NATIONAL QUALITY FORUM

# Meeting Summary

## Behavioral Health and Substance Use Standing Committee—Measure Evaluation Web Meetings

The National Quality Forum (NQF) convened the Behavioral Health and Substance Use Standing Committee for two two-hour web meetings on June 19 and one two-hour meeting on June 26, 2019 to evaluate six measures and to discuss the current composition of the Behavioral Health and Substance Use portfolio. This document summarizes those proceedings.

## Welcome, Introductions, and Review of Meeting Objectives

NQF welcomed the Standing Committee and participants to the web meeting. NQF staff reviewed the meeting objectives. Committee members each introduced themselves and disclosed any conflicts of interest.

## Topic Area Introduction and Overview of Evaluation Process

NQF staff provided an overview of the topic area and the current NQF portfolio of endorsed measures. There are currently 46 measures in the Behavioral Health and Substance Use portfolio. Additionally, NQF reviewed the Consensus Development Process (CDP) and the measure evaluation criteria.

## Measure Evaluation

During the meeting, the Behavioral Health and Substance Use Standing Committee evaluated six measures for endorsement consideration. A summary of the Committee deliberations will be compiled and provided in the draft technical report. NQF will post the draft technical report on July 25, 2019 for public comment on the NQF website. The draft technical report will be posted for 30 calendar days.

**Measure Evaluation Criteria Rating Key:** H – High; M – Medium; L – Low; I – Insufficient

### 0560 HBIPS-5 Patients Discharged on Multiple Antipsychotic Medications (The Joint Commission)

#### *Measure Steward/Developer Representatives at the Meeting*

- Elvira Ryan, The Joint Commission
- Dave Morton, The Joint Commission
- Stephen Schmaltz, The Joint Commission

#### *Standing Committee Votes*

- Evidence: H-1; M-2; L-6; I-7

#### *Standing Committee Recommendation for Endorsement: Yes-0; No-0*

The Standing Committee did not vote on a recommendation for endorsement because the measure did not pass the evidence criterion—a must-pass criterion. Committee discussion about this measure revealed concern that the evidence presented by the developer was dated and too

general to support a measure that looks at antipsychotic use for indications outside of schizophrenia. There was also concern that the measure might discourage multiple antipsychotic use that may be useful. Two specific studies[a,b] were cited during the discussion as examples demonstrating that while monotherapy is generally regarded as the best first-line approach, there is evidence to indicate that polypharmacy may be superior in some instances. Accordingly, the Committee was concerned that the current measure might work against the best option for some patients. Several Committee members noted the growing controversy in the field about the use of multiple antipsychotics as a treatment option in certain cases beyond the justifications included in the specifications.

In response, the developer noted that the measure is based on current guidelines, though new APA guidelines are anticipated in September. The developer also argued that its measure neither supports or discourages polypharmacy, but only requires that one of three evidence-based justifications be proffered if it is used.

At least one Committee member voiced support for these justifications and the measure as it was presented. Committee members expressed concern that the evidence presented did not fully consider the possibility that the measure may discourage polypharmacy when it is indicated. Committee members also held to their concern that the evidence presented was specific to schizophrenia even as the measure specifications were not similarly restricted to the use of antipsychotics for that single indication.

### 0640 HBIPS-2 Hours of Physical Restraint Use (The Joint Commission)

#### *Measure Steward/Developer Representatives at the Meeting*

- Elvira Ryan, The Joint Commission
- Dave Morton, The Joint Commission
- Stephen Schmaltz, The Joint Commission

#### *Standing Committee Votes*

- Evidence: M-14; L-2; I-0
- Performance Gap: H-3; M-9; L-3; I-1
- Reliability: M-13; L-3; I-0
- Validity: H-2; M-12; L-2; I-0
- Feasibility: H-2; M-11; L-3; I-0
- Use: Pass-16; No Pass-0
- Usability: H-3; M-10; L-2; I-1

[a] Tiihonen J, Taipale H, Mehtälä J, et al. Association of antipsychotic polypharmacy vs monotherapy with psychiatric rehospitalization among adults with schizophrenia. *JAMA Psychiatry*. 2019;76(5):499-507.

[b] Katona L, Czobor P, Bitter I. Real-world effectiveness of antipsychotic monotherapy vs. polypharmacy in schizophrenia: to switch or to combine? A nationwide study in Hungary. *Schizophr Res*. *152*(1)246-254.

### *Standing Committee Recommendation for Endorsement: Yes-14; No-2*

The Standing Committee recommended the measure for continued endorsement. Discourse here was lengthy, presumably because the use of restraints is an especially dramatic intervention. Regarding evidence and performance gap, the Committee heard and discussed that there are some inconsistent findings: sometimes the elderly and minorities are characterized as high risk for restraint use, but other studies show young white males are restrained most. Moreover, time trends presented by the developer in its submission indicated that restraint use was increasing with time, which contradicts that the quality measurement is useful.

In response, the developer argued that those increases were not adjusted for case-mix, and thus these trends should not be interpreted as a failure of the measures to have impact.

Committee members discussed the extent to which the measure ultimately correlated with both quality and improved outcomes. Generally, the Committee expressed sentiments in favor of the measure as impactful in changing the way restraints are used, though at least one member noted that decreased seclusion rates might correlate with increased and potentially inappropriate use of chemical restraints (e.g., sedatives). Physical restraint use was relatively rare (<1%) per the developer's data, and one member of the Committee suggested that the measure may have "topped out," i.e., there may be no further room for improvement across most facilities. The developer and Committee discussion clarified that the data showed approximately 36,000 occurrences of restraint use in 2018, suggesting that the method is still prominent.

Regarding reliability, the developers checked charts against their study records and found 100 percent concordance based on a limited number of records (<200). When asked if they felt a more direct test-retest reliability experiment was necessary—that is with "two timers" at the facility—the developers responded that this was not possible, but that they did travel to facilities to re-review charts, and there were procedures in place at each facility to get precise start and stop times for restraint use. Discussion of validity testing noted the "moderate" correlation (r=0.216) between restraint use and seclusion use (measure 0641). One public comment was tendered by a graduate student who claimed to have empirical data supporting a correlation between quality reporting and reduced restraint use.

## 0641 HBIPS-3 Hours of Seclusion Use (The Joint Commission)

### *Measure Steward/Developer Representatives at the Meeting*

- Elvira Ryan, The Joint Commission
- Dave Morton, The Joint Commission
- Stephen Schmaltz, The Joint Commission

### *Standing Committee Votes*

- Evidence: M-12; L-4; I-0
- Performance Gap: H-2; M-12; L-2; I-0
- Reliability: M-13; L-3; I-0
- Validity: H-1; M-14; L-1; I-0
- Feasibility: H-1; M-12; L-3; I-0
- Use: Pass-16; No Pass-0

- Usability: H-2; M-12; L-1; I-1

### *Standing Committee Recommendation for Endorsement: Yes-14; No-2*

The Standing Committee recommended the measure for continued endorsement. Committee discussion regarding this measure was similar to that for the restraints measure (0640). One Committee member noted that use of seclusion is arguably distinct from restraint use because it may be requested by a patient who desires a self-imposed "time out," thus there should be a distinction between "locked" and "unlocked" seclusion. The discussion noted disparities data that supported a gap. Outlier years (regarding use rates) were evident, though the time trends presented were generally downward (suggesting increased improvement with time). Reliability was noted as limited to 191 chart re-abstractions (as was the case with the restraint measure). Validity was predicated upon the same empirical correlation described for restraints, and face validity was noted as well based on responses from 36 hospital-based representatives—31 of which reported the measure to very good or good with regards to its validity.

At least one Committee member expressed concern, but otherwise accepted, that validity assessments submitted with NQF measures rarely include randomized studies that establish a link between a measure and complex, long-term quality outcomes (e.g., qualify of life years after symptoms first emerged). Another Committee member received confirmation from the developer that part of its validity presentation involved a target analysis, which allows users to assess or benchmark their relative performance against other hospitals reporting to The Joint Commission. Some concern was expressed that the measure may have "topped out," as much change in restraint and seclusion practices has already occurred in recent years. One Committee member suggested seclusion, physical restraints, and chemical sedation should all be considered in relation to one another as practices that may generally reflect undesirable outcomes. This suggestion was clarified and summarized by NQF staff who noted that these three approaches may either be complements or substitutes for one another. During the related measure discussion, one Committee member was interested in how measures might capture de-escalation options, and it was noted that seclusions and restraint approaches are related processes.

## 1922 HBIPS-1 Admission Screening for Violence Risk, Substance Use, Psychological Trauma History and Patient Strengths Completed (The Joint Commission)

### *Measure Steward/Developer Representatives at the Meeting*

- Elvira Ryan, The Joint Commission
- Dave Morton, The Joint Commission
- Stephen Schmaltz, The Joint Commission

### *Standing Committee Votes*

- Evidence: H-0; M-11; L-5; I-0
- Performance Gap: H-0; M-9; L-7; I-0
- Reliability: M-13; L-3; I-0
- Validity: H-0; M-11; L-4; I-1
- Feasibility: H-2; M-11; L-3; I-0
- Use: Pass-14; No Pass-2
- Usability: H-0; M-12; L-4; I-0

### *Standing Committee Recommendation for Endorsement: Yes-0; No-0*

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

Committee members expressed some concerns with the evidence, performance gap, and validity testing associated with the measure. The evidence presented by the developer generally achieved moderate and low ratings. One Committee member noted the difficulty of achieving a high grade for this sort of measure because randomized trials to assess the effect of such screenings (for violence, substance use, psychological trauma, and patient strengths on actual outcomes) are challenging to implement. At least two Committee members also expressed concern that the screenings and the evidence were not linked to treatment plans and actual outcomes in some more substantive way in the submission (other than with a logic model).

Committee members asked the developer to address the concern that the three-day period allowed for such screening was too long, especially to ascertain substance use or violence threats for a newly admitted patient. In response to the three-day window question, the developers said that this period of review was explicitly noted as a maximum time rather than an optimal time. Committee members expressed concern that the measure may have "topped out" as performance rates have increased from 87 percent in 2009 to 92.7 percent in 2018, and no disparities were presented. In 2018 the 20th percentile was 92 percent, and the 70th percentile was 99 percent. In response to the concern about whether an addressable gap remains, the developers said they were aware of inter-facility differences and noted that rates are not the same for free standing facilities versus an acute hospital unit or military/government facilities.

Validity testing presented by the developer included the absence of significant correlations between this measure and measures of hours of restraint and seclusion use (meaures 0640 and 0641 noted above), and a small, significant correlation between this measure and appropriate justification of multiple antipsychotic use (r=0.14, p=0.0002; measure 0560). This validity presentation, however, was regarded by some Committee members as reflecting only that some screening was occurring rather than that more specific risk screening was happening. Some Committee members felt the comparative standard used was a poor indicator of true risk screening as conveyed by this measure.

The Committee will re-vote on the measure on the post-comment web meeting on September 16, 2019.

## 3488 Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence (FUA) (National Committee for Quality Assurance)

### *Measure Steward/Developer Representatives at the Meeting*

- Junqing Liu, NCQA

### *Standing Committee Votes*

- Evidence: H-4; M-10; L-0; I-0
- Performance Gap: H-8; M-6; L-0; I-0
- Reliability: H-2; M-12; L-0; I-0
- Validity: H-2; M-12; L-0; I-0

- Feasibility: H-6; M-7; L-1; I-0
- Use: Pass-13; No Pass-1
- Usability: H-3; M-10; L-1; I-0

### *Standing Committee Recommendation for Endorsement: Yes-13; No-1*

The Standing Committee recommended the measure for continued endorsement. This measure is a maintenance measure because it is a "spin-off" from measure 2605, which had previously combined mental health and substance use disorder (SUD) emergency department (ED) follow-up visits into a single measure. Discussion surrounding this measure was generally favorable. The developer noted in its introduction that absence of follow-up care from a SUD ED event increased one's risk of readmission by six times.

Committee comments about evidence suggested reasonable quality, quantity, and consistency. The performance gap was substantial as data across ages and Medicare and Medicaid programs indicated that performance on this measure was below 20 percent of the ED discharges that are followed-up; moreover, those with the added risk of co-occurring psychiatric illness had even lower rates of follow-up. At least one member of the Committee was pleased that telehealth visits were included as qualifying numerator events. Committee members briefly raised the concern that persons with multiple ED visits might only be counted once in the denominator of this measure, but that concern was quickly set aside as minor. Committee members expressed support for the inclusion of SUD follow-up services delivered by primary care providers. Concerns that a simple survey of a patient, even an online survey, might be inappropriately counted as a follow-up event were assuaged by the developer who noted that only two-way communications (e.g., survey plus a response from provider) were counted as such, even as that two-way conversation might be asynchronous (e.g., via email). At least one Committee member asked the developer why it did not consider secondary diagnosis of SUD for the denominator. The developer responded that it was best to focus on the more certain SUD cases at this time.

Validation for this measure was demonstrated by comparing this measure to the mental illness follow-up measure and by comparing the 7- and 30-day follow-up measures to each other. The Committee found these validation approaches reasonable. Some concern was discussed that the current measure was competing with measure 0004 *Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment*, but the developer assuaged that concern by noting the current measure is uniquely specific to ED venues and immediate follow-up. Without requesting explanation from the developer, one Committee member expressed moderate concern that this measure does not include follow-up treatments outside of the traditional medical sphere (e.g., 12-step programs).

During the related measure discussion, one Committee member noted that this measure, which focuses on follow-up, may jumpstart initiation and engagement in care, the focus of another related measure (0004), for example. Note that 0004 differs from the current measure because it is not limited to detection of an SUD in an ED, and it also operationalizes "engagement" beyond a single follow-up.

### 3489 Follow-Up After Emergency Department Visit for Mental Illness (National Committee for Quality Assurance)

#### *Measure Steward/Developer Representatives at the Meeting*

- Junqing Liu, NCQA

#### *Standing Committee Votes*

- Evidence: H-4; M-10; L-0; I-0
- Performance Gap: H-7; M-7; L-0; I-0
- Reliability: H-2; M-11; L-1; I-0
- Validity: H-2; M-12; L-0; I-0
- Feasibility: H-5; M-8; L-1; I-0
- Use: Pass-13; No Pass-1
- Usability: H-3; M-10; L-1; I-0

#### *Standing Committee Recommendation for Endorsement: Yes-13; No-1*

The Standing Committee recommended the measure for continued endorsement. This measure was noted as similar to measure 3488. The Committee expressed some concern that the measure is based on a very general definition of follow-up, but the gap evident for delivery of even this 'low bar' service was said to be important to improve as fewer than 60 percent of the entities studied completed such follow-up activities with 30 days of an emergency department (ED) event. Telehealth was included in this measure as a numerator qualifying event. Signal-to-noise testing supported the reproducibility of the measure to differentiate between tested entities. Though the Committee articulated some concern that the two years of data presented did not demonstrate improvement trends, the developer said that such variability is not uncommon for newly deployed measures. The Committee expressed concern that follow-up was dependent on resources being available to deliver such care, though such availability is not directly assessed by the measure. In response, one Committee member noted that like a detoxification follow-up measure endorsed during the last NQF cycle and implemented in Delaware, this measure has the potential to spur the development of resources that facilitate bona fide follow-up events.

Much discussion surrounded the responsibility of the ED for this measure, discussion which seemed to converge on the expectation that EDs, per se, have limited responsibility and health plans are instead held responsible for performance on this measure. During the related measure discussion, one member reiterated that EDs generally don't have patients for that long. The Committee member shared that some facilities lack the resources to refer patients to adequate follow-up care, and patients may also lack access to follow-up care.

## Role of Scientific Methods Panel

At the request of the Committee, the NQF staff reviewed the role of NQF's Scientific Methods Panel (SMP) and the reasons that measure 3492 *Emergency Department Use Due to Opioid Overdose* did not pass the SMP review during the spring 2019 cycle. The principal reasons that measure 3492 did not pass were that the measure was not tested using ICD-10 codes and county-level validity testing was not provided. The SMP also had concerns that the measure conflated opioid overdose events with opioid-related events, the latter presumed to be less severe.

The Committee asked NQF why this measure is considered an outcome measure. Staff replied that is was considered as such because ED events are generally deemed avoidable negative events, rather than more standardized processes/pathways of care. During this discussion NQF staff also noted that the SMP mission is to focus on quantitative/statistical issues, and then to defer to the relevant standing committee as subject matter experts on specific clinical and content issues.

## Portfolio Gaps Discussion

NQF staff provided an overview of the portfolio gaps identified by the Committee in 2016-2017 and also most recently in fall 2018. Points described during that review included the Committee's previous expressed interest in seeing more measures that address the following issues: opioid use disorder, criminal justice issues, patient-reported outcomes, tailored treatments that specifically consider stage of illness and readiness for change, recovery measures that address long-term outcomes, social determinants (e.g., housing, employment), care coordination issues, cost of care issues, and functional outcomes. During the closing discussion, one Committee member emphasized the following issues: SMI should be considered as a special population or a disparities population; more patient-reported "experience" measures should be developed (in addition to patient reported outcomes); and more consideration should be given to considering if measures crystalize into standards of care (i.e., whether they become, as intended, broadly utilized).

Committee members also discussed the addition of burn-out to ICD-11, and staff presented additional information about behavioral health additions to that international reporting standard. One Committee member involved in the ICD-11 discussions noted that the new standard is about five years from full use but aims in part to facilitate more automated record keeping and analytics, and further includes some important innovations that permit linking of patient safety information to other medical records. NQF staff noted that ICD-11 advanced from ICD-10 by appending several novel brain-based illness changes such as creating new diagnostic categories for hording and gambling, making stroke a neurologic rather a circulatory disease, and permitting the use of the ADHD diagnosis across ages.

The gaps discussion included a question regarding the link between the priority list of gaps developed by the Standing Committee and NQF incubator activities. NQF staff explained that the key gaps identified here influence other activities at NQF including those focused-on measurement specifically, and quality improvement, more generally.

## Public Comment

A comment was received during the evaluation meeting from The Joint Commission, regarding its measure 0560. The developer expressed disappointment that the Committee did not pass the measure on the evidence criterion. The developer pointed out that one of the studies brought up by Committee members during the discussion was published in 2019 after submission materials were due and also noted that the developer intends to incorporate any updates from the updated guidelines, but the guidelines had not yet been updated. One Committee member, in response to this comment, encouraged the developer to appreciate that Committee members have a

responsibility to consider all relevant measure information they have access too, even if it emerges after the application is submitted.

Another public comment articulated support for measures 0640 and 0641 (restraints and seclusion) based on the commenter's professional observation that such methods have historically been exceptionally negative experiences for patients and their families, and they continue to be regarded as such.

## Next Steps

NQF will post the draft technical report on July 25, 2019 for public comment for 30 calendar days. The continuous public comment with member support will close on August 23, 2019. NQF will re-convene the Standing Committee for the post-comment web meeting on September 16, 2019.