

Measure Development for the  
Quality Payment Program  
(Mental Health/Substance Use Care)



## TECHNICAL EXPERT PANEL: SUMMARY OF ACTIVITIES

**Date of meeting: June 3, 2019**

Date of final summary document: July 29, 2019

This document was prepared by the American Psychiatric Association (APA) under cooperative agreement with the Centers for Medicare and Medicaid Services (CMS), an agency of the U.S. Department of Health and Human Services. The contents presented do not necessarily reflect CMS policy.

# Contents

TECHNICAL EXPERT PANEL: SUMMARY OF ACTIVITIES.....	1
Background .....	4
The APA Measure Development Initiative Technical Expert Panel .....	4
June TEP Meeting Summary .....	5
Table 1. The meeting’s five measure-specific subgroups. ....	5
Table 2. Pre-meeting information provided to panelists.....	5
Opening Remarks and Meeting Format.....	6
TEP Composition and Roll Call .....	7
Table 3. Panelist information.....	7
Overview of Discussion Procedures and Summary of Content .....	9
Summary of Discussion of MBC Process Measures .....	10
Standardized Assessment .....	10
Key discussion points and consensus: .....	10
Table 4: Results from the quantitative survey on Standardized Assessment.....	12
Monitoring .....	13
Key discussion points and consensus: .....	13
Table 5. Results from the quantitative survey on Monitoring.....	14
Treatment/Care-Plan Adjustment .....	15
Key discussion points and consensus: .....	15
Table 6. Results from the quantitative survey on Treatment/Care-Plan Adjustment.....	17
Summary of Discussion of MBC Outcome Measures .....	18
Functional Impairment.....	18
Key discussion points and consensus: .....	18
Table 7. Results from the quantitative survey on Functional Impairment .....	20
Recovery.....	21
Key discussion points and consensus: .....	21
Table 8. Results from the quantitative survey on Recovery.....	22
Other Information Arising from TEP Discussions.....	23
General Comments that Apply across Two or More Measures.....	23
Discussions on Selection of Assessment Tools .....	23
Standardized Assessment .....	23
Functional Impairment.....	24

Recovery..... 24

Summary of Discussion of Measure ‘Reduction or maintenance of symptoms for patients with opioid misuse’ ..... 25

    Key discussion points and consensus: ..... 25

Next Steps ..... 26

## Background

In September 2018, the Centers for Medicare and Medicaid Services (CMS) awarded the American Psychiatric Association (APA) funding for measure development as part of the Medicare Access and CHIP Reauthorization Act (MACRA) of 2015. APA entered into a cooperative agreement with CMS and contracted with the National Committee on Quality Assurance (NCQA) as the technical experts in measure development. The goal of the funding award is to support the development of meaningful quality measures that fill CMS-designated high priority areas, including mental health and substance use disorders. These measures are intended to reduce data collection burden for providers who wish to systematically track the provision and quality of care for individuals treated for mental and substance use disorders. Further, the measures are intended to provide useful information to both patients and providers for informing care or quality improvement. The proposed quality measures will be subject to federal rulemaking for inclusion in CMS's value-based payment program, the Merit-Based Incentive Payment System (MIPS).

### The APA Measure Development Initiative Technical Expert Panel

The measure development team convened the APA Measure Development Initiative Technical Expert Panel (TEP), comprising 18 individuals with a broad range of expertise that includes adult general psychiatry, substance use disorders, suicide, and psychotic disorders. These experts represent a diverse national perspective. The TEP consists of clinical experts in quality improvement who represent a variety of viewpoints and backgrounds (i.e., clinical care, clinical quality improvement and/or patient safety activities, medical education, research, practice administration, and health plan administration). The TEP participates in educational webinars and in-person meetings on measure development and meaningful discussions on the quality measure topics, feasibility of data-capture as part of the workflow, unintended consequences, burden, and other issues posed by the quality measures under development. The TEP also considers recommendations from the APA Measure Development Initiative's Consumer Family Panel (CFP) to ensure that the quality measures developed are useful for those under the providers' care.

The TEP provides feedback on key methodological and clinical decisions as they relate to the development of provider-level meaningful quality measures. Examples include informing on assessment tool selection for initial and continued use within the measurement-based care (MBC) quality measure framework, addressing potential problems with a measure's use within existing clinical workflows, or after the completion of each phase of testing. The TEP evaluates multiple assessment tools and iterations of the quality measures during the project.

The TEP also includes 3 CFP liaisons. Their role is to ensure that the CFP perspective is accurately communicated during the TEP meetings. This occurs as part of the materials presented by the measure developers during TEP webinars and in-person meetings, as well as the liaisons' engagement in TEP meeting discussions.

## June TEP Meeting Summary

This report summarizes the in-person TEP meeting convened June 3, 2019. The main goals for the meeting were for TEP members to:

1. Review and provide feedback on the draft specifications for the MBC process measures, which involved discussing and then, in an in-meeting electronic survey format, answering measure specific questions related to each aspect of the measure (i.e., denominator, numerator, exclusions, and exceptions)
2. Review and provide feedback on the draft specifications for 2 of the 5 MBC outcome measures (i.e., Reduction in [or Maintenance of] Functional Impairment and Improvement in [or Maintenance of] Recovery). Similar to the first goal, TEP discussion of measure specific questions related to each aspect of the measure (i.e., denominator, numerator, exclusions, and exceptions) was followed by the use of in-meeting surveys to get consensus to inform refinement of the specification of the measures.
3. Discuss the importance and usability of the proposed quality measure “Reduction in OUD Symptom Severity/Opioid Misuse” and its associated assessment tools. The discussion of importance and usability of this measure was carried forward to the June 3, 2019 meeting because time constraints at the March 23, 2019 TEP meeting prevented the measure developers from presenting this quality measure topic and its associated tools.

To meet these goals, the TEP worked through the measure developers’ key questions, including some based on feedback from the Post-meeting Follow-up surveys conducted shortly after the respective March 21, 2019 and March 28, 2019 CFP and TEP meetings. The purpose of the March 2019 Post-meeting Follow-up surveys was to collect information on the importance and usability of the proposed MBC measures under development.

In advance of the June 3, 2019 TEP meeting, the measure developers assigned each panelist, including CFP liaisons, to quality measure-specific subgroups (Table 1). Each subgroup consisted of approximately 6 panelists and were organized to reflect as much as possible the diversity of professional backgrounds of the full TEP membership. The three CFP liaisons were assigned separately to the subgroups on Standardized Assessment, Functional Impairment, and Recovery. Each subgroup was staffed by several APA and NCQA measure developers. Prior to the June TEP meeting, panelists were assigned to subgroups and provided pre-meeting information/materials specific to their respective subgroups as well as information/materials applicable to all panelists, regardless subgroup assignment (Table 2).

Table 1. The meeting’s five measure-specific subgroups.

<b>Process measure subgroups:</b>	<b>Outcome measure subgroups:</b>
Standardized Assessment	Reduction in (or Maintenance) of Functional Impairment
Monitoring of Symptoms, Functioning, and Recovery	Recovery
Treatment/Care-Plan Adjustment	

Table 2. Pre-meeting information provided to panelists.

<b>Full TEP:</b>	<b>Subgroups:</b>
TEP meeting agenda	Subgroup-specific MBC Business Case
March 28, 2019 TEP Meeting Summary	Draft quality measure specifications with questions
March 21, 2019 CFP Meeting Summary	All quality measures’ assessment tools
March Post-Meeting Survey Results from CFP and TEP	
Project update (March 29 – June 3)	

Full TEP:	Subgroups:
Opioid Misuse Topic and Reduction in Opioid Use Disorder (OUD) Symptom Severity Slides	

### Opening Remarks and Meeting Format

Kristin Kroeger, APA’s Chief of Policy, Programs, and Partnerships, welcomed attendees and conveyed APA’s appreciation for their participation and commitment to the Initiative. TEP co-chairs (Drs. Halverson and Ratzliff) and co-project director (Ms. Shugarman) stated the goals (listed above) and format of the meeting (i.e., measure-specific subgroup discussions, full TEP discussions, and measure-specific surveys at the end of each full TEP discussion to help fine-tune the specification of each measure covered in the meeting).

## TEP Composition and Roll Call

Meeting participants—the panelists, measure development team, and additional APA leadership staff—provided introductions and affirmed that there were no changes to their conflicts of interest.

Table 3. Panelist information.

<b>Name, Credentials, and Professional Role</b>	<b>Organizational Affiliation, City, State</b>	<b>Consumer Perspective</b>	<b>Clinical Content</b>	<b>Performance Measurement</b>	<b>Coding and Informatics</b>	<b>Conflict of Interest Disclosure</b>
<b>Andrew Sperling, MA, JD</b> <i>Liaison to the Consumer Family Panel</i> Director, Legislative Affairs Advocacy & Public Policy	National Alliance on Mental Illness (NAMI) Arlington, VA	Yes	No	No	No	No Conflict of Interest
<b>Anna Ratzliff, MD, PhD</b> <i>TEP Co-Chair</i> Associate Professor	University of Washington Seattle, WA	No	Yes	Yes	No	No Conflict of Interest
<b>Arthur Robin Williams, MD, MBE</b> Assistant Professor of Psychiatry/Research Scientist	Columbia University Dept of Psychiatry, Division of Substance Use Disorders New York, NY	No	Yes	No	No	No Conflict of Interest
<b>Caroline Carney, MD, MSc, FAPM, CPHQ</b> Chief Medical Officer	Magellan Health, RX Management Phoenix, AZ	No	Yes	Yes	No	No Conflict of Interest
<b>Elizabeth W. McKune, Ed.D., PCMH-CCE</b> Vice President, Health Integration	Passport Health Plan Louisville, KY	No	Yes	Yes	No	No Conflict of Interest
<b>Jerry Halverson, MD, DFAPA</b> <i>TEP Co-Chair</i> Chief Medical Officer	Rogers Behavioral Health Oconomowoc, WI	No	Yes	Yes	No	No Conflict of Interest
<b>Jill Harkavy-Friedman, PhD</b> VP, Research	American Foundation for Suicide Prevention New York, NY	No	Yes	Yes	No	No Conflict of Interest
<b>Jolene Rasmussen, MSCE</b> Director of Adult Behavioral Health	Texas Council of Community Centers Austin, TX	No	No	Yes	Yes	No Conflict of Interest
<b>Jose P Vito, MD, DFAPA</b> Forensic Tele-psychiatrist	New York State Office of Mental Health New York, NY	No	Yes	Yes	No	No Conflict of Interest
<b>Kari A. Stephens, PhD</b> Associate Professor	University of Washington Seattle, WA	No	Yes	Yes	Yes	No Conflict of Interest
<b>Kyaïen O' Quinn Conner, PhD, LSW, MPH</b>	University of South Florida	No	Yes	Yes	No	No Conflict of Interest

<b>Name, Credentials, and Professional Role</b>	<b>Organizational Affiliation, City, State</b>	<b>Consumer Perspective</b>	<b>Clinical Content</b>	<b>Performance Measurement</b>	<b>Coding and Informatics</b>	<b>Conflict of Interest Disclosure</b>
Assistant Professor	Tampa, FL					
<b>Lee Flowers, MD, MPH</b> Psychiatrist	Aspire Locums LLC Rutherford, NJ	No	Yes	Yes	No	No Conflict of Interest
<b>Lisa A. Ryer, LCSW</b> Mental Health Clinician II	Rutgers University Behavioral Health Care New Brunswick, NJ	No	Yes	Yes	No	No Conflict of Interest
<b>Nathaniel Z. Counts, J.D.</b> <i>Liaison to the Consumer Family Panel</i> Associate Vice President of Policy	Mental Health America (MHA) Alexandria, VA	Yes	No	No	No	No Conflict of Interest
<b>Volunteer Advocate 1</b> <i>Liaison to the Consumer Family Panel</i>	Volunteer Advocate Baltimore, MD	Yes	No	No	No	No Conflict of Interest
<b>Perry Meadows, MD, JD, MBA, FAFP</b> Medical Director, Government Programs	Geisinger Health Plan Danville, PA	No	Yes	Yes	No	No Conflict of Interest
<b>Robert Schloesser, MD</b> Executive Director, SPL Research Institute	Sheppard Pratt Health System Towson, MD	No	Yes	Yes	No	No Conflict of Interest
<b>Shuba Samuel, PhD, RN, FNP-BC, APNP, CEN, CNE</b> Family Nurse Practitioner	Oscar G. Johnson VA Medical Center Manistique, MI	No	Yes	No	No	No Conflict of Interest
<b>Tanni M. Bromley, MPAS, RPA-C</b> Associate Director of Behavioral Health, East	Landmark Health Batavia, NY	No	Yes	Yes	No	No Conflict of Interest
<b>Thomas Smith, MD</b> Deputy Chief Medical Officer	New York State Office of Mental Health New York, NY	No	Yes	Yes	No	No Conflict of Interest
<b>William W. Bruck, MSN, APN, FNP-BC, CARN-AP</b> Director of Medical Services	Seabrook-The Heart of Recovery Bridgeton, NJ	No	Yes	Yes	No	No Conflict of Interest

Note: Nathaniel Counts was unable to attend a portion of the afternoon full TEP discussion.

## Overview of Discussion Procedures and Summary of Content

Each subgroup met for an hour to have group-specific discussions of the draft measure specifications and questions related to each aspect of the measure (i.e., denominator, numerator, exclusions, exceptions). Each subgroup selected a representative who reported on the subgroup's discussion to the full TEP. Immediately following the full TEP's discussion of each measure, all panelists completed an in-meeting electronic survey about the measure (e.g., after the full TEP discussion on the Standardized Assessment measure, TEP members completed an electronic survey specifically about the Standardized Assessment measure). The survey, administered using REDCap survey software, included qualitative and quantitative questions about the measure. In the following sections, we summarize the key discussion points and consensus arising from each of the measure specification discussions. The qualitative and quantitative survey results are embedded within the summary text. The full set of quantitative survey questions and the accompanying results are also presented in a table at the end of each respective section.

## Summary of Discussion of MBC Process Measures

### Standardized Assessment

**Measure Name:** Measurement-Based Care: Initial standardized assessment for all patients seen for mental health and/or substance use care

**Brief Measure Description:** The percentage of individuals 18 years and older presenting with a mental and/or substance use disorder or indication, who have an initial assessment concurrent with or prior to an encounter in at least five (5) mental health domains including depression, anxiety, substance use, suicide risk and psychosis, as well as an initial assessment of global functioning and recovery.

Key discussion points and consensus:

#### Numerator Specification

- **Definition of an encounter.** In the quantitative survey results, 100% of respondents (Table 4) endorsed flexibly defining an ‘encounter’, which would include face-to-face office visit, any face-to-face visit (e.g., outpatient group session), telehealth visit, or other encounter.
- **Period of assessment.** Panelists discussed the cut-off period for conducting an initial assessment (and a monitoring assessment).
  - Panelists agreed multiple encounters could be allowed to count as an assessment in the numerator, up to two encounters per new episode. In terms of length of time, the TEP determined it was acceptable to allow up to two months from the first encounter to capture patient-reported data for the initial assessment.
  - The qualitative data affirmed allowing providers latitude in completing the initial assessment over two encounters. The qualitative data extended the discussion with several respondents suggesting the initial assessment workload could be shared with fellow providers (e.g., in the case of a care team).

#### Denominator Specification

- The TEP agreed a broad denominator should be used to ensure the capture of all mental health patients, rather than a denominator that is limited to a short list of specific diagnostic groups. In the qualitative data, several respondents suggested diagnoses (with appropriate use of diagnostic codes) should be part of the definition of the denominator, as a lack of information about diagnoses complicates data analysis/interpretation. Further, survey respondents stated diagnostic information would be useful across specialties. A new patient or an existing patient is one who has not been seen by the provider in the previous 6 months (also supported by the quantitative survey results: see Table 4).
- In discussion, the TEP reached consensus that the initial assessment may occur across up to 2 encounters that occur within at least 2 months. The qualitative data showed the TEP did not reach consensus about the need for 2 encounters (versus 1), but the qualitative data resonated with the TEP discussion that no more than 2 encounters should occur for completion of the initial assessment.
- Respondents indicated that a 2-month time frame seems sufficient to capture the data, acknowledging the fact that some patients will experience improvement in the initial weeks after the first encounter.

### **Exclusion/Exception Specification**

- **Patient refusal.** The TEP stated patient refusal should be documented but needs further discussion for consideration as an exclusion because a high rate of refusal should be a sign to the provider to reevaluate their patient engagement. Clinicians who employ patient engagement strategies can mitigate this challenge. Panelists noted reporting requirements should include counts of exclusions/exceptions to support improvements.
- **Psychiatric crisis and psychiatric/medical impairment.** The TEP recommended psychiatric crisis and psychiatric/medical impairment be considered as exclusions: psychiatric crisis as defined by the utilization of a crisis code; psychiatric and/or medical impairments, such as the inability to complete patient-reported outcome assessment due to acute symptoms of dementia, psychosis, intoxication, or medical conditions. Providers should have the option to write a note about their determination of cases of impairments. In the case of caregivers or other proxies, there should be an option for “don’t know”. The post-discussion survey qualitative data affirmed these points.

### **Other Discussion Points about this Measure**

- **Incentivizing MBC as a standard of care.** The TEP emphasized that implementation of MBC processes incentivizes continued use of MBC. For example, providers want to see positive patient health outcomes, which is known to be associated with the use of MBC. Also, standardized quantifiable outcomes gathered with MBC can help clinicians more precisely assess the impacts of prescribed interventions (e.g., medication and/or psychotherapy), including the need for more acute interventions and additional encounters for hard-to-treat patients.
- **Behavioral healthcare across settings.** The TEP discussed that behavioral health care outcomes may be the result of care received across various outpatient or ambulatory care settings and provider-types. The TEP discussed the potential requirement for appropriate attribution where psychiatric specialty practices are not the only setting (with the required specialty provider-types) in which an individual may have receive treatment for mental health pertinent to the outcomes observed.

Table 4: Results from the quantitative survey on Standardized Assessment.

<b>What defines baseline or an initial assessment?</b>	<b>N (%)</b>
1. New to provider	1 (4.8)
2. New episode not seen in six months	3 (14.3)
3. Both	17 (81.0)
<b>What defines an encounter?</b>	<b>N (%)</b>
1. Face-to-face office visit	0 (0.0)
2. Any face-to- face visit	0 (0.0)
3. Telehealth	0 (0.0)
4. Other (please specify)	0 (0.0)
5. All of above	21 (100)
<b>Who should be included in the denominator?</b>	<b>N (%)</b>
1. All patients seen for psychiatric care	3 (14.3)
2. Those with up to two encounters	9 (42.9)
3. Those with a diagnosis during the episode	6 (28.6)
4. Other (specify)	3 (14.3)
<b>What standardized assessments comprise an initial assessment?</b>	<b>N (%)</b>
1. Assessment of symptoms only	0 (0.0)
2. Assessment of function only	0 (0.0)
3. Assessment of recovery only	0 (0.0)
4. Assessment of symptoms and function but not recovery	7 (33.3)
5. Assessment of symptoms and recovery but not function	0 (0.0)
6. Assessment of function and recovery but not symptoms	0 (0.0)
7. Assessment of symptoms, function, and recovery but reported separately	14 (66.7)
8. Other (please specify)	0 (0.0)
<b>Should there be separate rates along with a composite (i.e. percent of patients with all relevant)?</b>	<b>N (%)</b>
1. No	2 (9.5)
2. Yes	16 (76.2)
3. Unsure	3 (14.3)
<b>Are the proposed exclusions/exceptions reasonable? (Noted on page 10.)</b>	<b>N (%)</b>
1. No	0 (0.0)
2. Yes	16 (76.2)
3. Unsure	5 (23.8)

## Monitoring

**Measure Name:** Measurement-Based Care: Monitoring of symptoms, functioning, and recovery for all patients seen for mental health and/or substance use care.

**Brief Measure Description:** The percentage of patients 18 years and older presenting with a mental and/or substance use disorder or indication who are monitored for improvement or maintenance of symptom severity, functional impairment, and recovery over a twelve-month period using standardized assessments.

Key discussion points and consensus:

### General Specifications

- As with the Standardized Assessment, the TEP ultimately communicated a preference for tools consisting of a wide range of symptoms rather than only select diagnostically specific tools. The subgroup had suggested limiting the number of diagnoses included in the denominator of this quality measure. The rationale was to focus on individuals with active symptoms for “treatable diagnoses”, including anxiety, depression, and schizophrenia. However, the majority of the TEP felt that limiting monitoring to specific diagnoses misrepresents the intent of the monitoring measure and could create disincentives, impacting negatively on the care of patients with other conditions. Most respondents in the qualitative data also indicated a preference for not limiting the monitoring measure to select diagnoses. (Some respondents felt select categories might be a ‘good starting point,’ but most respondents questioned why certain diagnoses were not named (e.g., bipolar disorder, substance use disorder) while comorbid diagnoses were named (e.g., schizophrenia).)
- The TEP stated it is important to document reasons why individuals refused to or could not complete an assessment. In the qualitative data, respondents expressed concern about provider burden and said collecting outcome information is sufficient.
- Regarding proxies, some panelists were conflicted on whether to collect information from a patient proxy (e.g., family member/caregiver) with tools that are not as strongly validated for this type of informant. Given the desire to collect this data as appropriate, panelists suggested that proxy data collection should be limited.

### Numerator Specification

- The quantitative survey results showed the majority of respondents (91%; Table 5) indicated a minimum of two assessments in a measurement year.

### Denominator Specification

- The panelists agreed that the Monitoring and Treatment/Care-Plan Adjustment [discussed below] measures would have the same denominator statements. In the quantitative survey, a majority of respondents (62%; Table 5) stated the denominator should include patients with an initial assessment (as opposed to all patients regardless of whether an initial assessment has been completed [38%]).
- A set of panelists commented that the utilization of evaluation and management billing codes may help ensure adequate assignment to the measures’ denominator. However, other panelists explained that not all practices attempt to limit an individual’s co-payment responsibility and do not bill for screening during an encounter. As an alternative, panelists suggested that diagnostic codes (e.g., ICD-10-CM) could inform when individuals are included in the quality measure’s denominator.

Measure developers agreed to examine this and other codes more closely so appropriate denominator assignments are made in the registry.

### Exclusions/Exceptions Specification

- The subgroup presented exclusion criteria details including “patient refusal or reason patient is unable to complete patient-reported outcome measures (suggested via key word mapping), and patient’s transfer of care.”
- The TEP agreed those requiring active treatment will automatically remain in the denominator; however, the panel questioned whether those in remission should remain in the denominator by asking at what point in longitudinal care would the measure no longer demonstrate meaning to the provider and patient?” The panel determined this should be examined during testing.
- Some panelists mentioned the potential for transfer of the patients’ care to another provider within the same practice or to an outside provider. The panel emphasized that patients who move between providers within the same practice will have the quality measure data collected during appropriate encounters. However, a new practice would start with an initial assessment or have the quality measure data transmitted.

Table 5. Results from the quantitative survey on Monitoring.

<b>Should the denominator include patients (as you indicated above) with an assessment (i.e. the standard initial assessment measure denominator) or all patients eligible for assessment regardless of whether the assessment was completed?</b>	<b>N (%)</b>
1. Patients (as you indicated above) with an assessment	13 (61.9)
2. All patients eligible for assessment regardless of whether the assessment was completed	8 (38.1)
<b>The small group recommended a minimum of 2 assessments in the measurement year, do you agree?</b>	<b>N (%)</b>
1. No	0 (0.0)
2. Yes	19 (90.5)
3. Unsure	2 (9.5)
<b>Should the monitoring measure require both a functioning and symptoms tool for all patients or one or the other?</b>	<b>N (%)</b>
1. Both symptoms and functioning	20 (95.2)
2. Symptom or functioning	1 (4.8)
<b>Should assessments be required to occur within a set time period?</b>	<b>N (%)</b>
1. No	3 (14.3)
2. Yes	14 (66.7)
3. Unsure	4 (19.0)
<b>Are the proposed exclusions/exceptions reasonable? (Noted on page 13 in text above.)</b>	<b>N (%)</b>
1. No	1 (4.8)
2. Yes	15 (71.4)
3. Unsure	5 (23.8)
<b>Are there others that should be considered?</b>	<b>N (%)</b>
1. No	17 (81.0)
2. Yes	1 (4.8)

## Treatment/Care-Plan Adjustment

**Measure Name:** Measurement-Based Care: Treatment or care plan adjustment for all patients seen for mental health and/or substance use care.

**Brief Measure Description:** The percentage of patients 18 years and older with a mental and/or substance use disorder or indication who had their treatment or care plan adjusted.

Key discussion points and consensus:

### General Specifications

- **Provider voice.** Panelists noted that Treatment/Care-Plan Adjustment should not be limited to behavioral health providers because some practices engage in bidirectional communication between primary and behavioral health care providers. Therefore, it is important the measure's workflow does not hinder the bidirectional relationship of the care providers.

### Numerator Specification

- **Intent of the measure.** In discussion, panelists emphasized the measure's intent to demonstrate provider provision of care. For instance, a medication or psychotherapeutic intervention adjustment was made, or the individual was maintained on the current medication or psychotherapeutic intervention, and the provider made a notation explaining the treatment decision. Panelists confirmed to the measure developers that additional interventions are not always the best decision. Further, treatments are unidentifiable through claims data due to lack of clinical details regarding what occurs during an encounter. This presents barriers when considering other treatment options. Further, an individual's course of treatment may involve subtle changes. The quality measure must be specified in a manner that is objective and quantifiable.
- **Assessment frequency.** Per the quantitative survey results (Table 6), a slight majority (51%) of panelists indicated the assessment should be conducted quarterly, and nearly one quarter of panelists (24%) indicated the assessment should be conducted monthly.
- **Language that reflects patient voice:** Panelists suggested incorporating language in the numerator reflective of patient-reported information. Examples include whether their care provider spoke with them about the results of the reassessment tool, if they received the care that they believed they needed, or if they were aware of their care plan.

### Denominator Specification

- During discussion, most panel members, but not all, stated that the Treatment/Care-Plan Adjustment measure's denominator should include all patients. The quantitative survey results (Table 6) reflected the discussion, with slightly more than half of respondents (57%) recommending the monitoring measure include all patients in the denominator.

### Exclusions/Exceptions Specification

- During discussion, the TEP determined that the Treatment/Care-Plan Adjustment measure should maintain the same exclusion criteria as the Monitoring quality measure.
- In the qualitative data, respondents suggested exclusions could include: refusals; individuals with inactive symptoms, current crises, cognitive impairment or who are otherwise unable to reasonably respond to patient-reported outcome measures; and individuals without an active treatment plan.

**Other Discussion Points about this Measure**

- **Data sources of this measure for testing.** The TEP recommended the measure developers consider using patient-reported information as the data source for the treatment/care plan adjustment measure. The measure developers will test this approach alongside the testing of the approach based on data collected with clinical documentation of adjustment to treatment/care plan. Specifically, both options will be tested in the alpha testing of the measure but only 1 of the 2 options (i.e., the more feasible and clinically useful) will move forward to beta testing.
- **Individual and provider factors.** The panelists discussed various factors (e.g., burden on providers to administer assessment tools, whether assessment items are structured or open-ended) associated with Treatment/Care-Plan adjustments (and Monitoring) for individuals with mental and/or substance use disorders who are seen within specialty and primary care treatment settings. Several panelists noted that adjustments to a given patient’s care plan depend upon provider and patient perception of whether adjustment is warranted and in what direction. For instance, among patients prescribed medication, panelists stated adjustment may entail maintenance, decrease, or increase in medication.
- **Incorporating provider input.** In the qualitative data, some respondents suggested incorporating provider input into assessment of clinical decision-making. Other issues raised about this recommendation included: additional burden on clinicians to administer many scales; the relevance of the construct of perception of care as a form of MBC; and whether items about clinical decision-making data would be structured or open-ended.
- **Other recommendations.** Other recommendations in the qualitative data included: conducting a scoping review on the most meaningful patient-reported outcomes; delaying the implementation of the measure to reduce burden on clinicians; and requesting more discussion about the meaning of treatment adjustment.
- **Potential benefits and harms.** This measure topic evoked conversation on potential benefits and harms, given recognition of the measure’s potential to elicit care practices that improve care quality. The table below provides a list of potential benefits and harms named by panelists.

Potential Benefits	Potential Harms
Prevention of treatment inertia by removing patients from psychotropic medications that aren’t treating clinical symptoms.	Creation of unintended consequences to patient safety by imposing unnecessary treatment changes.
Incentivization of data collected and utilized while individuals are under the providers’ care.	Imposition of unnecessary treatment adjustments that are not meaningful for patient outcomes.
Promotion of clinician-patient engagement.	Concerns for individuals whose symptoms fluctuate. For someone experiencing a short-term symptom decline, providers could hesitate to count the individual in their quality measure rate, particularly if they feel their performance is based on this outcome.
Incentivization of patient and clinician communication.	Perversion of incentives to providers who could update medications or other interventions to increase their quality measure score.

Table 6. Results from the quantitative survey on Treatment/Care-Plan Adjustment.

<b>How often/in what time frame should the functional assessment tool be administered?</b>	<b>N (%)</b>
1. Every visit	3 (14.3)
2. Monthly	5 (23.8)
3. Quarterly	11 (52.4)
4. Every 6 months	2 (9.5)
<b>The small group recommended including all patients in the monitoring measure in the treatment adjustment denominator (patients with a diagnosis of depression, anxiety, or schizophrenia with active symptoms). Do you agree?</b>	<b>N (%)</b>
1. No	7 (33.3)
2. Yes	12 (57.1)
3. Unsure	2 (9.5)
<b>Should the denominator be restricted to individuals who have not improved?</b>	<b>N (%)</b>
1. No	14 (66.7)
2. Yes	3 (14.3)
3. Unsure	4 (19.0)
<b>Should this measure apply to all or specific outcomes?</b>	<b>N (%)</b>
1. All outcomes	20 (95.2)
2. Specific outcomes	1 (4.8)
<b>[If option 2 in the previous question is chosen] Please select the specific outcome/s:</b>	<b>N (%)</b>
1. Symptoms	1 (4.8)
2. Functional impairment	0 (0.0)
3. Recovery	0 (0.0)

## Summary of Discussion of MBC Outcome Measures

### Functional Impairment

**Measure Name:** Reduction or maintenance of functioning for all patients seen for mental health and substance use care.

**Brief Measure Description:** The percentage of individuals aged 18 years or older with mental or substance use disorder who demonstrated a reduction in functional impairment (or maintained baseline level of functioning) based on results from a standardized assessment tool.

Key discussion points and consensus:

#### General Specifications

- **Change in measure name.** Panelists discussed modifying the outcome measure's title to "Improvement in or maintenance of functioning" or a similar wording that is more positive (i.e., by taking the focus off of "impairment"). Final phrasing was not determined during the discussion. The qualitative data also supported renaming the measure with more positive/neutral wording.
- **Assessments for inclusion.** Panelists considered whether process and outcome measures for functioning should be operationalized utilizing distinct assessment tools that are applied at specified time points, or whether the same assessment tool should be repeatedly implemented. In either case, the TEP agreed the specifications should include clear timeframes for when to administer the assessments.

#### Numerator Specification

- **Assessment period and frequency.** Panelists agreed a minimum of 2 assessments must be completed per measurement year, with more frequent assessments permissible, considering some individuals may have many encounters in a measurement year. Panelists discussed the consequences of an assessment completed during the fourth quarter of the year, including whether the performance rating of the assessment data rolls over to the following year. Panelists also called attention to the fact some patients' conditions improve rapidly (e.g., within 4 to 6 weeks). Because of this, these panelists suggested quarterly (or more frequent) assessments would be necessary for tracking symptoms of these patients. Less frequent assessments might not capture rapid changes in symptoms.

#### Denominator Specification

- **Broad-based denominator's potential to engage a greater number of patients.** The TEP agreed the specifications should include a broad-based denominator, comprising symptoms that cut across multiple psychiatric conditions as well as symptoms found only in specific mental and substance use disorders. This was perceived as important for ensuring the identification of primary care patients, who symptoms may otherwise go undetected. Panelists noted that since behavioral health providers do not often conduct assessments of functioning, the purpose of this quality measure is to increase providers' frequency of conducting assessments of functioning. Some panelists concurred that limiting the functioning measure to diagnoses alone could inappropriately skew toward conditions that are more commonly diagnosed and treated in primary care (e.g., depression).

- **Include all patients.** In the qualitative data, respondents suggested alignment of the denominator across measures: including all patients. Most respondents (71%; Table 7) on the quantitative survey endorsed using the same denominator as the Monitoring measure.
- **Diagnoses to include:** Most respondents (76%) on the quantitative survey indicated assessments should not be limited to select diagnoses. In discussion, it was noted that diagnoses should include only those for which a screening and monitoring tool were completed.

#### **Exclusions/Exceptions Specification**

- Recommendations included those agreed to in the Standardized Assessment Tool quality measure as well as an additional exclusion for: receipt of palliative care.
- In the qualitative data, respondents said no additional exclusions/exceptions necessary or that the exclusions could be the same across the MBC measures.

#### **Stratification and Risk Adjustment**

- Panelists noted risk adjustment is an attempt to account for individuals from different segments of the population, but it is difficult (statistically) for a provider to operationalize risk adjustment given current capabilities of EHR systems, data analytics, and other health information technology. Since functional impairment is an outcome measure, panelists stated the importance of considering social risk factors, as these may be linked to various domains of functioning (e.g., cognition, self-care).
- In the qualitative data, respondents suggested stratification variables could include: diagnosis, access to care, insurance type, employment status, food security, transportation, and housing. Normative modeling may be used to differentiate higher versus lower functioning groups. Regarding risk adjustment, respondents suggested including comorbid conditions, disability, and disorder severity. Other respondents stated they were not sure which variables to stratify/adjust on.)

#### **Other Discussion Points about this Measure**

- **Engaging providers in assessment:** Panelists commented that behavioral health providers do not often conduct assessments of functioning. Therefore, the intention of the quality measure is to engage providers in administering assessment tools that examine individuals' functioning. The TEP recommended assessment of functioning as it supports provider engagement with patients.
- **Assessing treatment impact on functional impairment changes.** The TEP emphasized the importance of demonstrating care was delivered, while acknowledging some individuals improve without receiving care. However, panelists cautioned that if the quality measures' objectives include advancing the utilization of evidence-based care, then measuring the outcome in the absence of care could prove counterproductive.
  - In the qualitative data, respondents stated change could be measured meaningfully as either maintenance or improvement in score. Some respondents suggested including t-scores and effect sizes.

Table 7. Results from the quantitative survey on Functional Impairment

<b>Should assessment of reduction in functional impairment use the same denominator as the monitoring measures?</b>	<b>N (%)</b>
1. No	4 (19.0)
2. Yes	15 (71.4)
3. Unsure	2 (9.5)
<b>Should the assessment be limited to specific diagnoses?</b>	<b>N (%)</b>
1. No	16 (76.2)
2. Yes	4 (19.0)
3. Unsure	1 (4.8)
<b>Between WHO-DAS and PROMIS, which should be considered?</b>	<b>N (%)</b>
1. WHO-DAS	5 (23.8)
2. PROMIS	16 (76.2)
<b>How should reduction or maintenance be defined?</b>	<b>N (%)</b>
1. Decline greater than X% of functioning scale	4 (19.0)
2. Decline greater than specific effect size on a functioning scale	10 (47.6)
3. Any change	7 (33.3)

## Recovery

**Measure Name:** Recovery for all patients seen for mental health and substance use care

**Brief Measure Description:** The percentage of individuals aged 18 years or older presenting with a mental and/or substance use disorder who demonstrated improvement or maintenance of recovery (as defined, prioritized, and/or reported by the individual) based on results from a standardized assessment tool.

Key discussion points and consensus:

### General Specifications

- **Guidance for providing input on the specifications.** The Substance Abuse and Mental Health Services Administration's (SAMHSA) definition of recovery will be used to guide TEP (and CFP) input on the measure's specifications. Panelists said it was important to think about improvement as well as maintenance of recovery. They further stated that a maintenance threshold should be specified and examined during testing.

### Numerator Specification

- **Assessment frequency.** As with the Functional Impairment measure, panelists recommended during discussion that assessments occur quarterly. This was reflected in the quantitative survey data, with a slight majority of respondents (62%; Table 8) endorsing quarterly assessments.
- **Assessment period.** In the qualitative data, most respondents indicated the assessment should be done after the initial assessment.

### Denominator Specification

- In the qualitative data, most respondents stated the denominator should include all patients.

### Exclusions/Exceptions Specification

- In the qualitative data, respondents suggested additional exclusions beyond those included in the Monitoring measure, such as: patient refusal, cognitive impairment, patients on hospice/palliative care, stable patients with chronic symptoms, and the same exclusions as applied to other measures.

### Stratification and Risk Adjustment

- Regarding factors for stratification/risk adjustment, survey respondents suggested including the following factors: age, gender, family support, adherence, severity, disability, complexity, non-compliance, socioeconomic status, diagnosis, and comorbidity.

Table 8. Results from the quantitative survey on Recovery.

<b>How should improvement or maintenance be defined? Options include anyone who has:</b>	<b>N (%)</b>
1. Decline greater than X% in recovery scale	4 (19.0)
2. Decline greater than specific effect size on a recovery scale	9(42.9)
3. Any change	8 (38.1)
<b>How often/ in what time frame should the recovery assessment tool be administered?</b>	<b>N (%)</b>
1. Every visit	0 (0.0)
2. Monthly	3 (14.3)
3. Quarterly	13 (61.9)
4. Every 6 months	4 (19.0)
5. Other	1 (4.8)

## Other Information Arising from TEP Discussions

*At various points throughout the day, panelists raised other important issues related to the measure development initiative. Here we summarize the key points arising from these discussions.*

### General Comments that Apply across Two or More Measures

- **Provider engagement.** Panelists discussed that greater engagement is proven to increase patient satisfaction, reduce the number of encounters needed to achieve a positive health outcome, and reduce costs. However, it should be tempered if the measure poses undue burden because the data collected by the providers are meaningless.
- **Defining objectivity and supporting individual-clinician communication.** A panelist commented that the specifications of the Monitoring and Treatment/Care-Plan Adjustment measures should not impact the bidirectional exchange between individuals and their providers. Some may demonstrate stability or subtle changes over the lifetime of their condition. However, disease courses vary among individuals. The broader panel agreed that this may be challenging for the measure developers when attempting to formulate an objective quantifiable outcome.
- **Consistency across settings.** Panelists agreed providers treating mental health disorders in primary care settings follow the initial assessment, monitoring, and treatment adjustment processes required of other mental health care providers. The alignment of standardized tools for the collection of these data should demonstrate consistency across all of those implementing the process measures regardless of care settings (i.e., specialty care, general psychiatry).

### Discussions on Selection of Assessment Tools

#### Standardized Assessment

- **The utility of cross-cutting measures.** Panelists agreed the Standardized Assessment tools (as all tools) are valuable for augmenting expert opinion/clinician judgment, not replacing it. Panelists noted comprehensive assessments that drill down to the specific, commonly seen mental health disorders and associated problems (e.g., depressive disorders, anxiety disorders, psychotic disorders, substance use disorders, and suicide) may have utility among providers in general. The Standardized Assessment subgroup suggested focusing on the specific disorders/problems listed above. However, in the full TEP discussion, panelists noted needing to embed broad-based assessment tools to capture symptoms that cut across diagnoses; cross-cutting measures can aid in the prevention of misdiagnosis or underdiagnosis.
- **Providing a list of tools from which providers can choose.** The TEP suggested including a list of approved, optional tools along with criteria for selecting which tool to use. This would enable providers to determine whether tools they currently utilize meet the selection criteria and may be used to satisfy the quality measure.
- **Including an assessment of recovery in the Initial Assessment.** The group discussed if an assessment of Recovery could be included as part of the Initial Assessment and decided it is important to differentiate between symptoms, functioning, and recovery. In the qualitative data, respondents stated recovery assessments might be inappropriate (e.g., too lengthy, of questionable validity) or

non-applicable (e.g., too early in treatment for recovery to be relevant) at the time of the initial assessment.

### Functional Impairment

- **WHODAS (a global assessment of functioning).** Panelists considered the benefits and risks associated with embedding the WHODAS assessment tool in the Functional Impairment quality measure. Because the WHODAS assesses global functioning, panelists argued it would be less sensitive to change among high-functioning individuals. Panelists also stated measuring discrete (or specific) domains of functioning is critical for tracking change in high-functioning individuals whose changes in level of functioning may be harder to detect with global assessments (e.g., WHODAS).
- **PROMIS (domain-specific assessments of functioning).** Panelists agreed the PROMIS is advantageous because it is already widely used in many health care systems. Panelists discussed the capacity for measure developers to include additional validated PROMIS metrics to uniquely specify quality measures for various types of outpatient settings. A panelist explained that when they treat individuals who appear to be doing well on several domains but are struggling in a single domain, it is beneficial for providers to be able to address particular domains. This allows the provider to work with the individual to apply new interventions or improve on an existing treatment that should elicit improvement in that domain.
- **Consensus:** While most panelists (76%; see Table 5) preferred the PROMIS, a discrete domain assessment, panelists also expressed willingness to start with the WHODAS as the functional impairment assessment tool. The TEP ultimately agreed the WHODAS is a reasonable tool and if during testing it is established that it is not sensitive enough, it can be removed and potentially replaced by another tool in the future. (APA's measure developers informed the panel that when the WHODAS was used during the DSM-5 field trials, its sensitivity to change was demonstrated across individuals and groups.)

### Recovery

- Some panelists stated the RAS-2 is the most “usable” tool for measuring recovery, particularly when assessing a single domain. However, other panelists disagreed, questioning whether the RAS-2 is feasible to implement given its complicated validation and scoring method. The subgroup affirmed their preference to use a single-score report rather than multiple scores parsed out by the recovery domains.
- Panelists listed the benefits of using the RAS-2 assessment tool. Benefits included: shortest instrument available among those previously considered; assesses the ‘whole’ patient; may be administered by the provider or self-administered; the individual domain scores can be used by clinicians to improve care;
- The TEP discussed using more than one Recovery assessment tool. However, some panelists cautioned against offering too many assessments tools as options, as this has the potential to complicate the implementation. Similar to TEP input in earlier discussions, provider judgement is preferred when deciding to apply the Recovery assessment tools.

## Summary of Discussion of Measure ‘Reduction or maintenance of symptoms for patients with opioid misuse’

**Measure Name:** Reduction or maintenance of symptoms for patients with opioid misuse

**Brief Measure Description:** The percentage of individuals aged 18 or older with opioid use disorder (OUD) or misuse who demonstrated a reduction in (or maintenance of) OUD severity/opioid misuse based on results from a standardized assessment tool.

*[Note: Due to the time constraints at the March 2019 in-person TEP meeting, neither this quality measure topic nor the associated assessment tools were discussed then. As such the APA and NCQA measure developers presented at the June 2019 TEP in-person TEP meeting the information originally developed for the March 2019 in-person TEP meeting. Panelists were not surveyed for their input on the feasibility of the assessment tools for this measure.]*

Key discussion points and consensus:

### **Differentiating Misuse and Abuse**

- Panelists inquired as to how the measure will differentiate between misuse and abuse. The measure development team clarified that the DSM-5 manual for OUD states that a patient must meet 2 of 11 symptoms in a single twelve-month period to meet OUD criteria. Once an individual meets the criteria, they may receive a severity assessment. It is at this point when a provider would monitor an individual with OUD. Because of the difficulty of distinguishing between disorder and misuse, clinicians are sometimes hesitant to diagnose a disorder because of the difficulty. The measure development team suggested focusing on capturing opioid use disorder, rather than misuse.

### **Denominator**

- The Tobacco Alcohol Prescription medications and other Substance (TAPS) 1 and 2 tools were presented as potential validated instruments for capturing opioid misuse or severity. Panelists communicated concerns with the quality measure’s denominator, particularly if TAPS 1 was used since the tool’s time frame refers to the past 12 months. The measure development team explained that the TAPS 2 may be completed every 3 months and is, therefore, a potentially viable assessment tool.
- To inform the denominator statement, the quality measure users will look for results that demonstrate the assessment tools’ scores have changed and signify that the individual no longer meets criteria for OUD. The measure developers explained that by specifying the tools, and using them in practice, the data can be examined for the tests of change.

### **Reduction or Abstinence**

- A panelist noted that whether this measure would assess abstinence as well as reduction in opioid use remains undecided
- Other panelist suggested that if the measure developers are considering broadening the quality measures to assess for additional substance use disorders, then the TEP and CFP should review tools that assesses multiple domains. A panelist commented that PROMIS has 7 “really good” assessment

tools, that if used, the providers and individuals might achieve reduction in symptoms and continued use.

#### **Alternatives to a PROM**

- Recommended alternatives to patient-reported measures included claims quality measure data collected under the CMS Merit-based Incentive Payment System (MIPS).
- It was also stated that while a urine drug screen is a standard of care for an individual with a substance use disorder, it can be humiliating for that individual. Panelists explained that they are interested in knowing how an individual is progressing between encounters. If urine testing is given at least every 2 weeks an outcomes assessment would be applied during the next 7 to 14 days.
- The TEP agreed that the proposed quality measure should assess beyond the symptoms. One panelist agreed that a composite score for function and symptoms would be beneficial.

#### **Next Steps**

- The information presented in this summary will guide the draft and refinement of the measure specifications for alpha and beta testing.
- The next TEP meeting is scheduled the first week of November 2019.