U.S. Department of Health and Human Services
Assistant Secretary for Planning and Evaluation
Office of Disability, Aging and Long-Term Care Policy

Development and Testing of Behavioral Health Quality Measures for Health Plans:

Final Report

March 2015

Office of the Assistant Secretary for Planning and Evaluation

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) is the principal advisor to the Secretary of the Department of Health and Human Services (HHS) on policy development issues, and is responsible for major activities in the areas of legislative and budget development, strategic planning, policy research and evaluation, and economic analysis.

ASPE develops or reviews issues from the viewpoint of the Secretary, providing a perspective that is broader in scope than the specific focus of the various operating agencies. ASPE also works closely with the HHS operating agencies. It assists these agencies in developing policies, and planning policy research, evaluation and data collection within broad HHS and administration initiatives. ASPE often serves a coordinating role for crosscutting policy and administrative activities.

ASPE plans and conducts evaluations and research--both in-house and through support of projects by external researchers--of current and proposed programs and topics of particular interest to the Secretary, the Administration and the Congress.

Office of Disability, Aging and Long-Term Care Policy

The Office of Disability, Aging and Long-Term Care Policy (DALTCP), within ASPE, is responsible for the development, coordination, analysis, research and evaluation of HHS policies and programs which support the independence, health and long-term care of persons with disabilities--children, working aging adults, and older persons. DALTCP is also responsible for policy coordination and research to promote the economic and social well-being of the elderly.

In particular, DALTCP addresses policies concerning: nursing home and community-based services, informal caregiving, the integration of acute and long-term care, Medicare post-acute services and home care, managed care for people with disabilities, long-term rehabilitation services, children's disability, and linkages between employment and health policies. These activities are carried out through policy planning, policy and program analysis, regulatory reviews, formulation of legislative proposals, policy research, evaluation and data planning.

This report was prepared under contract #HHSP2332010016WI between HHS's ASPE/DALTCP and Mathematica Policy Research. Additional funding provided by the HHS Centers for Medicare and Medicaid Services. For additional information about this subject, you can visit the DALTCP home page at https://aspe.hhs.gov/office-disability-aging-and-long-term-care-policy-daltcp or contact the ASPE Project Officer, D.E.B. Potter, at HHS/ASPE/DALTCP, Room 424E, H.H. Humphrey Building, 200 Independence Avenue, S.W., Washington, D.C. 20201. Her e-mail address is: D.E.B.Potter@hhs.gov.

DEVELOPMENT AND TESTING OF BEHAVIORAL HEALTH QUALITY MEASURES FOR HEALTH PLANS: Final Report

Jonathan Brown
Sarah Hudson Scholle
Sarah Croake
Miriam Rosenau
Junquing Liu
Nathan Darter
Kelsey Farson Gray
Menolly Hart
Bonnie Harvey
Rita Lewis
Dean Miller
Suzanne Morton
Milesh Patel
Christine Ranshous

Mathematica Policy Research

March 18, 2015

Prepared for
Office of Disability, Aging and Long-Term Care Policy
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
Contract #HHSP2332010016WI

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGMENTS

Mathematica Policy Research and the National Committee for Quality Assurance (NCQA) prepared this report under contract to the Office of the Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services (HHS) (HHSP23320100019WI/ HHSP23337001T). Funding support was provided by the HHS Substance Abuse and Mental Health Services Administration (SAMHSA). The authors appreciate the guidance of Kirsten Beronio, Joel Dubenitz, Richard Frank, D.E.B. Potter (ASPE), and Lisa Patton (SAMHSA). Jeremy Biggs, Jung Kim, Sean Kirk, and Jessica Nysenbaum (Mathematica) contributed to the data collection and analysis. Melissa Azur (Mathematica) and Mary Barton (NCQA) provided feedback on this report and guidance throughout the project.

The views and opinions expressed here are those of the authors and do not necessarily reflect the views, opinions, or policies of ASPE, SAMHSA, HHS, or the technical expert panel. The authors are solely responsible for any errors.

ABSTRACT

Summary: Many people with behavioral health disorders suffer comparatively poorer health outcomes, including premature death. Quality measures targeting this population utilized by states, health plans, providers and other stakeholders may improve the quality of their care. In this project, we developed and tested measures reported by health plans that focus on screening and monitoring of care for co-morbid conditions among people with serious mental illness (SMI) and/or alcohol or other drug dependency (AOD). For the SMI population, these measures focused on assessing comprehensive diabetes care; controlling high blood pressure; and screening for body mass index (BMI), high blood pressure, tobacco use, and unhealthy alcohol use. For the AOD population, the measures focused on screening for high blood pressure, depression, and tobacco use. We also developed a measure for health plan reporting to assess the extent to which people discharged from the emergency department for mental disorders or AOD receive timely follow-up care. In March 2015, the National Quality Forum (NQF) endorsed 11 measures from this project.

Major Findings: Measures that assessed diabetes care, high blood pressure control, BMI screening, and tobacco screening among the SMI population, as well as tobacco screening among the AOD population, demonstrated strong reliability and meaningful variation across health plans, suggesting they are suitable to differentiate the quality of care. The alcohol screening measure for the SMI population showed less variation across health plans but received support from stakeholders. The blood pressure and depression screening measures performed poorly and stakeholder support was divided. The follow-up after emergency department measure showed wide variation across state Medicaid programs and received strong stakeholder support. We identified several challenges for developing and using measures focused on behavioral health populations, including a lack of evidence to support some measure concepts and difficulty accessing data to calculate measures. Multistakeholder engagement throughout the project was critical to developing meaningful measures.

Purpose: We focused on developing measures for health plan reporting that address: (1) co-morbid conditions among SMI and AOD populations; and (2) follow-up care after discharge from the emergency department for a mental disorder or AOD. We tested the measures using quantitative and qualitative methods to assess attributes consistent with NQF endorsement criteria: importance, feasibility, usability, and scientific acceptability.

Methods: We reviewed existing measures and gathered input from consumers, providers, health plans, state agencies, and performance measurement experts to identify opportunities for new measures. After reviewing the evidence to support measure concepts, we specified and tested measures that addressed priority conditions and populations. We tested the follow-up after emergency department measure using

Medicaid claims data. All other measures were piloted at three diverse health plans. Quantitative testing of all measures involved calculating performance rates to examine variation across health plans or states, along with differences in performance between subpopulations. We examined the reliability of the measures using various psychometric tests. Finally, we solicited public comments and held focus groups with a range of stakeholders to get input on the measure specifications and to understand whether the measures yield findings that can be used to inform quality improvement efforts. We also sought their perspectives on practical barriers to implementing the measures. A technical expert panel provided guidance throughout the project. After the testing, we refined the measure specifications and submitted 11 measures to NQF for endorsement.

ACRONYMS

The following acronyms are mentioned in this report and/or appendices.

ACA	Affordable Care Act
ACO	Accountable Care Organization
AHRQ	HHS Agency for Healthcare Research and Quality
AMA-PCPI	American Medical Association Physician Consortium for Performance Improvement
AOD	Alcohol or Other Drug Dependence
AOD ED	Alcohol or Other Drug Dependence Emergency Department
ASPE	HHS Office of the Assistant Secretary for Planning and Evaluation
BH	Behavioral Health Record
BMI	Body Mass Index
BP	Blood Pressure
CDC	HHS Centers for Disease Control and Prevention
CHIPRA	Children's Health Insurance Program Reauthorization Act
CMS	HHS Centers for Medicare & Medicaid Services
CPT	Current Procedural Terminology
CQM	Clinical Quality Measures
D-SNP	Dual Special Needs Plan
EHR	Electronic Health Record
FFS	Fee-For-Service
FU ED	Follow-up Emergency Department
G-code	G Programming Language
HbA1c	Glycated Hemoglobin
HEDIS	Healthcare Effectiveness Data and Information Set
HHS	U.S. Department of Health and Human Services
HITECH	Health Information Technology for Economic and Clinical Health Act
HMO	Health Maintenance Organization
HRSA	HHS Health Resources and Services Administration

IET	Initiation and Engagement of Alcohol and Other Drug Dependence Treatment
IOM	Institute of Medicine
IPFQR	Inpatient Psychiatric Facility Quality Reporting
IQR	Interquartile Range
LDL	Low-Density Liporotein
MAX	Medicaid Analytic eXtract
MBHO	Managed Behavior Health Organization
MH ED	Mental Health Emergency Department
MH/SA	Mental Health and Substance Abuse
MU	Meaningful Use
NCQA	National Committee for Quality Assurance
NQF	National Quality Forum
ONC	HHS Office of the National Coordinator for Health Information Technology
PQRS	Physician Quality Reporting System
SAMHSA	Substance Abuse and Mental Health Services Administration
SD	Standard Deviation
SMI	Serious Mental Illness
TEP	Technical Expert Panel
USPSTF	U.S. Preventive Services Task Force
VA	U.S. Department of Veterans Affairs

EXECUTIVE SUMMARY

Given the prevalence of mental health and substance use disorders, and their toll on the health care system, national advisory groups have noted the dearth of behavioral health quality measures ready for implementation (AHRQ 2010). A recent National Quality Forum (NQF) committee identified several gaps in behavioral health quality measures that can be used to hold state agencies, health plans, providers, and other entities accountable for care. Specifically, the committee noted the need for measures that focus on transitions in care and that address co-morbid physical health conditions among individuals with serious behavioral health conditions (NQF 2012).

With the establishment of the National Behavioral Health Quality Framework, the U.S. Department of Health and Human Services (HHS), Substance Abuse and Mental Health Services Administration (SAMHSA) has articulated priorities for improving the quality of behavioral health care consistent with the National Strategy for Quality Improvement in Health Care. The framework defines goals for all aspects of care including preventing behavioral health problems, implementing and improving treatment, and promoting and supporting recovery. It also targets populations from young children to elderly and includes specialty behavioral health treatment settings as well as broader health care provider and community-based efforts. Within this framework, new measures are needed to monitor the quality of care and inform quality improvement efforts.

Purpose of Project

In September 2011, the HHS Office of the Assistant Secretary for Planning and Evaluation, with support from SAMHSA, contracted with Mathematica Policy Research and the National Committee for Quality Assurance to develop behavioral health quality measures. This three-year project began by reviewing existing measures and gathering input from consumers, providers, health plans, state agencies, and performance measurement experts to identify opportunities for new measures. We then specified and tested the measures listed in Table ES.1. Twelve of the measures focus on screening or monitoring of co-morbid conditions that are highly prevalent among individuals with serious mental illness (SMI) and/or alcohol or other drug dependency (AOD). These conditions include diabetes, hypertension, and alcohol use for the SMI population, and depression and tobacco use for the AOD population. In addition, we developed a measure to assess whether individuals who are discharged from the emergency department for mental health disorders or AOD receive timely follow-up care in the community. These measures were specified for health plan reporting, as such plans have an opportunity to ensure that individuals are connected with community providers and receive preventative screening and monitoring of chronic conditions.

TABLE ES.1. Measures Tested, Performance, and Submission to NQF			
Measure	**Variation in Measure Performance Across Health Plans or States (% of patients who met measure requirement)[1]**	**Reliability[2]**	**Received NQF Endorsement**
BMI Screening and Follow-up for People with SMI	11.6 - 55.0	0.84	✓
Alcohol Screening and Follow-up for People with SMI	1.5 - 58.4	0.79	✓
High Blood Pressure Screening and Follow-up for People with SMI or AOD	12.8 - 38.0 for SMI population 8.2 - 12.1 for AOD population	0.86	
Tobacco Use Screening and Follow-Up for People with SMI or AOD	9.8 - 64.1 for SMI population 8.8 - 30.4 for AOD population	0.74	✓
Clinical Depression Screening and Follow-up for People with AOD	1.7 - 20.6	0.77	
Comprehensive Diabetes Care for People with SMI[3]			
HbA1c Testing	15.7 - 65.4	0.65	✓
HbA1c Control (<8.0%)	6.0 - 48.8	0.51	✓
HBA1c Poor Control (>9.0%)	44.9 - 92.8	0.49	✓
Eye Exam	1.2 - 27.5	0.74	✓
Medical Attention for Nephropathy	6.0 - 61.4	0.76	✓
Blood Pressure Control	12.0 - 61.4	0.75	✓
Controlling High Blood Pressure for People with SMI	12.5 - 60.3	0.88	✓
Follow-Up After Emergency Department Use for Mental Health Conditions or AOD[4]	53.8 - 92.4 for mental health follow-up within 30 days 30.8 - 91.5 for AOD follow-up within 30 days	0.98	✓

NOTES:

1. Expressed as the proportion of patients who met the measure requirement. For the follow-up after emergency department measure, the table presents the variation across states. For all other measures, the table presents the variation across the 3 health plans that participated in measure testing. All the screening measures required that patients receive screening and, if positive, follow-up care.
2. Reliability for the follow-up after emergency department measure was calculated using beta-binomial statistic (score of 0.7 or higher indicates that the measure can reliably discriminate performance between states). Reliability for all other measures is the agreement between 2 chart abstractors (inter-rater agreement) for the numerator of the measure, calculated using Cohen's kappa statistic. Kappa scores of 0.61-0.80 indicate substantial agreement, suggesting that 2 abstractors independently had the same interpretation of the measure specification. All the measures demonstrated good reliability.
3. Although we refer to this conceptually as Comprehensive Diabetes Care, it includes 6 separate indicators/measures that were individually tested and submitted to NQF. This is not a composite measure.
4. The follow-up after emergency department measure has 4 rates: 7-day and 30-day follow-up for MH ED visits and 7-day and 30-day follow-up for AOD ED visits. We report the 30-day rates in this table for simplicity; there was also wide variation in the 7-day follow-up rates.

To align reporting for the SMI and AOD population with the general population, the measure specifications developed in this project were based on existing measures that health plans report as part of Healthcare Effectiveness Data and Information Set

(HEDIS®) or that providers report through the Physician Quality Reporting System (PQRS). Throughout the project, we sought input from a technical expert panel (TEP) and the PQRS measure developers and stewards to ensure that our specifications adhered to the original intent of the measure and to gather their feedback on our testing results.

Our testing of the measures was designed to gather information about their importance, feasibility, usability, and scientific acceptability, in accordance with NQF endorsement standards. We tested the follow-up after emergency department measure using Medicaid claims data. All the other measures (which use both administrative/ claims data and data abstracted from patient records) were piloted at three geographically diverse health plans: two Medicaid health plans and one Dual Special Needs Plan for individuals enrolled in both Medicaid and Medicare. Our quantitative testing involved calculating measure performance rates to examine variation across health plans or states, and differences in performance among subpopulations. We also examined the reliability of the measures using different psychometric tests depending on the data source (inter-rater agreement for measures that used data from patient records and beta-binomial testing for the follow-up after emergency department measure). Finally, we solicited public comment and conducted focus groups with a range of stakeholders to get input on the measure specifications and understand whether the measures yield findings that can be used to inform quality improvement efforts. We also sought their perspectives on practical barriers to implementing the measures. At the conclusion of the testing, we refined the measure specifications and submitted 11 measures to NQF in July 2014 (Table ES.1). After NQF review, all 11 measures were endorsed on March 6, 2015.

Measure Testing Results

Based on our testing, the measures with the strongest results and stakeholder support for the SMI population included those focused on comprehensive diabetes care, controlling high blood pressure, body mass index (BMI) screening, and tobacco screening. The tobacco screening measure also had strong performance and stakeholder support when applied to the AOD population. As summarized in Table ES.1, all these measures demonstrated strong reliability and meaningful variation across health plans, suggesting that they are suitable to differentiate the quality of care. For example, the proportion of individuals with SMI who met the requirement of the BMI measure (that is, they received BMI screening and follow-up care, if obese) ranged from 11.6 percent to 55.0 percent across health plans. There was a similar pattern for the other measures. In addition, the health plans, TEP, and other stakeholders reported that scores on these measures accurately reflected their expectations given the challenges associated with delivering care to these populations. When compared with either the overall 2012 Medicaid HEDIS rates or the rates of similar provider-level measures reported through PQRS, all these measures demonstrated much lower average rates in our testing -- suggesting disparities in care for the SMI and/or AOD population relative to the general population.

The alcohol screening measure demonstrated variation across health plans, but received less support from stakeholders and the TEP because an unusually low proportion of individuals with SMI were identified as unhealthy alcohol users. Nonetheless, they also perceived that this measure was important for health plans given the prevalence of alcohol use among the SMI population. Our analysis concluded that the measure had value for health plans and was suitable for submission to NQF.

Performance of the blood pressure screening measure was not as strong as the other measures. Health plans found that the measure specification (based on the PQRS measure) was overly complicated to implement. The TEP and other stakeholders echoed such concerns and perceived that screening for new cases of hypertension was less of a clinical and measurement priority than blood pressure control. There was little variation in the performance of the blood pressure screening measure for individuals with AOD, and stakeholders were not supportive of the measure for several reasons, including the lack of strong evidence to suggest that individuals with AOD are at greater risk for hypertension. Based on our analysis of the quantitative results and stakeholder feedback, we did not submit this measure to NQF.

Although there is evidence that depression is highly prevalent among people with AOD, and the TEP and stakeholders were generally supportive of the need for depression screening among this population, the depression screening measure did not yield information useful to health plans. Because the measure is intended to identify new cases of depression, individuals with a diagnosis of depression within the past year or who are already receiving depression treatment are excluded from the denominator of the measure. In our testing, nearly all individuals with depression had already been identified in the past year, and therefore, the measure resulted in a very low rate of identification and had limited value to health plans. Our analysis, based on only three health plans, suggested that a measure to monitor the quality of depression treatment among people with AOD may have more value for health plans than a measure designed to identify new cases of depression. Thus, we did not submit this measure for NQF endorsement.

Finally, when our follow-up after emergency department measure was tested using Medicaid claims data, it adequately distinguished performance between states and demonstrated very strong reliability. The proportion of individuals who received follow-up care after mental health and AOD emergency department visits varied widely across states. In addition, this measure received strong support from the TEP and stakeholders. Our analysis suggested that this is a useful measure to monitor follow-up care and therefore was submitted to NQF.

Other Lessons

This project identified several challenges and opportunities for developing and implementing quality measures focused on individuals with behavioral health conditions that may be useful for future efforts.

Multistakeholder engagement is critical to ensure that measures are meaningful and have the best chance for implementation. Our focus groups with consumers, providers, health plans, state officials, and performance measurement experts early in the project were critical to identify gaps in measurement, understand what entities could realistically be held accountable for performance on the measures, and identify data sources for measures. These stakeholders also provided valuable feedback to refine the measure specifications at several points in the project. They often have different perspectives, and finding common ground on quality measurement priorities can be difficult. In this project these stakeholders shared the concern that individuals with SMI and AOD have many co-morbid conditions that require better screening and monitoring, and that better monitoring of care transitions is needed. But they also proposed more controversial measurement concepts, including shared decision making, inappropriate use of psychotropic medications, monitoring of medication side effects, re-admissions, and others. For many of these concepts, there was no clear path forward to develop measures due to insufficient evidence or challenges identifying an entity accountable for the measure performance. Nonetheless, these are important concepts to consider for future work and it will be important to gain the input of all stakeholders to ensure that the final measures yield meaningful and actionable information.

Fragmentation of physical health and behavioral health coverage and services leads to fragmentation in accountability, creating obstacles for positioning and calculating measures. During the early stages of this project, for each measure concept that was proposed, we investigated the feasibility of existing data sources to calculate the measure and where the measure could be best positioned (providers, health plans, states, and such) to have the greatest impact on the quality of care. One of the major challenges we encountered is that no single entity is accountable for the quality of care for individuals with behavioral health conditions. Specialty mental health and substance abuse services are often carved out from general medical care or provided through special grant-funded systems of care that are not well connected with physical health plans, Medicaid, or other state agencies. This creates obstacles to accessing data across entities to calculate measures, and makes it difficult for these entities to act on the results of measures for which they perceive they have little influence. Many health plans initially volunteered to test our measures (indicating their interest in the health needs of individuals with SMI and AOD) but could not accurately calculate the measures because they did not have access to the full record of service utilization for their patients -- including both physical and behavioral health records and claims -- due to behavioral health carve-out arrangements or other limitations on data sharing. Stronger collaboration between the various entities responsible for providing the full array of services to the behavioral health population is necessary to facilitate the

widespread implementation of quality measures, and to promote shared accountability for performance on such measures.

Measures of psychosocial care would provide a more comprehensive understanding of the quality of care. Many stakeholders were concerned about the lack of NQF-endorsed measures focused on psychosocial care to complement existing measures that assess medication use and adherence. There was a particular concern among stakeholders that measures are needed to monitor the accessibility and outcomes of evidence-based psychosocial care, including various psychotherapies and other community-based mental health and social services. As we considered developing measures focused on psychosocial care, we discovered the lack of a data collection and reporting infrastructure to support such measures. As part of this project, we summarized the challenges involved in developing and implementing such measures, and proposed several avenues for future measure-development -- with an emphasis on advancing the measurement of outcomes (Brown et al. 2014). Further work is needed to move psychosocial measures forward.

Interpretation of data confidentiality hinders implementation of quality measures for behavioral health populations. During our testing, we found that even health plans that have responsibility for comprehensive physical health and behavioral health benefits have trouble accessing records for their patients with behavioral health conditions, particularly records for individuals with AOD. Some health plans interpret federal and state privacy laws as preventing them from accessing behavioral health records, and overcoming the legal hurdles to access such data is very burdensome and time consuming. In addition, the health plans that piloted our measures found that many behavioral health providers are unaccustomed to providing records for quality improvement purposes, and may not respond to such requests out of fear of violating privacy rules. Greater clarity of the privacy laws is needed to give health plans and providers confidence in their ability to share data for quality improvement purposes while protecting the rights and privacy of consumers.

Although the measures tested in this project fill critical gaps, more measures are needed to implement on a national scale to fully understand the quality of care provided to individuals with behavioral health conditions. Such measures must align with other federal and state initiatives (such as the electronic health record incentive program and Medicaid quality reporting) and take advantage of existing data sources and the evolving infrastructure for measurement.