NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Summary

Behavioral Health and Substance Use Standing Committee – Measure Evaluation Web Meeting

The National Quality Forum (NQF) convened the Behavioral Health and Substance Use Standing Committee for web meetings on January 29 and 31 and February 5, 2020 to evaluate seven measures.

Welcome, Introductions, and Review of Meeting Objectives

NQF welcomed the Standing Committee and participants to the web meeting. NQF staff reviewed the meeting objective, which was to determine recommendations for endorsement on the candidate measures. Committee members each introduced themselves and disclosed any conflicts of interests. Co-chair Harold Pincus, MD was recused on measures NQF 2800 and NQF 2801, and Mady Chalk, PhD, MSW was recused on NQF 3175. Quorum was met during the January 29 and 31 web meetings and during the first half of the February 5 web meeting.

Topic Area Introduction and Overview of Evaluation Process

NQF staff provided an overview of the consensus development process, including the procedures for measure discussion, voting, and achieving consensus. NQF staff reviewed the role of the Scientific Methods Panel (SMP). The SMP consists of methodological experts to help ensure a higher-level evaluation of the scientific acceptability of complex measures. For this project, the SMP reviewed the scientific acceptability of two measures, NQF 3538 All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries Who May Benefit from Integrated Physical and Behavioral Health Care and NQF 3492 Acute Care Use Due to Opioid Overdose.

Measure Evaluation

During the meeting, the Behavioral Health and Substance Use Standing Committee evaluated two maintenance measures, four new measures, and one measure for an early maintenance review. For the early maintenance review, NQF 3175 Continuity of Pharmacotherapy for Opioid Use Disorder was assessed to determine whether endorsement should be expanded to the individual clinician and clinician group/practice levels of analysis. Three maintenance measures and two new measures were recommended for endorsement. There were two new measures where consensus was not reached. These measures will be reviewed again during the post-comment call on April 22, 2020.

A summary of the Committee deliberations will be compiled and provided in the draft technical report. NQF will post the draft technical report on March 11, 2020 for public comment on the NQF website. The draft technical report will be posted for 30 calendar days.

Rating Scale: H – High; M – Medium; L – Low; I – Insufficient; NA – Not Applicable

2800 Metabolic Monitoring for Children and Adolescents on Antipsychotics (National Committee on Quality Assurance)

Measure Steward/Developer Representatives at the Meeting

Emily Morden (National Committee on Quality Assurance)

Standing Committee Votes

- Evidence: H-7; M-8; L-0; I-0
- Performance Gap: H-5; M-10; L-0; I-0
- Reliability: H-2; M-13; L-0; I-0
- Validity: H-2; M-11; L-2; I-0
- Feasibility: H-4; M-11; L-0; I-0
- Use: Pass-15; No Pass-0
- Usability: H-2; M-13; L-0; I-0

Standing Committee Recommendation for Endorsement: Yes-15; No-0

The Standing Committee recommended the measure for continued endorsement. This maintenance measure assesses the percentage of children and adolescents 1-17 years of age who had two or more antipsychotic prescriptions who had metabolic testing (i.e., glucose and cholesterol monitoring). Since the last review in 2016, the measure now has separate rates for each metabolic test in addition to the combined rate of receiving both tests. The measure also now combines two of the age strata. The Committee agreed that this metabolic monitoring measure is an important part of managing patient risk and is supported by evidence from guidelines. The Committee encouraged the developer to consider including more recent literature in the evidence presentation for the next submission,.

The Committee accepted the results of the score-level reliability and validity testing. There was discussion about whether the measures selected for construct validity, measures of adolescent well-care visits and well-child visits, are the appropriate measures to expect strong correlations.

The measure is used by commercial health plans and is in the Medicaid Child Core Set for 2020. Regarding usability, the Committee agreed there is opportunity for improvement, but questioned why there has not been more performance improvement over the past four years. Some members felt that the measure should not be very difficult to impact. The developer shared that splitting out results by testing type may provide better information to inform improvement. Other members expressed that required public reporting may further incentivize improvement.

2801 Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics (National Committee on Quality Assurance)

Measure Steward/Developer Representatives at the Meeting

Emily Morden (National Committee on Quality Assurance)

Standing Committee Votes

- Evidence: H-2; M-13; L-0; I-0
- Performance Gap: H-3; M-12; L-0; I-0
- Reliability: H-1; M-14; L-0; I-0
- Validity: H-13; M-1; L-1; I-0
- Feasibility: H-10; M-5; L-0; I-0
- Use: Pass-15; No Pass-0

- Usability: H-3; M-12; L-0; I-0

Standing Committee Recommendation for Endorsement: Yes-15; No-0

The Standing Committee recommended the measure for continued endorsement. This maintenance measure assesses the percentage of children 1-17 years with a new antipsychotic prescription without an indication who had documentation of psychosocial care as first-line treatment. The developer acknowledged that since 2016, the measure has been updated to combine the two lower age strata but is otherwise the same.

It was noted there are no new studies that contradict the evidence presented, and several Committee members emphasized that this topic area needs careful attention and monitoring. It was clarified that the measure allows for psychosocial care to be provided up to 30 days after the first prescription. At least one member noted that a strength of the measure is that it appropriately excludes serious mental illness. The Committee agreed there was significant room for improvement and that score-level reliability and validity testing results supported the measure's scientific acceptability. Construct validity was performed using four measures that reflect coordinated care across settings.

Regarding use, the measure is publicly reported in various applications. A potential unintended consequence discussed was that there may be other clinical scenarios, beyond those included in the specifications, in which you would not want to hold off starting medication while trying to help patients enroll in therapy.

The Committee emphasized the need for interventions that generate improvement on this measure as soon as there has not been enough improvement over time. Some members supported a version of this measure in the future that removes the indications and focuses on ensuring all children receive appropriate care coordination and psychosocial support.

3175 Continuity of Pharmacotherapy for Opioid Use Disorder (University of Southern California)

Measure Steward/Developer Representatives at the Meeting

Soeren Mattke (University of Southern California)

Standing Committee Votes

- Reliability: H-2; M-11; L-3; I-0
- Validity: M-10; L-6; I-0

Ad Hoc Review of Scientific Acceptability at the Clinician Level of Analysis

The Committee voted to pass this measure on reliability and validity at the clinician level based on the new testing provided. The endorsement of this measure is therefore recommended to be expanded to the clinician level of analysis. This measure focuses on the percentage of adults on medications for opioid use disorder who have at least 180 days of continuous treatment. The Committee conducted a targeted review of reliability and validity at the clinician level; criteria beyond reliability and validity were not re-adjudicated during this review. The measure was endorsed in 2017 at the health plan and state levels of analysis. It was presented to the Measure Applications Partnership (MAP) in 2018. The MAP encouraged the developer to test the measure at the clinician level of analysis before it is implemented in MIPS.

The Committee felt data collection was very consistent and the measure has clearly defined exclusions. They agreed that testing results passed the general threshold for reliability. The measure submission noted that two-thirds of an expert panel of nine individuals agreed or strongly agreed that the measure

has face validity. The Committee had some concern about the remaining panelists that dissented or were neutral. The developer shared that face validity results presented during the last review also support the measure's validity. NQF staff reminded the Committee that face validity is acceptable testing for the first evaluation (applicable to the current review) and that empirical testing will be necessary for maintenance review. The Committee generally agreed that the attribution approach presented was thorough and well-developed. The developer added that two-thirds of cases could be attributed to a single provider and that the "plurality rule" using days covered method was also used in attribution.

The Committee discussed the data used for testing and whether the sample was representative of the population in which the measure would be used. The developer clarified that the majority of patients in their testing sample were below 65 years of age and dual eligible, which is representative of the opioid use population. The developer noted they were limited to Medicare Fee-For-Service data for their analysis. It was suggested that using an all-payer database would allow for a larger patient volume per clinician, and more clinicians would have enough patients to calculate a performance score. It was discussed that the vast majority of the pharmacotherapy included in the measure was for buprenorphine rather than methadone. The Committee was very interested in reviewing empirical validity testing during the scheduled maintenance review.

The measure passed the reliability and validity criteria, so it is recommended that endorsement is expanded to the individual clinician and clinician group/practice levels of analysis. The measure will retain its existing maintenance review schedule.

3492 Acute Care Use Due to Opioid Overdose (Yale CORE)

Measure Steward/Developer Representatives at the Meeting

Ilana Richman (Yale CORE)
Lisa Suter (Yale CORE)

Standing Committee Votes

- Evidence: Pass-18; No Pass-1
- Performance Gap: H-7; M-10; L-0; I-0
- Reliability:
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel. SMP Subgroup Votes: H-1; M-2; L-1; I-0
    - Accept Scientific Methods Panel's rating for Reliability: Yes-18; No-0
- Validity:
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel. SMP Subgroup Votes: H-0; M-1; L-2; I-1
    - Accept Scientific Methods Panel's rating for Validity: Yes-8; No-11 (Consensus Not Reached)
    - Committee Vote: H-2; M-9; L-6; I-2 (Consensus Not Reached)
- Feasibility: H-2; M-11; L-2; I-0
- Use: Pass-13; No Pass-3

- Usability: H-2; M-10; L-4; I-0

Standing Committee Recommendation for Endorsement: Yes-X; No-X

The Standing Committee did not vote on the recommendation for endorsement at the meeting because the Committee did not reach consensus on validity—a must-pass criterion.

This new measure, specified at the population level (i.e. county and state), was developed for use in the Maryland Total Cost of Care model. The measure captures the rate of emergency department (ED) visits for opioid overdose among individuals in a specified geography over a one-year period. This measure was rated as low on the validity criterion by the SMP. The SMP concerns stemmed from the narrow scope of data used in testing and the lack of risk adjustment. The SMP also questioned the control of the healthcare sector in influencing the risk of opioid overdose. The measure was pulled by a Committee member for discussion and voting.

Regarding evidence, the Committee generally felt that there are obvious interventions that healthcare providers and systems can perform to reduce the risk of opioid overdose such as using medication-assisted treatment (MAT) and safer prescribing practices, increasing the capacity to provide MAT, and enacting mandatory prescription drug monitoring programs. The discussion included the benefit of this measure over the existing indicators of ED utilization and opioid overdose. The developer responded that the measure precisely addresses performance in the Medicare population, the population in which the measure is intended to be used. The Committee emphasized that overdose deaths varied four times across states and ten times between the lowest- and highest-performing counties.

The Committee agreed that the score-level and data element reliability testing results indicate the measure is reliable. The measure is supported by empirical validity testing that compares performance on the measure to performance on two other related indicators (i.e., opioid-related ED and hospital visits in an all-care population and age-adjusted rate of fatal overdose) in 25 states. Results showed the measure was highly correlated with the other measures. It was discussed that social risk factors are not distributed equally across states or counties. The developer shared that traditional factors like state-level poverty rates and state-level opioid overdose rates showed only a moderate relationship and based on the intended measure use, it is important to be able to calculate rates that consider all factors. Regarding the testing sample, the developer stated that performance in the Medicare Fee-For-Service population is not different from trends seen in the broader population. The Committee suggested that the developer consider expanding testing beyond the Medicare population in the future.

The Committee discussed that this is a public health measure and should only be used at the state or county level as specified. During the post-comment meeting, the Committee wanted to discuss whether overdoses are coded appropriately. The Committee will discuss and re-vote on the measure during the post-comment web meeting on April 22, 2020.

3538 All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries Who May Benefit from Integrated Physical and Behavioral Health Care (The Lewin Group)

Measure Steward/Developer Representatives at the Meeting

Colleen McKiernan, Lisa Alecxih, Meridith Eastman (The Lewin Group)
Roxanne Dupert-Frank (Centers for Medicare and Medicaid Services)

Standing Committee Votes

- Evidence: Pass-10; No Pass-9 (Consensus Not Reached)
- Performance Gap: H-1; M-14; L-3; I-2

- Reliability:
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel. SMP subgroup votes: H-5; M-1; L-0; I-0
    - Accept Scientific Methods Panel's rating for Reliability: Yes-18; No-2
- Validity:
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel. SMP subgroup votes: H-2; M-4; L-0; I-0
    - Accept Scientific Methods Panel's rating for Validity: Yes-16; No-5
- Feasibility: H-6; M-15; L-0; I-0
- Use: Pass-17; No Pass-4
- Usability: H-2; M-12; L-6; I-1

Standing Committee Recommendation for Endorsement: Yes-X; No-X

The Standing Committee did not vote on the recommendation for endorsement at the meeting because the Committee did not reach consensus on Evidence—a must-pass criterion.

The developer shared that the new measure is intended for use in state Medicaid to improve quality of care for beneficiaries with physical and mental health integration needs. The denominator includes four strata: beneficiaries with co-occurring physical health and mental health conditions, beneficiaries with co-occurring physical health conditions and a substance use disorder (SUD), beneficiaries with co-occurring mental health conditions and a SUD, and beneficiaries with serious mental illness (SMI). Evidence indicates state-level integrated care pilot programs have shown promise in reducing ED use for those with the need for integrated care. There was considerable discussion about how the outcome represents true quality of care. The developer reiterated that the denominator is a unique population and care coordination and communication are critical to managing these patients appropriately.

The Committee agreed that performance data can lead to multiple downstream outcomes and drive change, but they had some concern that there are factors other than adequate outpatient care and appropriate care coordination, including, but not limited to social determinants of health that play a significant role in why individuals frequent the ED. It was also noted that adequate use is determined by whether an ED visit results in an observation or inpatient psychiatric stay, but there are shortages of psychiatric inpatient beds available in some places.

The Committee discussed that data from 17 states showed high ED use and opportunity for improvement. In the future, the Committee is interested in reviewing normative rates of ED use or rates by a single diagnosis to better understand the appropriate measure benchmark. The scientific acceptability was reviewed by the SMP; the SMP supported that the measure is reliable and valid. The Committee discussed the impact of social determinants of health on performance. Some members expressed that social factors could unfairly impact measure rates, but others noted that the measure should not be adjusted in order to fully understand the factors driving results at the state level. The developer acknowledged that social risk factors were not included in the risk adjustment model, but noted that the population is relatively homogeneous and focuses on patients enrolled in Medicaid (a proxy for socioeconomic status).

The Committee agreed that the measure is feasible without additional discussion. The Committee accepted the measure's planned use to evaluate states and drive improvement in efficient ED use. Additional consideration shared by the Committee include whether there are better intermediate steps between outpatient care and ED visits that should be measured and the potential for unintended consequences (e.g., reducing access to ED care and inappropriately discouraging ED use).

The measure was consensus not reached on the evidence criterion. The Committee will discuss and re-vote on evidence during the post-comment web meeting on April 22, 2020.

3539e Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting

This is an electronic clinical quality measure (eCQM)

Measure Steward/Developer Representatives at the Meeting

Kirsten Barret, Michelle Dardis, Llew Brown (Mathematica)

Standing Committee Votes

- Evidence: H-1; M-18; L-2; I-0
- Performance Gap: H-3; M-18; L-0; I-0
- Reliability: H-1; M-14; L-5; I-0
- Validity: M-13; L-7; I-0
- Feasibility: H-3; M-11; L-6; I-0
- Use: Pass-18; No Pass-2
- Usability: H-1; M-14; L-5; I-0

Standing Committee Recommendation for Endorsement: Yes-19; No-1

The Standing Committee recommended the measure for NQF endorsement. This new eCQM calculates the proportion of hospitalizations for patients 65 years and older where an antipsychotic medication was prescribed in the absence of the threat of harm to self or others. This measure was originally submitted for NQF endorsement in late 2017. At that time the Committee recommended additional testing to examine the impact of the exclusions "antipsychotics prior to admission" and "antipsychotics for treatment resistant depression."

The evidence for the measure includes the American Geriatrics Society 2019 guidelines and literature that indicates harm from prolonged use of antipsychotics (e.g., higher mortality rate, risk of falls, and cerebral vascular events). Based on testing, the developer decided to exclude patients on antipsychotics prior to admission. The Committee was generally supportive of the added exclusion. One member cautioned that antipsychotics might be warranted in some individuals. The developer reiterated that many patients on antipsychotics for depression would be on these medications before the hospitalization and would be excluded. For performance gap, the Committee agreed data shows too many older patients are receiving these medications.

The reliability discussion focused on whether this measure would capture appropriate on-label prescribing since on-label and off-label indications can vary widely between medications. Overall, the Committee supported the measure's reliability. Regarding validity and feasibility, the developer noted that the "threat of harm" element is not collected in a structured field systematically across all sites, but increased implementation of the measure would drive better data collection. The Committee voiced

that during the maintenance evaluation, they would like to assess whether "threat of harm" is being captured more consistently. The measure was designed for use in the Inpatient Hospital Quality Reporting program. Members commented that a potential unintended consequence of the measure is increased restraint use, but overall the benefits of the measure outweigh the risks.

NQF staff shared that since this measure is applicable to both the patient safety and behavioral health topic areas, a subset of Patient Safety Standing Committee members were given the opportunity to provide comment. One member shared preliminary comments which generally supported the measure.

3541 Annual Monitoring for Persons on Long-Term Opioid Therapy (AMO) (Pharmacy Quality Alliance)

Measure Steward/Developer Representatives at the Meeting

Ben Shirley, Lisa Hines (Pharmacy Quality Alliance)

Standing Committee Votes

- Evidence: H-4; M-14; L-1; I-0
- Performance Gap: H-10; M-10; L-0; I-0
- Reliability: H-1; M-16; L-3; I-0
- Validity: M-15; L-3; I-1
- Feasibility: H-9; M-10; L-0; I-0
- Use: Pass-19; No Pass-1
- Usability: H-2; M-15; L-1; I-0

Standing Committee Recommendation for Endorsement: Yes-18; No-0

The Standing Committee recommended this new process measure for NQF endorsement. The measure is specified at the health-plan level and captures the percentage of individuals on long-term opioid therapy who failed to receive at least one drug test during the measurement year. The Committee questioned why one test per year was selected as the threshold, and the developer responded that guidelines recommend testing either one or one-to-two times per year. Their TEP thought once per year was reasonable and performance using this requirement indicated a substantial performance gap.

Regarding scientific acceptability, the Committee agreed testing results provided evidence of reliability and validity. The Committee discussed the definition of long-term opioid therapy – 90 cumulative days' supply of any combination of opioid medications indicated for pain. The definition is supported by the literature and aligns with the duration used to define chronic pain. It was noted that in the future the developer should explore different methodologies to account for prescriptions filled on the same day, but that it is expected same-day fills minimally impact performance rates.

The measure uses administrative claims data and is intended for use in the Quality Rating System for Qualified Health Plans. Use in other programs in the future, such as Medicare, would require additional testing. At least one member expressed that measurement related to opioid use should move towards more meaningful indicators of safe prescribing, but others emphasized that there is much room for improvement in this area as performance for this measure is very low. One member suggested that emerging drug screening technologies should be included in the future as evidence to support their use grows. The Committee agreed this measure is valuable and that its benefits far outweigh its harms.

Public Comment

Three comments were received in support of measure 3492: Acute Care Use Due to Opioid Overdose before the evaluation meetings. No public or NQF member comments were provided during the measure evaluation meetings.

Next Steps

NQF will post the draft technical report on March 11, 2020 for public comment for 30 calendar days. The continuous public comment period with member support will close on April 9, 2020. NQF will re-convene the Standing Committee for the post-comment web meeting on April 22, 2020.