

Public Comment Summary Report

Project Title: Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) Funding Opportunity: Measure Development for the Quality Payment Program (Mental Health/Substance Use Care)

Dates:

The Call for Public Comment ran from December 9, 2020 to January 18, 2021.

The preliminary Public Comment Summary was prepared in January and February 2021. The Public Comment Summary was posted on 8/10/21.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has entered a cooperative agreement with the American Psychiatric Association (APA) and the National Committee for Quality Assurance (NCQA) to develop provider-level measures for mental health and substance use. The cooperative agreement name is MACRA/Measure Development for the Quality Payment Program. The cooperative agreement number is #1V1CMS331640-02-00. As part of its measure development process, American Psychiatric Association (APA) and the National Committee for Quality Assurance (NCQA) requested interested parties to submit comments on measurement-based care measures.

Project Objectives:

The APA Initiative to Develop Meaningful Behavioral Health Quality Measures aims to form scientifically sound quality measures for use in pay-for-performance programs that will improve processes and outcomes for those treated for mental and substance use disorders. This includes five measure topics: measurement-based care processes, improvement or maintenance of functioning, improvement or maintenance of recovery, initiating and updating a suicide safety plan, and reduction in suicidal ideation or behavior.

Information About the Comments Received:

The measure developer solicited public comments by performing an extensive outreach effort via email notification to various stakeholders, including health plans, physicians, researchers, and clinical experts. Announcements were sent to the project's Technical Expert Panel (TEP), Consumer and Family Panel (CFP), NCQA's Behavioral Health Measurement Advisory Panel as well as the American Psychiatric Association's members.

We received a total of 63 responses on this set of measures, from 10 different entities. A large majority of comments received were contributed by 3 individuals.

Specifically, we received comments from:

- Three Specialty Societies
- Four Providers
- One Managed-Care Organization
- Two Other Individuals

Table of Contents

The team sought general feedback on the overall set of measures, as well as on each of the five measures individually. The seven sections in the report below include general and measure-specific summaries of the Public Comment feedback, the measure team's preliminary recommendations for each measure, and an overall analysis of the comments received. The report addresses these topics in the order listed below, beginning first with general feedback on the full set of measures:

1. General Stakeholder Comments
2. *Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment Measure*
3. *Improvement or Maintenance in Functioning for Individuals with a Mental and/or Substance Use Disorder Measure*
4. *Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder Measure*
5. *Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk Measure*
6. *Reduction in Suicidal Ideation or Behavior Symptoms Measure*
7. Overall Analysis of the Comments and Recommendations

1. General Stakeholder Comments

We received general feedback on the overall set of measures from six respondents during the Public Comment period. The majority of Public Comment respondents were in support of the measures (support or support with modifications). Overall, respondents indicated that the measure concepts have value and importance, but noted concerns regarding burden, workflow barriers, and limitations in evidence. Below are the key themes that emerged following a review of the general comments received.

- **Theme 1.1: Burden associated with data collection and workflow modifications required to report measures**
 - **Feedback:** Several respondents expressed concerns with the burden of administering the assessment tools specified in the measures, including the associated data entry and data collection requirements. Of note, one respondent representing a managed-care organization noted the need for additional administrative staff to successfully implement the measures. Another respondent, identifying as a provider, echoed concerns about how measures may strain current clinical workflow and indicated the measures would be very challenging to use and report in practices or clinical settings with little or no administrative support. Finally, one respondent provided feedback that the PsychPRO Registry currently has technical glitches and known challenges with the accuracy of the recorded data and overall user experience.
 - **Response:** The implementation of measurement-based care (MBC) can require significant changes in practice for clinicians. MBC entails routine use of assessment instruments, which may not be part of providers' usual workflow, and can require a different mode of interaction with patients. Patients also need to adjust to the need for

timely completion of patient-reported outcome measures (PROMs), and clinicians need to work closely with their patients to explain the purpose of assessment tools and how they will be used to inform and adjust treatment approaches. For these reasons, adoption of MBC in a practice may take several months and require multiple QI initiatives (e.g., PDSA cycles). APA and NCQA have conducted regular learning collaborative sessions during the development and testing of this measure set, providing technical assistance, answering questions, and working through challenges faced by participants. As MBC is more widely adopted as part of routine clinical practice, data collection difficulties are expected to become less of a barrier to implementation.

- **Theme 1.2: Validity and availability of selected assessment tools and instruments**

- **Feedback:** A few respondents emphasized the importance of constructing measures around well-studied and validated assessment tools. One respondent highlighted that assessment tools in quality measurement should be validated for use among different groups of individuals within the general population, including individuals with different psychiatric diagnoses and levels of condition severity. They expressed that, regardless of conceptual importance, quality measures should not utilize assessment tools if they are not widely used in the field or psychometrically tested and validated for specific patient populations. Respondents had concerns with the specific assessment tools included in the measures, and noted the importance of ensuring that the specified tools are freely available to encourage uptake and use of the measures.
- **Response:** The measure team is committed to using the best publicly-available tools for use in these measures. We understand the concerns about using assessment instruments that are not extensively studied. However, we believe these are critical domains for quality measurement, and these are the tools for which there is currently the best available evidence; broader implementation of MBC will provide opportunities for further improvement and refinement of tools used to assess mental health outcomes. If we are able to acquire sufficient data during the testing process, we will also be looking to crosswalk various tools to each other to identify where there may be multiple options for reporting the measure.

- **Theme 1.3: Recommendation to expand eligible population**

- **Feedback:** One respondent recommended lowering the age threshold for all measures to individuals age 12 and older due to the high prevalence of mental and substance use disorders among adolescents.
- **Response:** These measures were designed to focus on the adult population, but the measure team will explore incorporation of individuals age 12 and under for future versions of the measures.

- **Theme 1.4: Additional review of measure exceptions**

- **Feedback:** Multiple respondents provided feedback on draft measure exceptions. One respondent noted that the exceptions outlined in the specifications were sufficient, while another encouraged the team to consider additional exceptions such as language barriers, intellectual impairment, and invalid responses due to states of distress. Another respondent expressed concern with the exception due to “lack of motivation,” noting that clinicians could potentially game the measure by attributing scores from a

patient who does not demonstrate improvement in the assessment tools to lack of motivation, rather than ineffective treatment. Respondents also urged the measure developer to consider whether all the proposed exceptions are appropriate for all proposed measures, particularly the suicide process measure.

- **Response:** APA and NCQA will be assessing exceptions and potential exclusions as part of our testing plan, and will continue to review these issues with our Technical Expert and Consumer & Family Panels in light of public comments received as analysis is completed.

2. Measurement-Based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment Measure

Public Comment Response Themes

We received responses from four respondents on this measure. Respondents represented individual providers, a managed-care organization (MCO), and a specialty society. All respondents indicated support for the measure or supported the measure with modifications. The specific themes raised in public comment, along with the measure developer's responses, are summarized below.

- **Theme 2.1: Burden associated with measurement-based care**

- **Feedback:** In an effort to reduce added burden on clinicians and patients associated with this measure, one respondent recommended narrowing the scope of the measure to include three, rather than five, domains for assessment. Another respondent, representing an MCO, shared similar reflections, noting that covering multiple domains across multiple screening assessments would be difficult to complete within a single visit. The respondent indicated that additional personnel, including both administrative and clinical staff, would be needed to contact patients, and conduct the screening and follow-up assessments.
- **Response:** As noted in our supporting materials, the implementation of measurement-based care (MBC) can require significant changes in practice for clinicians. MBC entails routine use of assessment instruments, which may not be part of providers' usual workflow, and can require a different mode of interaction with patients. Patients also need to adjust to the need for timely completion of patient-reported outcome measures (PROMs), and clinicians need to work closely with their patients to explain the purpose of assessment tools and how they will be used to inform and adjust treatment approaches. For these reasons, adoption of MBC in a practice may take several months and require multiple QI initiatives (e.g., PDSA cycles). APA and NCQA have conducted regular learning collaborative sessions during the development and testing of this measure set, providing technical assistance, answering questions, and working through challenges faced by participants. As MBC is more widely adopted as part of routine clinical practice, data collection difficulties are expected to become less of a barrier to implementation. APA and NCQA discussed with our TEP and CFP members the potential to make psychosis and recovery assessments optional, keeping depression, anxiety, substance use, suicide ideation and behavior, and functioning as mandatory assessment domains to help reduce implementation burden. Both the TEP and CFP members agreed with this change.

- **Theme 2.2: Measuring treatment adjustment**

- **Feedback:** Several respondents pointed out challenges related to measuring treatment adjustment, both conceptually and logistically, and the potential unintended consequences that could result from a measure of treatment adjustment. While respondents understood the importance of treatment adjustment in high quality care, they expressed that it is not possible to know simply from a change in medication, therapy, or a referral, that adjustments or changes were driven by ratings from an assessment tool. Multiple respondents suggested that the data needed to determine the occurrence of treatment adjustment (as specified in the measure) may be available in patient records; however, significant administrative and clinical resources would be required to ensure that the data is comprehensive. Additionally, one respondent felt the criteria outlined for care plan adjustment in the current specifications were unclear, and that the language regarding care plan adjustment should be more explicit.

Finally, one respondent urged the measure development team to consider the unintended consequences of this measure, including premature treatment adjustment of certain medications (e.g., antidepressants) or psychotherapy that may require a longer period of time before demonstrating effectiveness. One respondent suggested the team consider implementing the first two indicators of the measurement-based care measure first (i.e., baseline assessment and monitoring), and consider adding the treatment adjustment indicator if the measure shows value and usability in clinical practice.

- **Response:** The measure team appreciates the feedback received on the measure, and will critically think through the importance, feasibility, and utility of all measure components. However, as emphasized by our Technical Expert Panel, we strongly believe in the importance of the treatment adjustment element of this measure. Patient assessment and monitoring are critical components of quality care, but do not in themselves have a significant impact on patient outcomes. As CMS and the National Quality Forum have repeatedly underscored, process measures should have a strong link to outcomes, to ensure that measures are not merely functioning as “checkbox” requirements that add to documentation burden without contributing to improvements in the quality of care. The evidence suggests that adjusting treatment as needed is the element of this measure that is most closely linked to improved patient outcomes. We would also note that the measure does not require a treatment adjustment if such an adjustment is unnecessary, only that there is documentation of a clinical decision to adjust or not adjust the care plan following monitoring. We will continue to explore the feasibility and appropriateness of this element of the measure as part of our testing and analysis.

3. *Improvement or Maintenance in Functioning for Individuals with a Mental and/or Substance Use Disorder Measure*

Public Comment Response Themes

We received comments from ten respondents on this measure. Respondents represented individual providers, specialty societies, a managed-care organization (MCO), and other individuals. All respondents indicated support for the measure (i.e., support or support with modifications). The specific themes raised in public comment, along with the measure developer’s responses, are summarized below.

- **Theme 3.1: Appropriateness of assessment tool**

- **Feedback:** Two respondents questioned the singular use of the 12-item WHODAS 2.0 as the tool for assessing outcomes in this measure. Several respondents suggested that the measure validity and usability could be improved by allowing for multiple assessment instruments (e.g., 36-item WHODAS and Sheehan Disability Index). With regard to the 12-item WHODAS 2.0, one respondent felt it lacked valuable information included in the 36-item WHODAS and that it was less appropriate for assessment of higher functioning groups, due to its strong focus on disability. With regard to interpreting the assessment scores, a respondent noted that the WHODAS 2.0 is subject to inconsistent and invalid score interpretations, due to the variability in elements being assessed and lack of scoring adjustments to account for these differences when measuring change.
- **Response:** The Sheehan Disability Scale was allowed to be substituted for the WHODAS 2.0 during beta testing of the measure. The frequency of tool use will be reviewed during analysis. The 36-item WHODAS version was not chosen due to the length of the tool which would increase completion time.

- **Theme 3.2: Assessment timeframe**

- **Feedback:** Public Comment participants were asked to provide feedback on whether the assessment of improvement or maintenance in functioning should be measured at 60 days (+/- 30 days), 90 days (+/- 30 days), or 180 days (+/- 30 days). Overall, feedback on this question was split, with 4 respondents supporting the use of a 90-day (+/- 30 days) period, 3 supporting a 180-day (+/- 30 days) period, 1 supporting a 60-day (+/- 30 day) period, and 1 recommending an additional option of 120 days (+/- 30 days). Although the measure team did not receive substantive feedback with rationales for the different threshold recommendations, the respondent indicating support for the 60-day threshold indicated that a shorter timeframe increases the actionability of the measure, as this follow-up timeframe allows earlier opportunity for modifying treatment. However, another respondent noted that the 60-day timeframe is too short to assess meaningful change for these outcomes. Another respondent who recommended 90 days indicated that 90 days is aligned with the treatment planning interval at many clinics, and that specifying the measure with this timeframe would align the follow-up assessment with the workflow and treatment adjustment. Overall, commenters agreed that, regardless of the assessment period, there should be consistency across the entire set of measures.

Response: The measure team will continue to explore the appropriate assessment timeframe as part of the testing and analysis of these measures. The measure team reviewed the public comments with the TEP and CFP. Panel members agreed with the recommendation to continue with the originally proposed 180-day (+/- 30 days) assessment timeframe. This aligns with the recovery measure timeframe.

- **Theme 3.3: Methodology for calculating change in assessment score**

- **Feedback:** Public Comment participants were asked to provide input on a recommended methodology for calculating change between the index and follow-up assessment. Feedback on the approach was evenly split between respondents who preferred calculating a score change using the index assessment and the last assessment a patient had during the 180-day (+/- 30 days) period, and those who preferred calculating a

score change between each index assessment and the subsequent assessment following the index (i.e., multiple index events possible per patient). One respondent who preferred the option of calculating change between the index and last assessment during the 180-day (+/- 30 days) period noted that calculating change scores between each index and subsequent assessment would create challenges in the context of reporting the measure.

- **Response:** The measure team will continue to explore the appropriate methodology and approach to calculating this measure. The measure team reviewed the public comments with the TEP and CFP. The panels agreed with the recommendation to calculate score change between index and the last assessment within the follow-up timeframe (i.e., at 180-days (+/- 30 days) after index assessment). This aligns with the recovery measure.
- **Theme 3.4: Threshold for assessment score reduction**
 - **Feedback:** Public Comment participants were asked to provide feedback on a potential alternative numerator which assesses if a WHODAS 2.0 score reduction threshold was met or surpassed at follow-up, indicating treatment improvement. The measure team posed several threshold recommendations based on findings from the literature, including 19% and 35%. Five out of the 6 respondents commenting on this topic recommended the 35% reduction threshold in the WHODAS 2.0 follow-up score and 1 respondent recommended the use of a 19% score reduction. One respondent did not recommend either proposed option, noting challenges with the interpretation of the WHODAS 2.0. The respondent acknowledged that each percentage decrease in WHODAS 2.0 does not represent comparable change in “function” without factoring in the actual questions that the changes were attributed to.
 - **Response:** The measure team will continue to evaluate the appropriate cutoff points for score changes as part of the testing and analysis of these measures. The measure team reviewed the public comments with the TEP and CFP. The panels agreed with the recommendation to use any positive change as the threshold for meeting measure requirements. As the measure is implemented and performance data becomes available, we will gain a better understanding of the appropriate threshold cutoffs.

Preliminary Recommendations

The measure team appreciates the constructive feedback received on this measure concept and will continue reviewing the individual elements of the specification (i.e., assessment tools and timeframes, and calculation methodology) to strengthen the validity and usability of the measure. To improve alignment across existing follow-up assessments, the measure team will explore an assessment timeframe of 90 days in testing.

4. *Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder Measure*

Public Comment Response Themes

We received responses from nine respondents on this measure. Respondents represented individual providers, specialty societies, an MCO, and an ‘other’ health care organization. Eight respondents indicated support for the measure (with or without modifications) and one respondent did not

support the measure. The specific themes raised in public comment, along with the measure developer's responses, are summarized below.

- **Theme 4.1: Appropriateness of assessment tool**

- **Feedback:** Three respondents questioned the singular use of the 24-item Recovery Assessment Scale (RAS) as the assessment tool for this measure. Comments focused on the lack of widespread adoption of this measure by providers and the unknown validity of the instrument among particular populations (e.g., individuals experiencing housing insecurity or homelessness, individuals with particular comorbidities). One respondent suggested that other tools, such as the PHQ-9 and GAD-7, should be considered for inclusion in the measure in addition to the RAS.
- **Response:** The measure team appreciates the desire to use assessment instruments that are in widespread use. In our assessment of the literature and use of recovery assessment tools in practice, the RAS was the most cited and used recovery tool. Among the studies examined was a systematic review of 12 longitudinal studies in which RAS scores were used as outcome variables, and which included marginalized populations as part of their samples. In these studies, the scores increased over time, which was associated with maintaining or showing improvement towards the goal of continued recovery (Salzer & Brusilovskiy, 2014). The RAS-24 instrument has also been evaluated for construct and content validity.

As feasible, the team intends to continue to compare and evaluate different assessment tools and scales as part of the testing process, conducting crosswalks across assessments with an eye toward identifying opportunities for use of multiple tools. The PHQ-9 and the GAD-7 do include components that assess reduction or remission of symptoms; however, we would note that symptom reduction is not equivalent to recovery, which includes broader concepts of wellness, health, and quality of life.

References:

Salzer, M. S., & Brusilovskiy, E. (2014). Advancing recovery science: reliability and validity properties of the Recovery Assessment Scale. *Psychiatric Services* (Washington, D.C.), 65(4), 442–453. <https://doi.org/10.1176/appi.ps.201300089>

- **Theme 4.2: Assessment timeframe**

- **Feedback:** Public Comment participants were asked to provide input on whether the assessment of improvement or maintenance in recovery should be measured at 60 days (+/- 30 days), 90 days (+/- 30 days), or 180 days (+/- 30 days). The recommendations were evenly split between those that suggested a 180-day period and those that suggested a shorter period (2 of 3 respondents recommended 90 days). Respondents who suggested the use of a 180-day (+/- 30 days) period felt that meaningful change in recovery was unlikely within a shorter timeframe and urged the measure development team to keep the assessment periods consistent across the proposed measures posted to Public Comment. One respondent, who identified as a provider, suggested the use of a 90-day period as the timeframe coincides with current treatment plan intervals and would improve the utility of the measure.
- **Response:** The measure team will continue to explore the appropriate assessment timeframe as part of the testing and analysis of these measures. The measure team

reviewed the public comments with the TEP and CFP. The panels agreed with the recommendation to continue with the originally proposed 180-day (+/- 30 days) assessment timeframe. This aligns with the functioning measure timeframe.

- **Theme 4.3: Methodology for calculating change in assessment score**

- **Feedback:** Public Comment participants were asked to provide input on the methodology used to calculate change if more than 1 follow-up occurred within 180 days (+/- 30 days) of the index assessment. The majority of respondents (4 out of 6) suggested calculating a change score using the index assessment and the last assessment a patient had during the 180-day (+/- 30 days) period. Only 2 respondents preferred calculating a change score between an index assessment and the subsequent assessment within the measurement period (i.e., multiple index events possible per patient). Respondents did not provide additional information to substantiate their recommendations. A respondent did note that making an informed recommendation is difficult without measure testing results.
- **Response:** The measure team will continue to explore the appropriate methodology and approach to calculating this measure as part of our testing and analysis. The measure team reviewed the public comments with the TEP and CFP. The panels agreed with the recommendation to calculate score change between index and the last assessment within the follow-up timeframe (i.e., at 180-days (+/- 30 days) after index assessment). This aligns with the functioning measure.

- **Theme 4.4: Definition of recovery**

- **Feedback:** One respondent noted that, although standardized measurement of recovery is important, this measure should not move forward in development until the definition of recovery is more concrete and an assessment tool is widely available and validated for use in various populations.
- **Response:** The Substance Abuse and Mental Health Services Administration (SAMHSA) defines recovery as “a process of change through which individuals improve their health and wellness, live a self-directed life, and strive to reach their full potential” (SAMHSA, 2019). We understand this definition to be widely used and it is the definition we are following for purposes of this measure.

Preliminary Recommendations

Public comment feedback indicated overall support for this measure concept, and the team proposes reviewing the recommendations received to strengthen the specification and further explore its development.

5. *Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk Measure*

Public Comment Response Themes

We received responses from nine respondents on this measure. Respondents represented individual providers, specialty societies, a health plan, and other health care organizations. The majority of respondents (eight out of nine) indicated support for the measure or supported the measure with

modifications. Public Comment respondents acknowledged that suicide safety planning (SSP) is important among patients at risk for suicide and is consistent with quality patient care, but noted the limitations of current evidence to support the effectiveness of SSP within outpatient and ambulatory care settings (i.e., the proposed measure care settings). The specific themes raised in public comment, along with the measure developer's responses, are summarized below.

- **Theme 5.1: Concerns with selection of assessment tool**

- **Feedback:** One respondent expressed hesitation with the measure's reference to the use of the C-SSRS tool for suicide risk assessment, noting that the incorporation of this tool in a measure may hinder progress on the tool's needed improvements.
- **Response:** The measure specifications do not limit the clinician to using the C-SSRS tool for this measure and allow for assessment of suicidal ideation or behavior symptoms based on results of any standardized assessment tool, or suicide risk based on the clinician's evaluation. As uptake of the measure increases, there will be potential for review and expansion of qualified assessment tools.

- **Theme 5.2: Suicide risk threshold for inclusion in measure denominator**

- **Feedback:** Public Comment participants were asked to provide input on the denominator population for this measure, which currently includes individuals with non-zero scores on any standardized assessment of suicidal ideation or behavior, or suicide risk based on the clinician's evaluation. Of the three respondents who provided feedback on the suicide risk score threshold for inclusion in the measure denominator, all three supported the use of a higher threshold that indicates "moderate" or "high" suicide risk. Respondents explained that a higher risk threshold could reduce measure-associated burden and improve the relevance of the measure by targeting a more critical population for intervention.
- **Response:** The measure team has received consistent feedback calling for a higher threshold for inclusion in the denominator for this measure; we will look to testing results to try to identify that threshold, and will continue to review these issues with our Technical Expert and Consumer & Family Panels in light of public comments received as analysis is completed.

- **Theme: Feasibility concerns**

- **Feedback:** Multiple respondents provided input on the feasibility of the measure. One respondent, representing a health plan, did not support the measure due to cited barriers related to the clinical and administrative resources and changes to workflow that would be necessary to report the measure. The measure team also received feedback on the timeframe for the SSP review period, where 1 respondent recommended allowing greater flexibility with the 120-day target to increase the usability of the measure. The respondent noted that users of the registry may not have active treatment-receiving patients in the same setting for up to 120 days, and requiring such a duration of time before the SSP review is conducted would limit the measure's usability.
- **Response:** This measure is proposed for use at the clinician level. Clinicians who do not have the training, resources, and ancillary staff needed to implement and conduct the processes assessed by the measure would not be included for assessment. With regard to the measure's timeframe for SSP review, the measure allows for the review to occur

at any point within 120 days after initiation of the SSP, rather than specifically at 120 days after initiation.

Preliminary Recommendations

The measure team is encouraged by the overall support for this measure concept. The team will review the feedback regarding score thresholds and timeframes to ensure the measure is feasible and useful to providers who elect to report this measure. We agree with increasing the suicide risk threshold in the denominator to “moderate or high” risk to focus and improve the measure’s utility, and will solicit additional expert feedback on the revised criteria before finalizing the specification.

6. Reduction in Suicidal Ideation or Behavior Symptoms Measure

Public Comment Response Themes

The measure team received responses from nine respondents on this measure. Respondents represented individual providers, specialty societies, a health plan, and other health care organizations. The majority of respondents (seven out of nine) supported the measure or supported it with modifications. The specific themes raised in public comment, along with the measure developer’s responses, are summarized below.

- **Theme 6.1: Concerns with appropriateness of assessment tool**
 - **Feedback:** Several respondents raised concerns about the validity of assessment instruments in measuring outcomes for the population defined in this measure. A few respondents suggested that the measure uses the C-SSRS+ tool in an unintended way, as suicidal ideation prompts within the tool are multi-part questions assessing different dimensions of risk. Respondents indicated that the interpretation of risk is limited when restricted to binary or numerical data (i.e. without understanding other documented risk factors). Another respondent raised additional concerns about the appropriateness of the tool for this measure, as the C-SSRS+ does not inquire about current suicidal ideation, but rather, suicidal ideation within the period of time since the patient’s last visit. The respondent suggested that this lack of specificity in risk period may introduce validity challenges into the measure.
 - **Response:** The measure team notes that, as advised by our Technical Expert Panel and others, use of the C-SSRS is not intended to be a de facto assessment of suicide risk, but rather an assessment of suicidal behavior and ideation. As with other assessment tools used in this set of measures, we will continue to evaluate additional instruments that assess suicidal thoughts and behavior (e.g., item 9 of the PHQ-9), and will conduct crosswalks with the C-SSRS to explore the possibility of using multiple tools as appropriate.
- **Theme 6.2: Calculating meaningful changes in score**
 - **Feedback:** Public Comment participants were asked to provide input on how to define a “meaningful change” in C-SSRS+ score in the measure numerator. The measure specification, as posted to Public Comment, indicates that the definition of a “meaningful change” in score for the C-SSRS+ assessment would be identified after measure testing. Respondents shared anticipated challenges with defining a meaningful score reduction, as “meaningful change” thresholds are likely to be dependent on the

severity of risk at the outset of treatment. Notably, respondents also reflected on the potential unintended consequences of assessing for “meaningful change” at follow-up, including potential patient abandonment, or avoidance of assessing suicidal ideation among those with fluctuating levels of suicidality.

- **Response:** The measure team will continue to explore the appropriate methodology and approach to calculating this measure as part of our testing and analysis, and will continue to review these issues with our Technical Expert and Consumer & Family Panels in light of public comments received as analysis is completed.
- **Theme 6.3: Methodology for calculating change in assessment score**
 - **Feedback:** Public Comment participants were asked to provide input on the methodology used to calculate change if more than 1 follow-up assessment with the C-SSRS+ was completed within 90 days (+30 days) of the index assessment. Feedback on this question was divided. One respondent recommended calculating change between index and subsequent assessments, citing that doing so would minimize potential unintended consequences of avoiding assessment among individuals with fluctuation in levels of suicidality. Another respondent recommended calculating change between the index assessment and lowest score on a repeat assessment during the measurement period, as this would ensure the maximum response to treatment achieved during the follow-up period was recognized.
 - **Response:** The measure team will continue to explore the appropriate methodology and approach to calculating this measure as part of our testing and analysis. The measure team reviewed the public comments with the TEP and CFP. The panels agreed with the recommendation to calculate score change between index and the last assessment within 90-days (+30 days).
- **Theme 6.4: Assessment timeframe**
 - **Feedback:** One respondent provided feedback on the timeframe for follow-up assessment and recommended shortening the timeframe to allow more individuals to meet eligibility requirements of this measure (i.e., continuous enrollment requirements).
 - **Response:** The measure currently allows for the follow-up assessment to occur at any time within the 90-day period, rather than at the 90-day interval.

7. Overall Analysis of the Comments and Recommendations

The Project Team appreciates the respondents’ thoughtful input, recommendations, and requests for clarification about the measures. All comments and responses were reviewed with the Technical Expert Panel (TEP) and Consumer and Family Panel (CFP) on March 1, 2021. The panel feedback was included in the above summary and recommendations. Overall, comments from the public were consistent with concerns and issues that have been raised in prior discussions with the TEP and the CFP. General takeaways from the Public Comment feedback included overall support for the measure concepts, but notable concerns related to increased burden, feasibility challenges, and limitations in the supporting evidence.

Public Comment Verbatim Report

Comment Number*	Date Posted/Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
31	12/29/2020	Mirean Coleman, LICSW, CT; National Association of Social Workers	Specialty Society	Cross-cutting Questions	N/A	Exception	The list is inclusive of exceptions.
32	12/29/2020	Mirean Coleman, LICSW, CT; National Association of Social Workers	Specialty Society	Cross-cutting Questions	N/A	Exclusion	No, not in this circumstance.
43	1/15/2021	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Cross-cutting Questions	N/A	Exception	Agree and suggest expanding list to include language barriers, lack of understanding of the questions through intellectual impairment, or invalid responses from over-rating or under-rating their distress
44	1/15/2021	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Cross-cutting Questions	N/A	Exclusion	Just as above [Agree and suggest expanding list to include language barriers, lack of understanding of the questions through intellectual impairment, or invalid responses from over-rating or under-rating their distress]
No62	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Cross-cutting Questions	N/A	Exception	How is the exception related to lack of motivation to improve being determined? What's to stop a provider from attributing the score from a patient who doesn't improve as due to their "lack of motivation" and not a lack of effective treatment? Also do these four exceptions make clinical sense for the suicide process measure? Or were they just being repeated across all measures without that particular consideration?
2	12/14/2020	Jessica Holton, MSW, LCSW, LCAS	Provider	General – Project or Measure Suite	N/A	Project	The current platform is not intuitive and has several glitches that could interfere with consistent follow-through from providers and clients.

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
7	12/18/2020	Julie Ohnemus, MD; Open Door Community Health Centers	MCO - Managed Care Organization	General – Project or Measure Suite	N/A	Project	I certainly would be interested in tracking what these initiatives have to offer but funding would need to accompany to administrate and have staff hired to do such work
28	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	General – Project or Measure Suite	N/A	Project	The overall intent of this set of measures is good. Although evidence is still evolving, some research suggests that use of standardized rating scales is likely to improve the assessment, monitoring, and treatment of individuals with psychiatric illness. Recent clinical practice guidelines are also in support of such an approach. However, as discussed in greater detail, some of the specific approaches that are being taken to measurement are likely to be problematic. Revisions to the measures and/or elimination of some of the measures would be important at this juncture until more robust information is available on the utility and benefits of some of these rating scales.
33	12/29/2020	Mirean Coleman, LICSW, CT; National Association of Social Workers	Specialty Society	General – Project or Measure Suite	N/A	Project	There are many children suffering from a mental illness or substance use disorder. I recommend lowering the age qualification from 18 to 12.
46	1/15/2021	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	General – Project or Measure Suite	N/A	Project	A useful effort
63	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	General – Project or Measure Suite	N/A	Project	Thank you for the opportunity to provide feedback on these measures. Given that the measures in their current format have several undefined fields that are dependent on testing, which will likely influence the final format of the measures, comprehensive commenting is difficult at this time. In addition to the comments made on specific measures, we wonder if there was a rationale to exclude the Health Behavior Assessment and Intervention Codes from the code set. These would seem to be appropriate to include for various integrated care settings where psychologists are located.

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
8	12/18/2020	Julie Ohnemus, MD; Open Door Community Health Centers	MCO - Managed Care Organization	General - Unintended Consequences following implementation of any measures in this suite	N/A	Unintended Consequences	Costs & staffing; accuracy of tools in our patient population
29	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	General - Unintended Consequences following implementation of any measures in this suite	N/A	Unintended Consequences	<p>Most of the potential unintended consequences have already been discussed in the context of individual measures, but significant unintended consequences include:</p> <ul style="list-style-type: none"> • formalizing use of scales such as the RAS-24 or the C-SSRS as indicators of "quality care" when they still have substantial weaknesses. • spending substantial proportions of visit time on safety planning or C-SSRS administration when other aspects of care such as assessment, psychoeducation, psychotherapy, or treatment planning may be more valuable for the individual patient creating administrative burdens and burdens on clinicians that seem to outweigh any clinical benefit to patients or financial benefit to clinicians (e.g., via MIPS or value-based care). <p>If measures are designed in a way that is cumbersome and logically complex, clinicians may actually become less inclined to use measurement-based care than they are at present.</p>
34	12/29/2020	Mirean Coleman, LICSW, CT; National Association of Social Workers	Specialty Society	General - Unintended Consequences following implementation of any measures in this suite	N/A	Unintended Consequences	It would be important to re- assess for suicide as frequently as possible because this is a high risk population.

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
10	12/22/2020	Laura J. Fochtmann, MD; Stony Brook University	Provider	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support w/ Mod	Overall Support w/ Mod	<p>1. Aspects of rating scale selection for inclusion in the measure:</p> <p><i>a. Appropriateness for outpatient use</i> Any rating scales (whether patient reported or clinician reported) that are suggested for use as part of this measure should have clear data supporting utility and usability in outpatient populations of patients, including individuals with a range of psychiatric diagnoses and a range of illness severities. Data should also be available that assesses its appropriateness across demographic groups and in individuals with varying levels of health literacy. If such data is not available for a given scale, it should not be included until additional information can be gathered on the scale's properties.</p> <p><i>b. Sensitivity to clinically relevant changes</i> Since the measure is examining changes in scale scores as well as baseline levels of symptoms, there should also be data available on what constitutes a clinically significant change in the rating scale score and whether scale increments are consistent in their meaning. (For example, is going from 30 to 25 on a scale about the same as going from 15 to 10 on the scale? Does it depend on which items are rated most highly?)</p> <p><i>c. Clinically meaningful scoring thresholds</i> If thresholds are used to determine whether specific actions need to be taken, those thresholds should have been assessed for their discriminant ability (e.g., with ROC curves).</p> <p><i>d. Scale availability</i> Any rating scales that are suggested for use as part of this measure should be freely available for clinical use and incorporation into electronic record systems. For scales that are freely available with permission of the scale copyright holder, administrative processes for requesting use should be straightforward to complete. Mandatory paid training as a requirement for using a "freely available" scale should be viewed as akin to paying for scale use. Rating scales that</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							do not meet these requirements should not be included as options in the measure.
11	12/22/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support w/ Mod	Overall Support w/ Mod	<p>2. Bundling of multiple clinical outcomes into a single measure</p> <p><i>a. Inclusion of symptoms that are part of the DSM Cross-Cutting Measure seems appropriate to this measure</i></p> <p><i>b. Inclusion of measures that also exist as freestanding measures</i></p> <p>It is not clear why the functioning and recovery measures are incorporated here but also exist as freestanding measures. These should be removed from the cross-cutting measure. When measures have multiple bundled requirements either as part of the measure itself or as part of multiple different numerators and denominators, it becomes much less likely that the measure will be feasible to use in clinical practice. Clinicians could still choose to do multiple measures and, if so, they should receive credit for doing so.</p>
12	12/22/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support w/ Mod	Overall Support w/ Mod	<p>3. Removal of recovery scale as part of this measure</p> <p>If a decision is made to continue to have multiple bundled measures as part of this initial cross-cutting measure, the recovery element should still be eliminated at the present time. Recovery is obviously an important issue; however, it has been conceptualized in multiple ways and its measurement accuracy using the recommended scale is unclear. Until further consensus is available on definitions of recovery and until robust measures are available, it should not be incorporated into this overarching measure.</p>
13	12/22/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support w/ Mod	Overall Support w/ Mod	<p>4. Removal of the third numerator/denominator for this measure</p> <p>The third part of the measure is defined in a way that would be difficult to implement in a meaningful and reliable fashion. The definition of a "care plan" is unclear but seems broad and non-specific. Ostensibly, the goal of measurement-based care is for information about patient symptoms, functioning, and other variables to drive clinically indicated changes in the treatment</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>plan. However, simply noting that a change in medications, therapy, referral or consultation occurred does not mean that it was driven by the rating scale findings. In addition, a change in treatment could have been made just before the ratings were completed and an additional change could be inappropriate at that juncture. For example, some medications (e.g., lamotrigine, clozapine) requires slow titration and another dose increase may not be appropriate. Other medications (e.g., lithium) may require serum monitoring before making further adjustments in dose. If treating with psychotherapy, antidepressants, or antipsychotics, a stable approach to treatment may be needed for several weeks before an effect would be expected; premature adjustments to treatment may be counterproductive. Thus, it would not be sufficient to look for orders (e.g., referrals, consultations) or changes in medication to fulfill this measure. Instead, meaningful information about a change in the care plan (or a rationale for no change) would likely require manual chart review or additional entry of structured information by the clinician. Each of these strategies is associated with substantial administrative and/or clinician burden. Given the complexities of this third portion of the measure, it would be preferable to adopt the initial 2 portions of the measure first and consider adding the third part of the measure at a later time, depending on the utility and usability of the first two parts of the measure in actual practice.</p>
14	12/22/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support w/ Mod	General	<p>The overall principle of using standardized self-report ratings to inform assessment and treatment is a good one. However, there are a number of modifications that would be important to make to this measure and supporting materials</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
35	12/29/2020	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support w/ Mod	Overall Support w/ Mod	Suggest 3 domains instead of 5. Regular assessment of 5 domains presents a burden on both the patient and the health system in busy clinical practices
47	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Support	General	The adjustment in care plan criterion is unclear. It also possible that responsiveness to measure information involves staying the course (e.g., the data may not indicate that an adjustment is needed). The language regarding adjustment and no adjustment is inconsistent and may be confusing to the user.
3	12/18/2020	Julie Ohnemus, MD; Open Door Community Health Centers	MCO - Managed Care Organization	Measurement-based Care Processes: Baseline Assessment, Monitoring and Treatment Adjustment (Process Measure)	Neither support or do not support	General	Would have challenges in tracking. Doing some of the screening which already is a challenge to obtain within a visit. Would likely need to hire someone to contact patients to perform these screens and do follow-ups.
9	12/22/2020	Katherine Ast, MSW, LCSW; American Academy of Hospice and Palliative Medicine (AAHPM)	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support	Timeframe	Most recent assessment

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
15	12/22/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Overall Support w/ Mod	<p>1. Rationale for support Assessing and monitoring the extent of an individual's functioning is an important aspect of care. The WHODAS 2.0 12 item measure seems appropriate for use in this regard since it is well-validated. It does include a mix of physical and mental health related items, but this also seems appropriate given the high degree of physical health related comorbidities in individuals with mental health conditions.</p> <p>2. Support for freestanding measure rather than bundled measure As noted above, it seems better to have this as a freestanding measure rather than bundled with the measurement care processes measure so that clinicians could receive appropriate credit for using both.</p> <p>3. Scale version used Since the 36 item version of the WHODAS provides more detailed information, particularly for a baseline assessment, the measure should permit initial use of the WHODAS 36 with followup assessments using the shorter scale if desired.</p> <p>4. Denominator identification period The fact that the denominator identification period starts before the measurement period seems potentially problematic. Software vendors typically push to have EMR software upgrades in place before the start of a measurement year but not sufficiently beforehand to support such measures. User education programs similarly are planned to coincide with the start of the measurement period. The denominator identification periods don't seem consistent for all of the measures of this measure set, which is likely to add to the confusion. Each of these factors would tend to reduce adoption of the measure by clinicians.</p> <p>5. Scale availability The WHODAS information associated with the DSM5 measures definitions notes that, for the 36 item version, "This material can be reproduced</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							without permission by clinicians for use with their own patients. Any other use, including electronic use, requires written permission from WHO." As discussed above in the section on measurement-based care processes, it would be important for users to have a straightforward mechanism for requesting permission for incorporation into electronic systems. Otherwise, clinicians or health organizations will not be inclined to incorporate these measures.
16	12/22/2020	Laura J. Fochtmann, MD; Stony Brook University	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	General	Using a specific % of change in WHODAS scores seems fraught with difficulties for the reasons discussed above. In addition, it is not clear what data supports these specific cut-points.
17	12/22/2020	Laura J. Fochtmann, MD; Stony Brook University	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Timeframe	Since the measures are constructed as a set, it is important that all of the time frames be consistent for all of the measures. For the functioning measure, 60 days seems too short to note a meaningful change in function for many patients, since significant functioning changes would typically lag behind symptom changes. 90, 120, or 180 days seem acceptable (each +/- 30 days).
18	12/22/2020	Laura J. Fochtmann, MD; Stony Brook University	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Calculating Change	It is difficult to give an answer to these questions without understanding how a change score would be used. It seems unlikely that each integer increment (or decrement) on the change score would necessarily represent a comparable change in "function" depending on what element or elements of the WHODAS was rated differently. (See https://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.2014.14050587 for one commentary). This would make a change score hard to interpret and could lead to erroneous inferences about the clinical significance of the change. Allowing patients to be counted more than once would

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							make matters even more complicated and confusing.
30	12/29/2020	Mirean Coleman, LICSW, CT; National Association of Social Workers	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support	Timeframe	A shorter time frame of 60 days allows time to modify treatment goals early on especially if goals are not working.
36	12/29/2020	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Overall Support w/ Mod	The WHODAS 2.0 may not be an optimal measure for mental health functioning since it has a large focus on physical functioning. I suggest the Sheehan Disability Scale to be either substituted or added as an option given that it takes less time for patients to complete and focuses on the areas most pertinent to behavioral disorders, i.e. social functioning, family functioning, and occupational functioning.
37	12/29/2020	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Timeframe	90 days is the same interval as most treatment planning intervals at clinics. The outcomes can inform treatment planning.
38	12/29/2020	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	General	19% is more realistic for a clinical population, as opposed to a research population, which may include treatment-resistant individuals

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
48	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Overall Support w/ Mod	We recommend staying with the currently proposed follow-up period. However, similar to the CSSRS outcome measure, we suggest for calculating the change between an index assessment and a subsequent assessment within the measurement period. In addition, we encourage an alternative numerator that includes a percentage reduction. However, we are confused by the options that are being proposed and where they come from. We reviewed the Ustun et al. (2010a) article, which reports some effect sizes, but do not know enough to suggest a specific percentage on this measure. Further, due to it's stronger focus on disability, sensitivity to change may be best in more severely disabled populations and may not be appropriate for higher functioning groups.
49	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	General	See above
50	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Timeframe	See previous comments.
51	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder	Support w/ Mod	Calculating Change	See previous comments.

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
				(Outcome Measure)			
52	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance of Functioning for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	General	See previous comments.
4	12/18/2020	Julie Ohnemus, MD; Open Door Community Health Centers	MCO - Managed Care Organization	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Overall Support w/ Mod	Have these assessment tools been proven in the population we serve (e.g. homeless)? How does the RAS assess with other co-morbidities involved?
19	12/22/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Do not support	General	As noted above recovery is an important issue, however, it has been conceptualized in multiple ways and its measurement accuracy with the recommended scale is unclear. Until further consensus is available on definitions of recovery and until robust measures are available, it should not be incorporated into a quality measure.
20	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Do not support	Timeframe	Please see above comments on the functioning measure relating to timing of this measure if it is adopted at the present time. [Functioning: Since the measures are constructed as a set, it is important that all of the time frames be consistent for all of the measures. For the functioning measure, 60 days seems too short to note a meaningful change in function for many patients, since significant functioning changes would typically lag behind symptom changes. 90, 120, or 180 days seem acceptable (each +/- 30 days).]

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
21	12/29/2020	Laura J. Fochtmann, MD; Stony Brook University	Provider	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Do not support	Calculating Change	<p>Please see above comments on the functioning measure relating to incorporating multiple measurements if this measure is adopted at the present time.</p> <p>[Functioning: It is difficult to give an answer to these questions without understanding how a change score would be used. It seems unlikely that each integer increment (or decrement) on the change score would necessarily represent a comparable change in "function" depending on what element or elements of the WHODAS was rated differently. (See https://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.2014.14050587 for one commentary). This would make a change score hard to interpret and could lead to erroneous inferences about the clinical significance of the change. Allowing patients to be counted more than once would make matters even more complicated and confusing.]</p>
39	12/29/2020	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	Overall Support w/ Mod	I would not prioritize this measure over other better established and more widely used measures of improvement and recovery, such as the PHQ-9 and GAD-7, which are the most widely used measures in the world for response and remission from mental disorders, and allow cross-cultural comparisons. Clinicians should be able to choose among a variety of approved measures, instead of locked into just one.
40	1/15/2021	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support w/ Mod	General	Coincides with treatment plan intervals
53	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance in Recovery for Individuals with a	Support	General	We recommend sticking with the 180 days (+/- 30 days). We do not have a super compelling rationale for this other than this is consistent with other existing outcome measures, and we simply

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
				Mental and/or Substance Use Disorder (Outcome Measure)			need more data to determine the optimal window. Similar to the other measures, it is challenging to comment sufficiently with the measures being currently under testing and therefore no change percentage value being offered. There might not be sufficient evidence to support a specific value, but again, trivial changes would presumably "count."
54	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support	Timeframe	See previous comments.
55	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Improvement or Maintenance in Recovery for Individuals with a Mental and/or Substance Use Disorder (Outcome Measure)	Support	Calculating Change	See previous comments.
1	12/14/2020	Jessica Holton, MSW, LCSW, LCAS	Provider	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support	General	The current Suicide Safety Plan on the APA Dashboard has several typos and errors. The errors do not interfere with the content.
22	12/29/2020	Laura J. Fochtmann, MD; Stony Brook University	Provider	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	Overall Support w/ Mod	The evidence of benefits of safety planning is relatively weak and the strongest studies are not specific to outpatient practice or general outpatient samples. Nevertheless, there is some available evidence and many of the steps of safety planning are consistent with good patient care in patients at risk of suicide. Thus, I would support use of this measure with modifications as outlined below.

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							Although the measure notes that any standardized assessment of suicidal ideation is acceptable to identify a need for a safety plan, it specifically refers to the C-SSRS Screen Version but not to the names of other possible measures. Although the C-SSRS is frequently mentioned in the context of suicide risk assessment, there are multiple problems with it. Once scales are locked into being part of a quality measure, it becomes less likely that additional research will be conducted to improve on or correct for flaws in existing measures.
23	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	Completion Threshold	As suggested by the developers of the safety plan, use of a higher threshold seems warranted. To collaborate with a patient in developing a safety plan can take 15-20 minutes or longer and would subsume the majority of time in a typical outpatient visit or a significant fraction of the time in an initial evaluation. While a safety plan may be useful, there are also significant opportunity costs involved. In other words, there will be other aspects of assessment, psychotherapy, or treatment planning that cannot be addressed at the visit due to time spent on safety plan development. It is possible that some of those aspects of care could be equally or more beneficial to addressing the underlying problems that the patient is facing. If the safety plans are not done in a thorough fashion, the benefits are likely to be much less robust and the goals of the measure would not be met.
56	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	Overall Support w/ Mod	The screener really does seem to be getting at the most critical assessment items, so we are sympathetic to the logic of sticking with the lower threshold. However, the higher threshold may be more consistent with how it is used in practice in a large psychiatric hospital. It does not help us reliably identify people who are at more immediate or acute risk to have the non-zero answer. The non-zero rule would essentially equate to at least answering yes to the question of wishing they were dead or could go to sleep

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							and not wake up (we use the threshold of the last 24 hours) and so folks may always answer yes to that question, but not ever have intentions or plans, and may not need the more formal intervention of the suicide safety plan. In utility, its more meaningful for us to intervene with updated safety plan and further assessment when people ALSO endorse "actually had thoughts of killing self and how they might do so, bringing them into the moderate risk category. Safety plan deployed at that level of risk makes more sense as there is a risk to intervene with.
57	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	Completion Threshold	See previous comments.
58	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	General	<ul style="list-style-type: none"> The time frame for follow-up for the safety planning measure (120 days +/- 30 days) is a very long time frame and would limit the utility of the measures in settings that provide briefer treatment. At least from the experience of users of the MBHR, this would be a limiting factor as the largest users of the APA registry would not have patients in active treatment in the same setting for this length of time. Increasing the flexibility of time frame would allow for utility in a broader range of settings, including those that may be doing brief crisis management interventions. The proposed measure denominator identifies any individual with non-zero score on the "C-SSRS Screen Version - Recent" or any other standardized assessment of suicide ideation or behavior as needing a Suicide Safety Plan. The developers of the Suicide Safety Plan have suggested that there should be a higher threshold for requiring the completion of a Suicide Safety Plan (e.g., those indicating having thought of how they would kill themselves) to prevent "watering down" of the effect of the brief intervention or

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>lack of fidelity to its use. Should APA and NCQA consider either a lower or higher threshold to define the denominator? o In terms of the non-zero score on the CSSRS, CRSPR or other standardized rating scale triggering the need for suicide safety plan, I agree with the developers of the safety plan intervention. Focusing on individuals with acute, moderate to high risk of suicide as those who will benefit most from a safety plan clinical intervention. Targeting all patients who have a lifetime history of suicide behaviors as requiring active safety plan intervention may create excessive burden and limit utility and intent of the measure. Recommend setting threshold at a higher score aligned with clinical indication of suicidal ideation and behavior being a focus of treatment. • The safety planning measure (process) identifies a non-zero score on the CSSRS as a trigger for requiring safety plan, OR clinician assessment of need, OR other standardized measure, providing a list of additional measures but without specifying scores that would trigger need for safety planning. This would need to be quantified in order for the measure to be operationalized and implemented. • Maintaining two numerators for initial process measure makes sense in order to support promotion of safety planning as a clinical intervention. • A couple of qualitative thoughts o Does the specification need to include requirement that the safety plan be initiated within 24 hours of completion of suicide risk evaluation indicating that safety planning is warranted? This is referenced, but not clearly specified in the definition specification. o In order to promote the clinical aspects of safety planning, it may be warranted to include documentation that the patient participated in the development of the plan and was provided a copy so as to minimize this being strictly a documentation and administrative measure.</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
5	12/18/2020	Julie Ohnemus, MD; Open Door Community Health Centers	MCO - Managed Care Organization	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Do not support	General	We refer out patients who have active suicide plans and care for those at mild & moderate risk
41	1/15/2021	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	Overall Support w/ Mod	Although safety planning is helpful for many individuals, the research evidence supporting it is still very preliminary with only one RCT. I have seen excessive reliance on safety planning sometimes hinder treatment and recovery, rather than decrease suicide risk. This intervention needs better study.
42	1/15/2021	Robert Gregory, MD; Upstate's Psychiatry High Risk Program	Provider	Initiation and Update to Suicide Safety Plan for Individuals with Suicidal Ideation, Behavior or Suicide Risk (Process Measure)	Support w/ Mod	Completion Threshold	See previous comments.
6	12/18/2020	Julie Ohnemus, MD; Open Door Community Health Centers	MCO - Managed Care Organization	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Do not support	General	Don't have the staffing to carry this out
24	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Do not support	Calculating Change	If this measure is adopted at the present time, please see above comments on the functioning measure relating to timing of this measure and incorporation of multiple assessments.
25	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Do not support	General	This measure is problematic in focusing only on suicidal ideas and behaviors as measured by the C-SSRS. As noted above, the C-SSRS is frequently mentioned in the context of suicide risk assessment; however, there are multiple problems with its design and if it becomes accepted as a mandatory measure of "quality

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>care," it is unlikely for additional research and improvements in assessment to take place.</p> <p>As with the safety plan measure, completion of the CSSRS, including the intensity measures, can be quite time consuming, if done correctly, leaving little time for any other assessment, psychotherapy, psychoeducation, treatment planning or other aspects of care during a visit.</p> <p>If the "since last visit" version of the scale is used, it would need to be completed at each visit to avoid missing key details about suicide attempts, including aborted or interrupted attempts. The "since last visit" approach may be fine for research studies in which individuals are typically excluded upfront for suicide risk, but will lead to a disproportionate focus on the C-SSRS items if used in many high-risk clinical settings.</p>
26	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Do not support	General	<p>Other problems with the C-SSRS include, but are not limited to, the following: The screening questions apply to a one month period or a time since last visit with no opportunity on the scale for documentation of current suicidal ideation.</p> <p>Although a longer look back is relevant to overall risk, current ideation is also crucial to risk assessment and for planning of need for a higher level of care.</p> <p>Many of the screening questions are actually multi-part questions, which makes it impossible for someone reviewing the scale to interpret a "Yes" answer to one of the questions.</p> <p>Patients (and staff) often have trouble distinguishing between questions about method (e.g., Have you been thinking about how you might do this?) and plan (e.g., Have you started to work out or have you worked out the details of how to kill yourself?). Adding to the confusion is that most clinicians view a suicide method as one</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>part of the plan not as entirely distinct from a plan.</p> <p>Distinguishing between some intent and intending to carry out a plan is challenging. Even when an individual has acted on suicidal ideas and made an attempt, intent is not always easy to determine.</p> <p>The numerical ratings and low/moderate/high risk ratings on the screening version can be misleading since they do not consider the whole host of other risk factors that are documented to be associated with increased suicide risk. This can, in turn, lead to false positives and false negatives, if specific numerical ratings are used as a threshold.</p> <p>In rating the intensity measures, providing ratings on the most severe type of ideation seems logical but may be misleading if the most severe type of ideation is rare and fleeting but a less severe type of ideation is persistent.</p> <p>With a significant look back period, it can be hard to select a single rating for frequency or duration, particularly if suicidal ideation has varied during the interval.</p>
27	12/29/2020	Laura J. Fochtman, MD; Stony Brook University	Provider	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Do not support	General	<p>In terms of the controllability of thoughts, patients often have trouble making these distinctions. If thoughts did not result in an attempt, they were obviously controllable to some degree and the fine points of controllability may not be clear to them.</p> <p>Similarly, it can be hard to elicit the extent to which deterrents prevented suicidal behavior. If no attempt was made, then some sort of deterrent was likely to be present, particularly in the context of frequent suicidal ideas.</p> <p>Eliciting the reasons for ideation is not straightforward either. Since individuals can make a serious suicide attempt or die by suicide with either set of reasons, the inclusion of this item as</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>a dimensional measure of intensity seems questionable.</p> <p>The use of the intensity score has been less well studied than other aspects of the scale and there is less information available about the ways that these elements contribute to risk than there is information about many other risk factors (e.g., specific diagnoses, prior suicide attempts, recent hospitalizations, psychosocial stressors, demographic factors, family history of suicide).</p> <p>For situations in which an attempt has occurred, the potential lethality is rated only if no injury took place, however, there could be significant potential for lethality yet only sustain a minor actual injury (e.g., holding gun to head and firing but only sustaining a graze injury).</p>
59	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Support w/ Mod	Overall Support w/ Mod	<p>We recommend calculating the change between an index assessment and a subsequent assessment within the measurement period. Regardless of this approach or using the last assessment within 90 days, the magnitude of change does not seem to be addressed (so, a change of 1 point is sufficient?). The latter proposal might deter folks from taking on people who tend to have fluctuations in their level of suicidality over a treatment course, making people with personality disorders for example, less likely to be assessed when providers are thinking ahead to this reporting ?</p>
60	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Support w/ Mod	Calculating Change	<p>See previous comments.</p>
61	1/17/2021	Vaile Wright, PhD; American Psychological Association	Specialty Society	Reduction in Suicidal Ideation or Behavior Symptoms (Outcome Measure)	Support w/ Mod	General	<ul style="list-style-type: none"> The time frame for follow-up for the outcome measure using the CSSRS measure (90 days +/- 30 days) is a very long-time frame and would limit the utility of the measures in settings that provide briefer treatment. At least from the experience of users of the MBHR, this would be a limiting factor

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>as the largest users of the APA registry would not have patients in active treatment in the same setting for this length of time. • Potential issues with using the CSSRS as a symptom monitor for suicidal ideation and behaviors arise from its structure as a qualitative risk assessment tool. It was not originally designed to have quantitative scores assigned, and experience with attempting to quantify and assign a score and risk level to individual items have proven to be fairly complex. Thus, problems may arise in different systems creating different types of branching logic and assigning different values to some of the intermediate level items given that a single uniform scoring algorithm has not been validated and published as a "gold standard" for operationalizing a scoring system for the instrument. • Determining a meaningful score reduction to indicate adequate response to treatment will be challenging, and will differ depending on case mix and severity of acute risk at the outset of treatment. While clinicians providing treatment for suicidal behaviors may appreciate this and be able to use reassessment using the CSSRS with individual patients, it will be much more challenging to capture this in a MIPS type measure that is reviewed at a population level by payors and other non-MH reviewers. • We have not been able to identify existing literature (in very brief search) identifying use of the CSSRS as a PROM. Validation of the utility of the tool for this purpose may be needed before full usage in registry is indicated. • Denominator Identification Period: The period in which individuals can have an encounter with a baseline assessment using the C-SSRS+. The denominator encounter period is the 10-month window starting on November 1 of the year prior to the measurement year and ending on September 1 of the measurement year. While the denominator identification needs to ensure adequate timeframe for follow-up during the reporting period is needed, not using rolling periods does</p>

Comment Number*	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization*	Measure Set or Measure	Measure Support	Measure Element	Text of Comments
							<p>arbitrarily eliminate people who enter treatment during the period between Sept 1 and Nov 1. Also, see earlier bullet about length of time for follow-up. Recommend more flexibility to not exclude brief treatment settings. • If a patient has more than one follow-up assessments completed with the C-SSRS+ in the numerator time period with different scores, how should this be handled in the measurement? Should APA and NCQA consider calculating the change between the baseline assessment and the last assessment within that period? Or should APA and NCQA consider and index assessment for each measurement and calculate a change between the index assessment and any subsequent assessment within the measurement period, which would allow for the patient to be counted more than once. o Recommend calculating the change score between the baseline assessment and the lowest score on repeat administration during the measurement period. Do not think it is best to count individual patients multiple times during one measurement period/episode of care, but important to recognize maximum response to treatment. This can be challenging for some patients who will have fluctuating levels of suicidal ideation and behaviors, and want to avoid suppressing use of measurement-based care for those individuals</p>

*Optional