncertainty surrounding the topic

Good practice statement

- Statements by the Guideline Governance Group (GGG) that are not supported by a body of evidence as listed above but considered significant for clinical practice

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?

Moderate or similar

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

3.1 Implement a skin care regimen that includes:

- Keeping the skin clean and appropriately hydrated
- Cleansing the skin promptly after episodes of incontinence
- Avoiding use of alkaline soaps and cleansers
- Protecting the skin from moisture with a barrier product.

Strength of Recommendation: Strong Positive

Level of Evidence: B2

We have identified this recommendation statement due to its high level of evidence, relevancy to prevention of pressure injuries, and strong recommendation. For a list of additional EPUAP/NPIAP/PPPIA: (2019) and ACP: (2015) clinical practice guidelines that are relevant to this measure concept, see Section 1 in the 2022 MUC List Attachment: Pressure Injury.

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?

Yes

Estimated Impact of the Measure: Estimate of Annual Denominator Size

30,071,363

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); Not conceptually or empirically indicated (enter here):: Higher risk patients require more intervention to prevent Pressure Injuries, but there is no empirically observed association between pre-existing risk and perceived avoidability.

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.

10

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

10

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?

20

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;Random Split-Half Correlation

Signal-to-Noise: Name of statistic

Adam's signal-to-noise ratio (SNR)

Signal-to-Noise: Sample size

18

Signal-to-Noise: Statistical result

0.97

Signal-to-Noise: Interpretation of results

Evaluating against the conventional standards, test statistics showed almost perfect measure score-level reliability and suggested that the measure, as currently specified, can distinguish performance of one hospital from another. A sizable portion of th

Random Split-Half Correlation: Name of statistic

Intra-class correlation coefficient (ICC) via the split-half sample approach

Random Split-Half Correlation: Sample size

18

Random Split-Half Correlation: Statistical result

0.916

Random Split-Half Correlation: Interpretation of results

Evaluating against the conventional standards, test statistics showed strong measure score-level reliability and suggested that the measure, as currently specified, can distinguish performance among hospitals. A sizable portion of the variability between hospitals appears to be attributable to real differences in quality of care.

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

Convergent validity evaluated by the Spearman's rank correlation coefficient. By correlating hospital performance in PI with their performance in the independently collected and NQF-endorsed measures that share a similar conceptual framework regarding how safe care is produced and defined, we assessed measure score level validity using the correlation coefficient that ranges from 0 to 1.

Empiric Validity: Sample size

18

Empiric Validity: Statistical result

-0.68

Empiric Validity: Methods and findings

Convergent validity is a concept that refers to if multiple measures of an underlying same concept are positively correlated. For this exercise, we collected pilot sites' patient safety outcome from a set of related measures (e.g., healthcare associated infections and nursing care) in Hospital Care Compare (data.cms.gov) and estimated the Spearman's rank correlation coefficients between HH PI and each of the related measures at the hospital level. Positive (or negative, pending context) correlations then

provide support for the score level validity as they document a similar quality construct of patient safety. Given that PI is nursing care sensitive, we hypothesized that there will be an inverse relationship between hospital performance in PI and patients' rating of hospital care around nursing (e.g., nurse communication, staff responsiveness, etc.)

Using Hospital Compare Data from data.cms.gov, the Spearman rank correlation between pilot sites' PI scores and 12 quality measures reflecting patients' perspectives of hospital care provide evidence for moderate measure construct validity. For example, higher rate of PI is inversely related to patients' perspective of hospital care, such as nurse communications, staff responsiveness, discharge information, and overall rating of hospital care. In row 41 (Empiric Validity: Statistical result), we report the rank correlation between our PI measure and the HCAHPS measure about staff responsiveness (H-COMP-3-STAR-RATING extracted from Hospital Compare at data.cms.gov). We provide a full list of measures for which we calculated Spearman's rank correlation coefficients in Tables 7-8 (Section 3) of the 2022 MUC List Attachment: Pressure Injury.

Empiric Validity: Interpretation of results

Yes

Face Validity

Yes

Face Validity: Number of voting experts and patients/caregivers

9

Face Validity: Result

9

Patient/Encounter Level Testing

Yes

Type of Analysis

Agreement between eCQM and manual reviewer

Sample Size

310

Statistic Name

Other (enter here):: Percent Agreement, Kappa, Positive Predicative Value, and Sensitivity (listed below in order)

Statistical Results

0.98,0.8,0.97,0.98

Interpretation of results

Overall, testing results indicate strong measure data element reliability and validity.

The minimum Kappa (0.8) occurred to one measure critical data element and the disagreement happened in one pilot site. Kappas for the remaining critical data elements are no less than 0.9 and generally equal to 1.0.

The modes of PPV and sensitivity are 1.0 (or 100%) even if the lowest value is 0.97 and 0.98, respectively.

Across the 18 pilot sites, measure denominator ranged from 553 to 38,476 qualified inpatient encounters and measure rate ranged from zero to a high of 2.02 per 100 qualified inpatient encounters. The wide variability indicates ample room for quality improvement in hospital inpatient setting. Exhibit 1 (Section 3) 2022 MUC List Attachment: Pressure Injury translates the distribution of observed measure rate and its 95% confidence interval into a histogram and further displays the overall weighted average measure rate (green dashed horizontal line). Testing data showed that several hospitals' performance rates are consistently below the overall mean, while a few others are above that mean.

Comparing to the variation in measure performance rates, variation in Kappa, PPV, and sensitivity is much more limited. The minimum Kappa equals 0.8, but the second smallest Kappa has reached 0.91, which denotes very strong concordance. The minimum PPV equals 0.97 and the second smallest PPV equals 0.98. The median and mode of PPV are both 1.0. The minimum sensitivity equals 0.98 and the second smallest sensitivity equals 1.0. The median and mode of sensitivity are both 1.0. Limited variation substantiates the conclusion that the measure has strong data element-level reliability and validity.

In Tables 9-14 (Section 3) of the 2022 MUC List Attachment: Pressure Injury we provide the list of data elements tested, along with their percent agreement, Kappa, or PPV and sensitivity.

Measure performance – Type of Score

Proportion

Measure Performance Score Interpretation

Lower score is better

Mean performance score

1.06

Median performance score

0.61

Minimum performance score

0

Maximum performance score

2.02

Standard deviation of performance scores

0.56

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

Donta Henson
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

Hannah Klein
1400 Crystal Drive
10th floor
Arlington, VA 22202
hklein@air.org
(206) 939-4978

Secondary Submitter Contact Information

Anna Michie
1400 Crystal Drive
10th floor
Arlington, VA 22202
amichie@air.org
(443) 259-5180

Submitter Comments

N/A

MUC2022-084 COVID-19 Vaccination Coverage among Healthcare Personnel (HCP) (2022 revision)

Program

Ambulatory Surgical Center Quality Reporting Program; Hospital Inpatient Quality Reporting Program; Hospital Outpatient Quality Reporting Program; Hospital Value-Based Purchasing Program; Hospital-Acquired Condition Reduction Program; Inpatient Psychiatric Facility Quality Reporting Program; Inpatient Rehabilitaion Facility Quality Reporting Program; Long-Term Care (LTC) Hospital Quality Reporting Program; Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program; End-Stage Renal Disease (ESRD) Quality Incentive Program

Section 1: Measure Information

Measure Specifications and Endorsement Status

Measure Description

Percentage of healthcare personnel who are considered up to date with recommended COVID-19 vaccines.

Numerator

The numerator for this measure consists of the cumulative number of HCP in the denominator population who are considered up to date with recommended COVID-19 vaccines.

Facilities should refer to the definition of up to date as of the first day of the quarter.

https://www.cdc.gov/nhsn/pdfs/hps/covidvax/UpToDateGuidance-May2022-508.pdf

As of April 1, 2022, up to date includes:

1. Individuals who received their second dose in a two-shot primary vaccination series, (Pfizer-BioNTech or Moderna vaccines) less than 5 months ago

2. Individuals who received a J&J/Janssen as their primary vaccination less than 2 months ago

3. Individuals who have received a primary series and one booster dose when recommended.

Numerator Exclusions

None

Denominator

The target population is the number of healthcare personnel (HCP) eligible to work in the healthcare facility for at least one day during the one-week data collection reporting period, excluding persons with contraindications to COVID-19 vaccination.

This measure includes at least one week of data collection a month for each of the 3 months in a quarter.

The denominators are reported by aggregating the categories below:

There are four categories of HCP:

1. Employees: includes all persons who receive a direct paycheck from the reporting facility (i.e., on the facility's payroll).
2. Licensed independent practitioners (LIPs): This includes physicians (MD, DO), advanced practice nurses, and physician assistants only who are affiliated with the reporting facility who do not receive a direct paycheck from the reporting facility.
3. Adult students/trainees and volunteers: This includes all students/trainees and volunteers aged 18 or over who do not receive a direct paycheck from the reporting facility.
4. Other contract personnel: Facilities may also report on individuals who are contract personnel. However, reporting for this category is optional. Contract personnel are defined as persons providing care, treatment, or services at the facility through contract who do not fall into any of the above-mentioned denominator categories.

Denominator Exclusions

Denominator-eligible individuals with contraindications to COVID-19 vaccination. Medical contraindications are listed in a vaccine's FDA authorization or labeling and include severe allergic reaction. The current list of contraindications as well as exclusions may be found at https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html and includes:

1. Severe allergic reaction (e.g., anaphylaxis) after a previous dose or to a component of the COVID-19 vaccine
2. Known diagnosed allergy to a component of the COVID-19 vaccine

Denominator Exceptions

None

State of development

Field (Beta) Testing

State of Development Details

Beta testing was conducted by assessing if the collection of information on additional/booster vaccine doses received by healthcare personnel (HCP) was feasible, as information on receipt of booster vaccine doses is required for determining if HCP are up to date with the current COVID-19 vaccination recommendations.

Feasibility was assessed by calculating proportion of facilities which reported additional/booster doses of COVID-19 vaccine.

This assessment was conducted in the following facility types based on vaccine coverage data for the first quarter of 2022 (January - March) reported through the National Healthcare Safety Network (NHSN):

Ambulatory Surgery Centers (ASCs)

Dialysis Centers

Hospitals

Inpatient Psychiatric Facilities (IPFs)

Inpatient Rehabilitation Facilities (IRFs)

Long Term Acute Care (LTACs)

Skilled Nursing Facilities (SNFs)

Reliability and validity testing of the measure as specified is planned based on vaccine coverage data for the third quarter of 2022 (July- September) reported through the National Healthcare Safety Network (NHSN) in the same facility types listed above.

What is the target population of the measure?

Healthcare Personnel

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

Other: All Healthcare Personnel

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 Clinical Data (non-EHR);Electronic Health Record;Paper Medical Records;Registries;Other: The source may vary by facility. Data may be collected from electronic sources or paper-based sources. It may be obtained from existing records or a system specifically designed for COVID-19 vaccination tracking.

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?

Facility

In which setting was this measure tested?

Not yet tested

Multiple Scores

No

What one healthcare domain applies to this measure?

Safety

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

N/A

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

Yes

CMIT ID

08062-C-ASCQR, 08062-C-HOQR, 08062-C-IRFQR, 08062-C-ESRDQIP, 08062-C-PCHQR, 08062-C-SNFQRP, 08062-C-HIQR, 08062-C-LTCHQR, 08062-X-LTCHC

Alternate Measure ID

N/A

What is the endorsement status of the measure?

Submitted

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

3636

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

No

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

Numerator

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

The CDC recommendations for COVID-19 vaccination have changed since the initial formulation of the measure COVID-19 Vaccination Coverage among Healthcare Personnel (CMT 08062) which was originally titled: SARS-CoV-2 Vaccination Coverage Among Healthcare P

If endorsed: Year of most recent CDP endorsement

2020

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?

Yes

Previous Measure Information

In what prior year was this measure published?

2020

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

2020: MUC20-0044: SARS-CoV-2 Vaccination Coverage Among Healthcare Personnel

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

2020 Coordinating Committee, Hospital, Post-Acute Care/Long-Term Care, Rural Health

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

2020-2021 Hospital Outpatient Quality Reporting Program (Hospital OQR); Hospital Inpatient Quality Reporting Program (Hospital IQR); Ambulatory Surgical Center Quality Reporting Program (ASCQR); Inpatient Psychiatric Facility Quality Reporting Program (IPFQR); PPS-Exempt Cancer Hospital Quality Reporting Program (PCHQR); End-Stage Renal Disease Quality Improvement Program (ESRD QIP); Inpatient Rehabilitation Facility Quality Reporting Program (IRF QRP) SARS-CoV-2 Measure; Long-Term Care Hospital Quality Reporting Program (LTCH QRP) SARS-CoV-2 Measure; Skilled Nursing Facility Quality Reporting Program (SNF QRP) SARS-CoV-2 Measure

What was the MAP recommendation in this year?

Conditional Support

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

N/A

MAP report page number being referenced for this year:

MAP Report for 2020, pages 24-25

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

Measure currently used in a CMS program, but the measure is undergoing substantial change

Range of years this measure has been used by CMS Programs

Quality Reporting Programs indicated in question below, 2022 to present (except Skilled Nursing Facility Quality Reporting Program [last quarter of 2021-present])

What other federal programs are currently using this measure?

Ambulatory Surgical Center Quality Reporting Program;Hospital Inpatient Quality Reporting Program;Inpatient Psychiatric Facility Quality Reporting Program;Inpatient Rehabilitation Facility Quality Reporting Program;Long-Term Care (LTC) Hospital Quality Re

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?

Other: National Healthcare Safety Network (NHSN)

Stratification

No

Feasibility of Data Elements

Some data elements are in defined fields in electronic sources

Feasibility Assessment

CMS quality reporting programs have already required facilities to report data on COVID-19 vaccination coverage among healthcare personnel (HCP) for primary vaccination. Feasibility of reporting additional/booster doses of vaccine is evident by the proportion of facilities nationwide that have

already reported vaccination additional/booster coverage data to CDC's National Healthcare Safety Network (NHSN).

Even though the deadline to report vaccination coverage data for the first quarter of 2022 is not until August 2022 (except for dialysis and nursing homes which have additional reporting requirements), the proportions of facilities already reporting vaccine coverage data including additional/booster coverage as of May 2022 are:

- Ambulatory Surgery Centers (ASCs): 64.4%
- Dialysis Centers: 97.0%
- Hospitals: 74.6%
- Inpatient Psychiatric Facilities (IPFs): 74.3%
- Inpatient Rehabilitation Facilities (IRFs): 63.9%
- Long Term Acute Care (LTACs): 90.3 %
- Skilled Nursing Facilities (SNFs): 99.2%

These high reporting rates indicate reporting the measure is feasible.

Method of Measure Calculation

Other (enter here)::
Data Collection:

1. Identify all healthcare personnel (HCP) eligible to work during the selected week. The week always begins on a Monday at 12:00 midnight and ends on Sunday at 11:59 PM.

2. Categorize all eligible HCP into one of four HCP categories (see "Measure Information" #012)

3. Among eligible HCP, identify those who are considered up to date with recommended COVID-19 vaccines.

4. Among eligible HCP who are not considered up to date with recommended COVID-19 vaccines, identify those who have a contraindication to COVID-19 vaccination.

5. Among eligible HCP who are not considered up to date with recommended COVID-19 vaccines, and who do not have a contraindication to COVID-19 vaccination, identify those who have refused or declined vaccination.

6. Among eligible HCP are not considered up to date with recommended COVID-19 vaccines, identify those whose COVID-19 vaccination status cannot be determined.

Measure Calculation:
The weekly coverage rate is the numerator divided by the denominator (minus exclusions) for a particular week:

1. For each one-week period, tabulate the denominator by summing the number of HCP in each of the categories of HCP minus the number of HCP with contraindications to COVID-19 vaccination.

2. Calculate the weekly COVID-19 up to date vaccination coverage percentage by dividing the number of HCP in the denominator who are considered up to date with recommended COVID-19 vaccines by the

number of HCP in the denominator and multiplying by 100.
The measure is reported for a quarter (3-month period). Quarterly up to date COVID-19 vaccination coverage is determined by selecting one weekly coverage rate per month, then averaging 3 weekly coverage rates (one week from each of the 3 months in the quarter).
For facilities that report more than one week per month, the latest week of data for the reporting month will be used.

Hybrid measure: Methods of measure calculation

N/A

Evidence of Performance Gap

There are clinically significant differences in booster/additional dose vaccination coverage rates among facilities, indicating that facilities have room for improvement and implementing the revised measured would be meaningful.

The following performance scores are the reported booster/additional dose coverage rates for the first quarter of 2022 (January 1 - March 31, 2022) by facility type:

ASCs: median 34.0%; interquartile range 16.4% - 55.6%

Dialysis Centers: median 14.7%; interquartile range 5.4% - 31.3%%

Acute Care Hospitals: median 22.5%; interquartile range 9.1% - 38.7%

IPFs: median 19.1%; interquartile range 8.7% - 37.9%

IRFs: median 20.3%; interquartile range 8.9% - 37.7%

LTACs: median 22.6%; interquartile range 10.8% - 36.9%

SNFs: median 31.8%; interquartile range 18.9% - 49.7%

Unintended Consequences

None

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

Observational studies of real-world effectiveness of COVID-19 vaccination

Summarize the evidence

The CDC recommendations for COVID-19 vaccination have changed since the initial formulation of the measure COVID-19 Vaccination Coverage among Healthcare Personnel (CMT 08062) which was originally titled: SARS-CoV-2 Vaccination Coverage Among Healthcare Personnel (MUC20-0044). It is now recommended that individuals stay up to date with COVID-19 vaccination (https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html).

This revision of measure to include reporting of up to date vaccination is informed by a search of the published literature. There are no published data on the impact of reporting up to date COVID-19 coverage reporting among healthcare workers; however, the following real-world observational data support the positive impact of COVID-19 vaccination, healthcare personnel vaccination, and additional/booster COVID-19 vaccine.

1. COVID-19 vaccine uptake in the U.S. is associated with reduced COVID-19 incidence and mortality:

- -Suthar AB, Wang J, Seffren V, et al. Public health impact of covid-19 vaccines in the US: observational study. BMJ 2022 Apr 27;377:e069317. doi: 10.1136/bmj-2021-069317.
- December 2020-December 2021 cross-sectional analysis of US county level surveillance and vaccine administration data from 48 states.
- It was observed that 10% improvement in vaccination coverage was associated with an 8% reduction in mortality rates and a 7% reduction in incidence.

2. Among U.S. healthcare workers, COVID-19 vaccine effectiveness has been found to be high:

- Pilishvili T, Gierke R, Fleming-Dutra KE, et al. Effectiveness of mRNA Covid-19 Vaccine among U.S. Health Care Personnel. N Engl J Med 2021 Dec 16;385(25):e90. doi: 10.1056/NEJMoa2106599.
- This was a test-negative case-control study of US healthcare personnel from 25 states conducted from December 2020-May 2021.
- Vaccine effectiveness against infection was 88.8% (95% CI, 84.6 to 91.8) for BNT162b2 vaccine and 96.3% (95% CI, 91.3 to 98.4) for the mRNA-1273 vaccine.

3. U.S. Healthcare worker COVID-19 vaccination was associated with reduced patient COVID-19 infections and deaths:

- McGarry BE, Barnett ML, Grabowski DC, et al. Nursing Home Staff Vaccination and Covid-19 Outcomes. N Engl J Med 2022 Jan 27;386(4):397-398. doi: 10.1056/NEJMc2115674.
- This study was a cross-sectional analysis of US nursing home staff vaccination and resident infection data reported to the US Centers for Medicare and Medicaid Services from June 2021-August 2021.
- In the presence of high community prevalence of Covid-19, nursing homes with low staff vaccination coverage had COVID-19 infection and death rates 132% and 195% higher, respectively, than those with high staff vaccination coverage.

4. With the COVID-19 Omicron variant, despite continued protection against invasive mechanical ventilation and death, a decrement in COVID-19 vaccine effectiveness has been observed for Emergency Department visits and hospitalizations:

- Tenforde MW, Self WH, Gaglani M, et al. Effectiveness of mRNA Vaccination in Preventing COVID-19-Associated Invasive Mechanical Ventilation and Death - United States, March 2021-January 2022. MMWR Morb Mortal Wkly Rep 2022 Mar 25;71(12):459-465. doi: 10.15585/mmwr.mm7112e1.
- This study was a case-control study of mRNA vaccine effectiveness (VE) against COVID-19 associated invasive mechanical ventilation (IMV) and in-hospital death among adults hospitalized at 21 US hospitals from March 2021-January 2022.
- VE against IMV or in-hospital death was 90% overall; 88% for 2 doses and 94% for 3 doses, and 94% for 3 doses during the Omicron-predominant period.
- Ferdinands JM, Rao S, Dixon BE, et al. Waning 2-Dose and 3-Dose Effectiveness of mRNA Vaccines Against COVID-19Associated Emergency Department and Urgent Care Encounters and Hospitalizations Among Adults During Periods of Delta and Omicron Variant Predominance VISION Network, 10 States, August 2021-January 2022. MMWR Morb Mortal Wkly Rep 2022 Feb 18;71(7):255-263. doi: 10.15585/mmwr.mm7107e2.
- This study was a test-negative case-control study evaluating VE against COVID-19 emergency department/urgent care (ED/UC) visits and hospitalizations among adults at sites across 10 states from August 2021-January 2022.
- During the Omicron period, VE against ED/UC visits was 87% in the first two months after a 3rd dose and decreased to 66% among those vaccinated 4-5 months prior; VE against hospitalizations was 91% during the first two months following a 3rd dose and decreased to 78% >= months after a 3rd dose.

5. Additional or booster dosing has been associated with reduced infections in both patients and healthcare workers:

- Prasad N, Derado G, Nanduri SA, et al. Effectiveness of a COVID-19 Additional Primary or Booster Vaccine Dose in Preventing SARS-CoV-2 Infection Among Nursing Home Residents During Widespread Circulation of the Omicron Variant -United States, February 14- March 27, 2022. MMWR Morb Mortal Wkly Rep 2022 May 6;71(18):633-637. doi: 10.15585/mmwr.mm7118a4.
- This report is a cross-sectional analysis of data reported to CMS from 15,000 nursing homes from January-March 2022.
- Compared with primary series vaccination only, an additional or booster dose provided greater protection (relative VE = 46.9%) against SARS-CoV-2 infection during Omicron variant predominance.
- Oster Y, Benenson S, Nir-Paz R, et al. The effect of a third BNT162b2 vaccine on breakthrough infections in health care workers: a cohort analysis. Clin Microbiol Infect 2022 May;28(5):735.e1-735.e3. doi: 10.1016/j.cmi.2022.01.019. Epub 2022 Feb 7.
- This two-hospital cohort study evaluating COVID-19 infection rate among healthcare workers (HCWs) receiving a 3rd vaccine dose (booster) compared with those who had received only a two-dose regimen in August 2021.
- HCWs who received only the two-dose regimen had an infection rate of 21.4% (85 of 398), compared with 0.7% (35/4973; relative risk 30) among the boosted group.

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

9999

Type of Evidence to Support the Measure

Other (enter here):: Individual peer-reviewed observational studies of waning effectiveness indicate "up to date" metric is needed.

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 conceptually or empirically indicated

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

No

Reliability: Type of Reliability Testing

N/A

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

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

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

Proportion

Measure Performance Score Interpretation

Higher score is better

Mean performance score

9999

Median performance score

9999

Minimum performance score

9999

Maximum performance score

9999

Standard deviation of performance scores

9999

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 Disease Control and Prevention

Measure Steward Contact Information

Natasha Poudyal
1600 Clifton Road NE
Atlanta, GA 30329
qpp1@cdc.gov
(707) 975-9356

Long-Term Measure Steward

N/A

Long-Term Measure Steward Contact Information

N/A

Primary Submitter Contact Information

Natasha Poudyal
1600 Clifton Road NE
Atlanta, GA 30329
qpp1@cdc.gov
(707) 975-9356

Secondary Submitter Contact Information

Andrew Geller
1600 Clifton Road NE
Atlanta, GA 30329
wia0@cdc.gov
(404) 498-0639

Submitter Comments

Performance Scores by Facility Type are provided here. These have been indicated in the following sections by "9999" since only numerical values were allowed in the entry.

Row 064: Mean performance score

The scores in this subsection (Measure Performance) evaluate reporting of additional information that previously was not required to be collected under the measure SARS-CoV-2 Vaccination Coverage Among Healthcare Personnel (MUC20-0044). Since the implementation of MUC-0044, CDC has recommended that individuals stay up to date with COVID-19 vaccination, which requires vaccine booster data (https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html).

The following performance scores are the reported booster/additional dose coverage rates for the first quarter of 2022 (January 1- March 31, 2022) by facility type:

ASCs: 38.3%

Dialysis Centers: 21.9%

Acute Care Hospitals: 26.3%

IPFs: 25.1%

IRFs: 25.4%

LTACs: 25.3%

SNFs: 36.2%

Row 065: Median performance score

ASCs: 34.0%

Dialysis Centers: 14.7%

Acute Care Hospitals: 22.5%

IPFs:19.1%

IRFs: 20.2%

LTACs: 22.6%

SNFs: 31.8%

Row 066: minimum performance score

Dialysis centers: 0%

Acute Care Hospitals: 0%

IPFs: 0%

IRFs: 0%

LTACs: 0%

SNFs: 0%

Row 067: maximum performance score

ASCs: 100%

Dialysis Centers:100%

Acute Care Hospitals: 93.1%

IPFs: 95.1%

IRFs: 96.8%

LTACs: 96.2%

SNFs: 100%

Row 068: standard deviation of performance scores

ASCs: 27.0%

Dialysis Centers: 22.2%

Acute Care Hospitals: 21.2%

IPFs: 21.3%

IRFs: 21.2%

LTACs: 18.5%

SNFs: 22.7%

Row 075: estimated impact of the measure: estimate of annual denominator size

ASCs: 1,096 facilities; 92,820 HCP

Dialysis Centers: 7,369 facilities; 217,348 HCP

Acute Care Hospitals: 2,589 facilities; 5,078,202 HCP

IPFs: 760 facilities; 258,190 HCP

IRFs: 769 facilities; 247,321 HCP

LTACs: 329 facilities; 91,470 HCP

SNFs: 14,250 facilities; 1,971,405 HCP

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

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 references, 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://pubmed.ncbi.nlm.nih.gov/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 references, 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 Stree
Suite 4400
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 Stree
Suite 4400

Raleigh, NC 27601
zbutt@phreesia.com
(773) 209-8183

Secondary Submitter Contact Information

Zeeshan Butt
434 Fayetteville Stree
Suite 4400
Raleigh, NC 27601
zbutt@phreesia.com
(773) 209-8183

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.