



**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
Junqing 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 #HHSP2332010016W1

The opinions and views expressed in this report are those of the authors. They do not necessarily reflect the views of the Department of Health and Human Services, the contractor or any other funding organization.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	vi
ABSTRACT	vii
ACRONYMS	ix
EXECUTIVE SUMMARY	xi
I. PROJECT RATIONALE	1
A. Project Purpose	2
B. Report Roadmap	3
II. SELECTION OF MEASURE CONCEPTS	4
A. Scan of Measures	4
B. Focus Groups	6
C. Evidence Review	7
D. Technical Expert Panel Meeting	8
E. Final Measure Concepts	8
III. SPECIFICATION OF MEASURES	10
A. Specification of Screening and Monitoring Measures	10
B. Specification of Follow-up after Emergency Department Measure	13
IV. APPROACH TO MEASURE TESTING	15
A. Testing Questions	15
B. Quantitative Testing of Screening and Monitoring Measures	16
C. Characteristics of Health Plans that Participated in Testing	17
D. Health Plan Data Collection	19
E. Approach to Quantitative Testing of Follow-up after Emergency Department Measure	21
F. Approach to Gathering Stakeholder Feedback for All Measures	23
G. Data Security	25
V. TESTING RESULTS FOR SCREENING AND MONITORING MEASURES	26
A. Characteristics of Denominator Population Selected for Testing of Screening and Monitoring Measures	26
B. Number of Patients Included in Denominator for Screening and Monitoring Measures	26
C. Service Utilization among Denominator Populations for Screening and Monitoring Measures	28
D. Screening and Monitoring Measure Testing Results	30
E. Variation in Screening Measure Rate by Medical and Behavioral Health Data Sources	44

F. Inter-rater Agreement for Screening and Monitoring Measures.....	45
G. Conclusion from Testing of Screening and Monitoring Measures: Revisions to Measure Specification and National Quality Forum Submission.....	47
VI. TESTING RESULTS FOR FOLLOW-UP AFTER EMERGENCY DEPARTMENT MEASURE	50
A. Characteristics of the Denominator Populations	50
B. Measure Exclusions.....	50
C. State Variation in Follow-up after Emergency Department Performance using Different Numerator Options.....	54
D. Follow-up after Emergency Department by Beneficiary Characteristics	57
E. Relationship Between Follow-up after Emergency Department and Inpatient Stays.....	59
F. Reliability of Follow-up after Emergency Department Measure	60
G. Stakeholder Feedback on Follow-up after Emergency Department Measure	61
H. Final Outcome of Testing: Revisions to Measure Specification and National Quality Forum Submission.....	61
VII. OTHER LESSONS.....	63
REFERENCES	65
APPENDICES	
APPENDIX A. Technical Expert Panel Members.....	A-1
APPENDIX B. Measure Specifications	A-2

LIST OF FIGURES AND TABLES

FIGURE I.1. Measure Development Process	3
FIGURE V.1. BMI Screening and Follow-up for Patients with SMI	31
FIGURE V.2. Alcohol Screening and Follow-up for People with SMI	33
FIGURE V.3. High Blood Pressure Screening for People with SMI.....	35
FIGURE V.4. High Blood Pressure Screening for People with AOD	35
FIGURE V.5. Tobacco Use Screening and Follow-up for People with SMI	37
FIGURE V.6. Tobacco Use Screening and Follow-up for People with AOD.....	37
FIGURE V.7. Clinical Depression Screening and Follow-up for People with AOD	39

TABLE ES.1. Measures Tested, Performance, and Submission to NQF.....	xii
TABLE II.1. Data Sources and Key Search Terms for the Environmental Scan	5
TABLE II.2. Ambulatory Measure Concepts Selected for Specification and Testing.....	9
TABLE III.1. Measures Tested for SMI and/or AOD Population.....	11
TABLE IV.1. Quantitative Testing and Analysis of Screening and Monitoring Measures	17
TABLE IV.2. Characteristics of Health Plans that Participated in Pilot Test	18
TABLE V.1. Characteristics of Denominator Populations Selected for Screening and Monitoring Measures by Health Plan.....	27
TABLE V.2. Denominator Size for Screening and Monitoring Measures before Exclusions by Health Plan.....	28

TABLE V.3a.	Health Care Utilization in 2012 for Patients in SMI Denominator of Screening and Monitoring Measures.....	29
TABLE V.3b.	Health Care Utilization in 2012 for Patients in AOD Denominator of Screening and Monitoring Measures.....	29
TABLE V.4.	BMI Screening and Follow-up for People with SMI by Health Plan.....	30
TABLE V.5.	Alcohol Screening and Follow-up for People with SMI by Health Plan	32
TABLE V.6.	High Blood Pressure Screening and Follow-up for People with SMI or AOD by Health Plan.....	34
TABLE V.7.	Tobacco Use Screening and Follow-up for People with SMI or AOD by Health Plan	36
TABLE V.8.	Clinical Depression Screening and Follow-up for People with AOD by Health Plan	38
TABLE V.9.	Overall Measure Rate among People with SMI by Patient Characteristics	40
TABLE V.10.	Overall Measure Rate among Patients with AOD by Patient Characteristics	40
TABLE V.11.	Comparison of Health Plan Screening Measure Results with ACOs that Report Through PQRS.....	41
TABLE V.12.	Comprehensive Diabetes Care for People with SMI by Health Plan.....	42
TABLE V.13.	Controlling High Blood Pressure for People with SMI by Health Plan.....	42
TABLE V.14.	Comprehensive Diabetes Care Indicators and Controlling Blood Pressure Rates by Patient Characteristics	43
TABLE V.15.	Service Utilization among People with SMI Who did not Meet Measure Requirements.....	44
TABLE V.16.	Service Utilization among People with AOD Who did not Meet Measure Requirements.....	44
TABLE V.17.	Overall Measure Rate by Medical or Behavioral Health Data Sources	45
TABLE V.18.	Inter-rater Reliability for Screening and Monitoring Measures	46
TABLE V.19.	Summary of Testing Results and Stakeholder Feedback for Screening and Monitoring Measures.....	48

TABLE VI.1.	Characteristics of Beneficiaries in the Follow-up after Emergency Department Measure Denominator after Exclusions	51
TABLE VI.2.	Proportion of Eligible Discharges Excluded from the Follow-up after Emergency Department Measure	52
TABLE VI.3.	Follow-up after Emergency Department Rates after Applying Denominator Exclusions	53
TABLE VI.4.	Number and Percent of Follow-up after Emergency Department Denominator Remaining after Exclusions, by State.....	54
TABLE VI.5.	Performance of Follow-up for MH ED Measure by Numerator Options	55
TABLE VI.6.	Performance of Follow-up for AOD ED Measure by Numerator Options	55
TABLE VI.7.	Follow-up after Emergency Department Performance Rates by State	56
TABLE VI.8.	7-day and 30-day Follow-up Rates after Mental Health Discharge from the Emergency Department, by Patient Characteristics	58
TABLE VI.9.	7-day and 30-day Follow-up after AOD Discharge from the Emergency Department, by Patient Characteristics	59
TABLE VI.10.	Relationship Between Follow-up after Emergency Department Measure Performance and Inpatient Stays	60
TABLE VI.11.	Summary of Testing and Stakeholder Feedback for Follow-up after Emergency Department Measure	62
TABLE A.1.	Technical Expert Panel Members	A-1
TABLE B.1.	Specifications of Parent Measures and New Measures Submitted to NQF	A-2
TABLE B.2.	Specification of Parent Measure and Measures Tested but not Submitted to NQF	A-10

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 Lipoprotein
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	Reliability²	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 SMI³			
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:			
<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>			

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.

I. PROJECT RATIONALE

A number of barriers have contributed to the lack of progress in the measurement of quality for behavioral health care (IOM 2006; Pincus et al. 2011). The lack of objective measures for diagnosis, poor documentation by providers, and limited implementation of evidence-based treatments make specifying and reporting quality measures challenging. Responsibility for behavioral health care is divided among providers and multiple funding streams. For low-income and disabled patients, it is split between federal and state funding streams, including Medicaid, Medicare, state MH/SA agencies, and other state programs. Further, although quality improvement efforts among health plans and providers have spurred attention to the accessibility, costs, and outcomes of care, these are largely unconnected to public sector efforts such as U.S. Department of Health and Human Services (HHS) Substance Abuse and Mental Health Services Administration's (SAMHSA's) Uniform Reporting System and its national surveys of MH/SA programs.

The current focus on quality in federal health reform initiatives presents compelling opportunities to redress this lack of attention to quality in behavioral health. The expansion of coverage through Medicaid and exchanges has resulted in larger enrollment of low-income adults, among whom behavioral health problems are common. The Affordable Care Act (ACA) established the HHS Centers for Medicare & Medicaid Services (CMS) Inpatient Psychiatric Facility Quality Reporting (IPFQR) program. It also authorized demonstrations of new care models designed to improve integration of care between primary care and MH/SA services. For the first time, standardized reporting by states on the quality of care for children enrolled in Medicaid and the Children's Health Insurance Program, as well as for adults in Medicaid, is occurring through provisions of ACA and the Children's Health Insurance Program Reauthorization Act (CHIPRA). Through the Health Information Technology for Economic and Clinical Health (HITECH) Act, thousands of providers receive incentives for implementing and demonstrating "meaningful use" (MU) of electronic health records (EHRs). To date, few behavioral health measures are included in these landmark efforts. For example, only two measures with a behavioral health focus (that is, Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention and Preventive Care and Screening: Screening for Clinical Depression and Follow-up Plan) are included in the 2014 list of measures for the CMS EHR incentive program (CMS 2014).

With the establishment of the National Behavioral Health Quality Framework, 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 defined goals for all aspects of care, including preventing behavioral health problems, implementing and improving treatment, and promoting and supporting recovery; targets populations from young children to elderly; and includes MH/SA

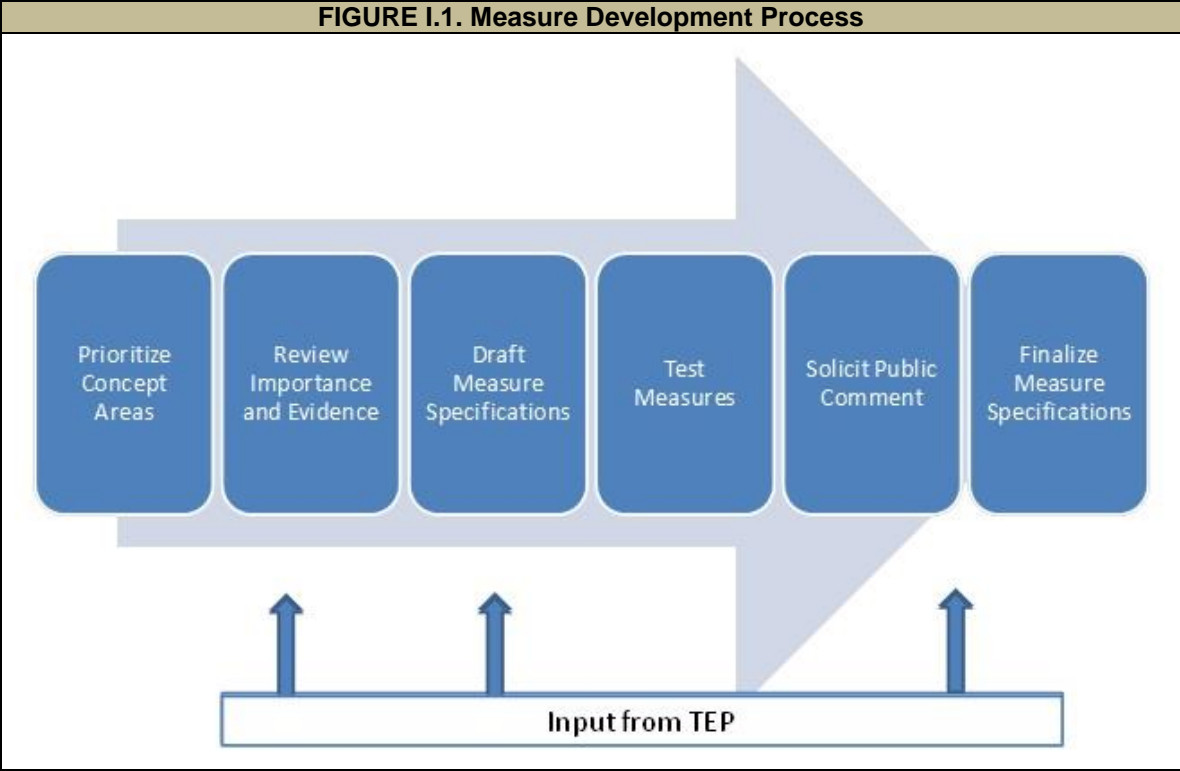
settings as well as broader health care provider and community-based efforts. Although this framework has the potential to drive quality improvement in behavioral health care, new measures are needed to monitor the quality of care and inform quality improvement efforts. Further, it is essential that efforts to identify, test, and implement new measures are aligned with other federal and state initiatives (such as the CMS EHR incentive program and Medicaid quality reporting) and take advantage of existing data sources and the evolving infrastructure for measurement (such as SAMHSA's ongoing reporting initiatives, health plan quality reporting, and new capabilities of health information technology).

A. Project Purpose

In September 2011, the HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE), with support from SAMHSA, contracted with Mathematica Policy Research and the National Committee for Quality Assurance (NCQA) to develop quality measures focused on populations who receive behavioral health care. Although this project did not begin with a mandate to develop measures for a specific public reporting program, ASPE and SAMHSA wanted the measures to be broadly applicable to Medicaid and other populations, and be suitable to potentially incorporate into national reporting programs such as the Medicaid Adult Core Set and others.

As illustrated in Figure I.1, the first step in this three-year project involved prioritizing importance measure concepts, which was informed through an environmental scan and focus groups with a range of stakeholders to identify measure gaps and priorities. The process identified several potential measure concepts, including measures that focused on preventative care and co-morbid conditions among people with serious mental illness (SMI) and/or alcohol or other drug dependency (AOD), as well as measures that focused on transitions between settings of care. We then reviewed the strength of evidence supporting each measure. After final measure concepts were selected, we developed measure specifications and pilot tested the measures. The pilot testing involved both quantitative data collection to examine the performance and psychometric properties of the measures, and qualitative data collection, including focus groups and a public comment period. Based on the findings from the testing, we refined the measure specification and submitted the strongest measures to the National Quality Forum (NQF) for endorsement in July 2014. A Technical Expert Panel (TEP) provided guidance throughout the project.

In July 2012, the contract for this project was modified to support the development of measures for the CMS IPFQR program. The development of the IPFQR measures has a different history and time line from the measures that began in September 2011, and therefore are not included in this report. A separate report for the IPFQR measures is available from ASPE.



B. Report Roadmap

This report summarizes the development and testing of the ambulatory quality measures. Chapter II describes the process for selecting measure concepts. Chapter III describes the process for specifying the measures. Chapter IV describes the methods used to test the measure, and Chapter V and Chapter VI summarize the findings. The final chapter offers lessons learned from this project that may be applicable to future measure development and implementation efforts.

II. SELECTION OF MEASURE CONCEPTS

The selection of measure concepts involved several steps: (1) conducting an environmental scan of existing measures to identify gaps; (2) holding focus groups with stakeholders to gather input on measurement priorities and where to position measures; (3) reviewing the strength of the evidence to support the measure concepts; and (4) convening a TEP to provide input on measure concepts and the evidence supporting those concepts. This chapter briefly describes these steps and how they influenced the development of the measures.

A. Scan of Measures

After initial meetings with ASPE and SAMHSA to identify priority measure-development areas, we conducted a review of existing behavioral health measures and measure-development initiatives to identify opportunities for potential measure concepts. The review was organized according to the SAMHSA Behavioral Health Quality Framework's six domains. This task involved multimode data collection drawing upon various data sources to identify the gaps in quality measurement.

Develop search criteria and taxonomy. We first developed definitions of terms used to search for and categorize data according to the domains in the framework. To align our review with other federal initiatives, we also included other domains in the taxonomy, such as those recommended by the HHS Office of the National Coordinator for Health Information Technology (ONC) Policy Committee Quality Measures Workgroup for Meaningful Use measures and categories used to organize the core sets for the CHIPRA of 2009 and Adult Medicaid. We built on the taxonomy NCQA developed for organizing measures for consideration for the Medicaid Adult Core set.

Identify and collect measures. We then searched the three most widely used sources of measures: the National Quality Measure Clearinghouse, NQF, and the online inventory maintained by the Center for Quality Assessment in Mental Health. We also reviewed measures used or developed by SAMHSA (including those developed under the Mental Health Statistics Improvement Program), the U.S. Department of Veterans Affairs (VA), and the National Association of State Mental Health Program Directors. The scan did not include measures developed from international sources, measures pertaining to dementia (per feedback from ASPE and SAMHSA), or measures for all comorbid physical conditions that could be applied to people with behavioral health conditions. Table II.1 provides a summary of the data sources and key search terms used for the scan.

To identify any additional measures that may not have been captured in these sources, we supplemented the search by reviewing findings from prior environmental scans conducted for the following projects:

- Subcommittee of the HHS Agency for Healthcare Research and Quality’s (AHRQ’s) National Advisory Council: Identifying Quality Measures for Medicaid Eligible Adults.
- American Recovery and Reinvestment Act HITECH Eligible Professional Clinical Quality Measures (CQM).
- ONC CHIPRA Electronic CQM Development.
- Development, maintenance, and support of hospital outpatient, outpatient imaging efficiency, psychiatric inpatient, and cancer hospitals quality of care measures.

TABLE II.1. Data Sources and Key Search Terms for the Environmental Scan	
Data Source	Key Search Terms/Categories
National Quality Forum	<ul style="list-style-type: none"> – Mental health – Alcohol use – Substance use or abuse – Depression – SMI
National Quality Measures Clearinghouse	<ul style="list-style-type: none"> – Disease or condition-based measure – Behavior and behavior mechanisms – Behavioral disciplines and activities – Mental disorders – Psychological phenomena and processes – Treatment or intervention-based measure
Center for Quality Assessment in Mental Health	<ul style="list-style-type: none"> – Major depressive disorder – Schizophrenia – Personality disorders – Substance abuse or dependence – AHRQ Level A: Good research evidence – AHRQ Level B: Fair research evidence – AHRQ Level C: Clinical consensus or opinion

We created a detailed spreadsheet categorizing each measure by name, description, numerator, denominator, exclusion populations or criteria, NQF identification number, domain, data source, level of specification, type of measure, condition, and age range for the relevant population. We also assigned measures to a priority area of the SAMHSA quality framework and created domains and subdomains within the framework to provide greater specificity. The domains and subdomains were based on a categorization scheme developed for the International Initiative for Mental Health Leadership project, which conducted a scan of international initiatives in mental health quality measurement (Fisher 2012).

B. Focus Groups

The focus groups were intended to obtain stakeholder input on the most relevant topics for measure development. We conducted discussions with six groups of stakeholders in February and March 2012 to identify priorities and gaps in behavioral health quality measurement. These stakeholders included: (1) consumers and consumer representatives; (2) researchers and performance measurement experts; (3) representatives of state MH/SA agencies; (4) state Medicaid program representatives; (5) health plans; and (6) providers, including specialty MH/SA providers and primary care and family practice providers. Each discussion included 6-8 individuals and had a slightly different focus given the expertise and knowledge of different stakeholder groups.

During each discussion, we asked participants to identify their priorities for behavioral health quality measurement. We asked participants to focus on quality measures for working-age adults who receive behavioral health services (mental health or substance abuse treatment) in primary care or specialty behavioral health care settings. The discussion facilitators emphasized that participants should not limit their consideration of measures to any particular payer (for example, Medicaid or private plans); accountable entity (such as the state, Medicaid, health plans, or providers); or diagnostic group. We encouraged participants to suggest measure areas and specific measures that could apply to a range of populations and service settings based on what they viewed as the most pressing quality concerns or gaps in quality measurement. Below is a brief summary of the focus of each discussion:

- In our first discussion, we asked consumers and consumer representatives about the challenges that consumers encounter when trying to access care and points in the service system that require quality improvement.
- We held a second discussion, with researchers and performance measurement experts, to gather their input on gaps in quality measurement, existing measures that could be refined, or areas in which new measures are needed.
- Our third discussion, with representatives of state MH/SA agencies, focused on gathering input on the types of measures that would help their agencies monitor and improve the quality of care.
- Our fourth meeting, with representatives of state Medicaid programs, focused on the types of measures that would help Medicaid programs monitor and improve care.
- Our fifth meeting, with representatives from health plans, allowed us to gather information about their experiences with existing quality measures, the types of new measures that would help them monitor and improve care, and the feasibility of reporting different types of measures.

- Our final meeting, with providers, focused on the saliency and clinical relevance of selected measure concepts, and the feasibility of reporting certain measures from the provider perspective.

Following the focus groups, the team reviewed the transcripts and notes to find convergence on key themes and identify divergent viewpoints. We determined to what extent the measure priorities differed across stakeholder groups. We also summarized challenges that participants identified related to development, adoption, or implementation of the measures. We submitted a memo to ASPE summarizing key findings from all the focus groups and then debriefed ASPE and SAMHSA to select measure concepts with the strongest support.

C. Evidence Review

Based on the environmental scan and focus groups, we prioritized measure concepts for ASPE and SAMHSA to review. We then conducted evidence reviews on the prioritized concepts in order to assess whether there is clear guidance to specify the denominator and numerator of a measure. The evidence reviews also addressed a critical component of NQF review -- the importance of a measure, including the extent to which it reflects a high-impact aspect of the national health care system and the evidence base supporting the measure.

We conducted reviews in five areas: (1) screening and monitoring of general health conditions among individuals with SMI and substance use disorders; (2) preventive services for risky sex behaviors among high-risk substance using populations; (3) discharge planning and post-discharge follow-up from inpatient, emergency department, or residential care; (4) shared decision making in behavioral health care; and (5) medication-assisted opioid treatment.

The reviews drew on evidence-based clinical guidelines, systematic reviews (including meta-analyses), and the recommendations of authoritative government agencies and task forces, including the U.S. Preventive Services Task Force (USPSTF), the HHS Centers for Disease Control and Prevention (CDC), and others.

The approach and methodology of each review varied depending on the measure concept. For all measure concepts, we began with a search of evidence-based clinical guidelines and systematic reviews. For some concepts, we did not find clinical guidelines or systematic reviews focused on our target condition or specifically on the SMI or AOD populations. For the measure concepts of screening and follow-up for general health conditions and infectious diseases, we reviewed USPSTF, CDC, and other recommendations for the general population. We also examined whether there is evidence of higher prevalence of certain health conditions or disparities in screening or treatment for those conditions among individuals with mental health disorders or AOD to determine whether it would be sensible to adapt existing measures for our target population. For the measure concept focused on discharge planning and post-discharge

follow-up, in the absence of clear guidelines or systematic reviews focused on our target population, we examined existing quality measures to identify opportunities for adapting them or developing new measures.

D. Technical Expert Panel Meeting

The TEP was convened to provide input on the selection of measure concepts and offer feedback on the measure specifications and testing results throughout the duration of the project. The TEP included experts in behavioral health quality measurement, the treatment of behavioral health disorders, and the organization and financing of behavioral health services. It also included representatives from consumer and family organizations, state MH/SA agencies, provider organizations, health plans, and state Medicaid programs. Representatives from several federal agencies also attended all the TEP meetings. (See Appendix A for the list of TEP members.)

The initial TEP meeting was held in July 2012 and focused on reviewing the findings from our environmental scan and focus groups. We reviewed the evidence summaries, which were provided to TEP members prior to the meeting. The TEP then prioritized measure concepts for further specification and testing.

E. Final Measure Concepts

At the conclusion of this process, ASPE and SAMHSA selected seven measure concepts for further specification and testing (Table II.2). These measures fell into three broad categories: (1) screening and follow-up for physical health and co-morbid conditions among people with SMI and AOD; (2) monitoring of chronic physical health conditions among individuals with SMI; and (3) follow-up after discharge from an emergency department for individuals with mental health conditions and AOD.

There were several factors that influenced the selection of these measure concepts, as described below:

- **Focus group and TEP support.** The focus groups and TEP strongly supported improving screening for co-morbid conditions and monitoring of chronic conditions among individuals with SMI and AOD. They also strongly supported measures that focus on transitions between different settings of care because these transitions present opportunities for individuals to lose contact with the health care system.
- **Measures fill an important gap.** The recommendations of the focus groups and TEP were consistent with our review of measure gaps. There is a lack of NQF-endorsed measures that assess whether individuals with SMI and AOD receive preventive care and monitoring of chronic conditions. In addition, their recommendations are consistent with the recommendations that NQF “examine

its portfolio of existing outcome measures and consider stratification for the MHSU [mental health and substance use] populations, thereby allowing these measures to be applied to persons with various MHSU conditions across care settings” (NQF 2011).

- **Strong evidence for the measures.** Certain co-morbid physical health conditions (obesity, diabetes, and hypertension) and health behaviors (tobacco use) are more common among individuals with SMI and AOD. These conditions are often undetected or poorly managed in these populations; individuals with SMI die, on average, two decades early due in part to these co-morbid conditions and health behaviors.

Measure Concept	Specified and Tested for SMI Population	Specified and Tested for AOD Population
BMI assessment and follow-up	✓	
Alcohol screening and follow-up	✓	
Blood pressure screening and follow-up	✓	✓
Tobacco assessment and follow-up	✓	✓
Depression screening and follow-up		✓
Comprehensive diabetes care (includes 6 indicators) ¹	✓	
Blood pressure control	✓	
Follow-up after discharge from emergency room ²	✓	✓
NOTES:		
1. Although we refer to this conceptually as Comprehensive Diabetes Care, it includes 6 separate measures. This is not calculated or reported as a composite measure.		
2. Measure was specified and tested for people discharged from the emergency department with any mental health or AOD diagnosis, not just those with SMI diagnoses.		

The process used to select measure concepts also identified a need for measures focused on the delivery and outcomes of psychosocial care. Although there are evidence-based psychosocial treatments for a number of conditions, there is a lack of quality measures to track the uptake and outcomes of these treatments. Such quality measures could help encourage greater use of evidence-based practices by providing tools for monitoring and rewarding the adoption and implementation of effective psychosocial treatments. Unfortunately, data systems commonly used for quality measurement (claims and medical records) have limited ability to capture information on the use or outcomes of effective psychosocial treatments.

In the fall of 2012 we received input from members of our TEP and several leading experts to help us identify next steps for developing quality measures focused on psychosocial care. Based on that feedback, we wrote a white paper that described the strengths and limitations of various strategies for measuring the quality of psychotherapy, and proposed next steps for the development of such measures (Brown et al. 2014).

III. SPECIFICATION OF MEASURES

The next step in the project involved developing measure specifications. After gathering feedback from our TEP and other stakeholders, we determined that our measures were most suitable for health plan reporting. Health plans have an opportunity to ensure that their patients receive preventive care and monitoring of chronic conditions as well as follow-up during care transitions.

As described below, in an effort to align our specifications for the SMI and AOD populations with measures used for the general population, we modeled the specifications on existing measures reported by health plans and providers (referred to as the “parent” measures in Table III.1). The process for specifying the screening and monitoring measures was somewhat different than the process for specifying the follow-up after emergency department measure given the different data sources used for the measures. Here we summarize the steps in the specification process and the major adaptations that were made to the parent measures. Appendix B includes the final measure specifications.

A. Specification of Screening and Monitoring Measures

Identification of data sources for measures. Based on feedback from stakeholder focus groups and our TEP, we determined that patient record review was necessary to accurately capture the numerator of these measures because several numerator components are not reliably reported using claims data. All of the screening measures, and the measures to monitor diabetes and hypertension, were specified using “hybrid” data sources. These measures use health plan administrative/claims data to identify the denominator-eligible for the measure, and primarily medical records (paper or electronic) to calculate the numerator. Depending on the measure, some numerator components can also be identified using claims data (for example, claims codes for smoking cessation treatment count toward the numerator of the tobacco screening measure).

We modeled our measures on health plan measures that are reported as part of the Healthcare Effectiveness Data and Information Set (HEDIS[®]) and on provider-level measures that are included in the CMS Physician Quality Reporting System (PQRS) (Table III.1). Here we briefly describe the overarching approach to the specification and adaptation process.

Adaptation of PQRS measures. We sought to align our health plan specifications for the tobacco, body mass index (BMI), depression, alcohol, and blood pressure screening measures with the provider-level measures that are included in PQRS. These provider-level measures are reported using G-codes or Current

Procedural Terminology (CPT) Category II codes. These are non-payment codes that can be submitted on claims forms. All the measures we adapted have also been specified for electronic reporting and several are included in the CMS Meaningful Use Incentive Program. Because the CPT II and G-codes are not routinely used, we used the narrative specification developed as part of the electronic specifications to guide our health plan specifications.

TABLE III.1. Measures Tested for SMI and/or AOD Population			
New Measure Developed for This Project (NQF # assigned for review)	Parent Measure that Served as Model (NQF #)	Parent Measure Steward	Parent Measure Use in Federal Programs (PQRS measure # where applicable)
BMI Screening and Follow-up for People with SMI (2601)	Preventive care and screening: BMI screening and follow-up (0421)	CMS	PQRS (128), MU Stage 2, NQF Duals
Alcohol Screening and Follow-up for People with SMI (2599)	Unhealthy alcohol use: Screening and brief counseling (2152)	AMA-PCPI	None (screening component of measure is similar to PQRS 173)
High Blood Pressure Screening and Follow-up for People with SMI or AOD (not submitted to NQF)	Preventive care and screening: Screening for high blood pressure and follow-up documented (not NQF-endorsed)	CMS	PQRS (317)
Tobacco Use Screening and Follow-up for People with SMI or AOD (2600)	Preventive care and screening: Tobacco use screening and cessation intervention (0028)	AMA-PCPI	PQRS (226), MU Stage 2, NQF Duals
Clinical Depression Screening and Follow-up for People with AOD (not submitted to NQF)	Screening for clinical depression and follow-up plan (0418)	CMS	PQRS (134), Adult Medicaid Core Set, MU Stage 2, NQF Duals
Comprehensive Diabetes Care for People with SMI: – HbA1c testing (2603) – A1c poor control (>9.0%) (2607) – A1c control (<8.0%) (2608) – Eye exam (2609) – Medical attention for nephropathy (2604) – Blood pressure control (2606)	Comprehensive diabetes care: – HbA1c testing (0057) – A1c poor control (>9.0%) (0059) – A1c control (<8.0%) (0575) – Eye exam (0055) – Medical attention for nephropathy (0062) – Blood pressure control (0061)	NCQA	Adult Medicaid Core Set Medicare Stars, MU Stage 2 MU Stage 2 Medicare Stars, MU Stage 2 Medicare Stars, MU Stage 2 MU Stage 2
Controlling High Blood Pressure for People with SMI (2602)	Controlling high blood pressure (0018)	NCQA	Adult Medicaid Core Set, Medicare Stars, SNP, MU Stage 2
Follow-Up After Emergency Department Use for Mental Health Conditions or AOD (2605)	Follow-up after hospitalization for mental illness (0576)		Adult Medicaid Core Set, SNP, NQF Duals

We made four main adaptations to the existing PQRS specifications, described below:

- Expanding the data source for measure exclusions.** We sought to keep the original measure exclusions. The parent PQRS measures identify exclusions through the medical record. Given that health plans have access to administrative and claims data, our health plan specification allows for the exclusions to be identified with these data sources to reduce the data collection and reporting burden on health plans, and to ensure that exclusions not documented in medical records are captured. For example, for the BMI screening measure, our specifications allow for individuals to be excluded due to pregnancy

if documented in medical record or using claims codes for pregnancy. For some measures we also changed the exclusion criteria when the original exclusion was not appropriate for health plan reporting. For example, the parent specification for the alcohol screening measure allowed for patients to be excluded from the denominator if there was a medical reason that interfered with screening during the visit (as would be appropriate for a provider-level measure) but such an exclusion is not necessary or appropriate for health plans, which have a longer time period to ensure that their patients receive screening and follow-up care.

- **Refining denominator population to focus on SMI and/or AOD and require continuous enrollment in health plan.** The denominator for each measure was limited to the SMI and/or AOD population (depending on the measure) rather than the general population. We specified the denominator based on evidence that the target condition was either more prevalent among the SMI and/or AOD population or that these populations experience disparities in care for the target condition. The specification of the SMI and AOD denominators aligned with other NQF-endorsed HEDIS measures reported by health plans. Consistent with other health plan measures, the denominator also required that the patient was consistently enrolled in the health plan for the time frame required to assess both the numerator and denominator. The period of continuous enrollment varied by measure. For example, the continuous enrollment period for the HbA1c testing measure was the measurement year with no more than one gap in enrollment of up to 45 days. This is consistent with the parent measure as this allows for identification and testing during the same measurement year.
- **Refining time frame for numerator to reflect health plans' level of accountability.** The numerator was modified to recognize the opportunity that health plans have to ensure that their patients receive care over a longer time period. The existing provider-level PQRS measures mainly assess whether screening and follow-up care was delivered by a reporting provider at the visit (or a previous visit to the same provider). In contrast, health plan measures (including those reported for HEDIS) typically use a look-back period of one year or longer to capture whether *any* provider delivered the service. We selected the appropriate time frame for our specifications by reviewing clinical guidelines and USPSTF recommendations for screening and follow-up care, and examining the evidence supporting the parent measures. We also sought to align the time frame for the numerator across our measures when appropriate to minimize confusion for health plans that may implement these measures as a group.
- **Strengthening the numerator requirements to reflect health plans' level of accountability and the intensity of services necessary for SMI and AOD populations.** Based on stakeholder feedback and guidance from the TEP, we also modified the numerator for several of the screening measures to recognize that health plans have an opportunity to ensure that their patients receive more intensive follow-up care over a longer period of time than could be expected of individual providers, and to recognize that the SMI and/or AOD populations may

require more intensive intervention than the general population given the complexity of their health, mental health, and psychosocial needs. For example, the existing provider-level PQRS numerator specification for the unhealthy alcohol use screening measure requires evidence of brief counseling (because this measure is specified for the general population and it is reasonable to hold providers accountable for delivering brief counseling during the visit). Our stakeholder focus groups and TEP recommended that brief counseling was insufficient follow-up for individuals with SMI and encouraged us to strengthen the measure by requiring two events of counseling over the measurement year following the positive screening. All of the screening measures were revised in this fashion.

At various point in the project, we reviewed our specifications with the PQRS measure developers and measure stewards (including CMS and American Medical Association Physician Consortium for Performance Improvement [AMA-PCPI]) to understand the evidence supporting the existing measures and ensure that our specification adhered to the original intent of the measure. We gathered their feedback on the adaptations described above prior to submission to NQF. We also discussed future stewardship of the measures developed as part of this project, and determined that NCQA would serve as the steward of the new health plan measures.

Adaptation of HEDIS measures. Given that the diabetes and hypertension control measures are already specified for health plan reporting for the general population, and have strong evidence to support their applicability to the SMI population, we did not make any adaptation of the exclusions or numerator of these measures. Rather, we limited the denominator to the SMI population and used the existing exclusions and numerator specifications to facilitate comparisons with the general population. During the period of our testing, new guidelines for cholesterol management for people with cardiovascular disease were published. As a result, NCQA retired two diabetes care HEDIS indicators, LDL screening and LDL control. For this reason, we removed these indicators from consideration for this measure set and do not report the results. The HbA1c <7 percent indicator of the diabetes care measure was removed from consideration and is not reported because it is not NQF-endorsed.

B. Specification of Follow-up after Emergency Department Measure

Our measure of follow-up care after emergency department discharge is calculated using only claims data. The specification was modeled on the NQF-endorsed Follow-up After Hospitalization for Mental Illness measure (NQF #0576), for which NCQA is the steward.

Defining the denominator. We sought for this measure to be broadly applicable to all mental health emergency department (MH ED) and AOD emergency department (AOD ED) visits. Based on feedback we received early in the project from our TEP and stakeholders, we limited the denominator to emergency department visits with a primary

mental health or AOD diagnosis. There are two denominator populations: MH ED visits and AOD ED visits. The mental health diagnosis codes aligned with NQF measure 0576 while the AOD diagnosis codes aligned with the Initiation and Engagement of Alcohol and Other Drug Dependence Treatment (IET) measure (NQF #0004) to reduce confusion for health plans that may report these measures in the future.

Defining the numerator. The numerator requires an outpatient or partial hospitalization visit with a primary diagnosis of mental health or AOD (mental health diagnosis at follow-up for MH ED discharges and AOD diagnosis at follow-up for AOD ED discharges). We did not restrict the numerator to visits with mental health or AOD practitioners because the TEP and other stakeholders reported that primary care visits should be considered as meeting the numerator requirement given the broad denominator population, and because primary care providers are increasingly providing behavioral health care. The measure yields the following four rates:

1. 7-day follow-up after MH ED discharges.
2. 30-day follow-up after MH ED discharges.
3. 7-day follow-up after AOD ED discharges.
4. 30-day follow-up after AOD ED discharges.

IV. APPROACH TO MEASURE TESTING

Following the specification of the measures, we pilot tested the measures using quantitative and qualitative methods. The testing was designed to assess the performance and psychometric properties of the measures and to gather information to inform their eventual implementation. Moreover, the testing was intended to gather information about the *importance*, *scientific acceptability*, *usability*, and *feasibility* of the measures, as defined in the following NQF measure criteria:

- **Importance.** The strength of evidence supporting that a measure concept promotes high-quality care and allows for differentiation in performance.
- **Scientific acceptability.** The verification that the psychometric properties of a measure -- validity and reliability -- are strong enough to justify its use to assess quality of care.
- **Validity.** The ability of measure specifications to promote *accuracy* in data collection and measure score calculation to ensure appropriate characterization of performance.
- **Reliability.** The ability of measure specifications to promote *consistency* in data collection and aggregation to ensure that variability in measure score reflects actual variation in performance.
- **Usability.** The value of a measure in informing quality improvement activities.
- **Feasibility.** The availability of data elements required for the calculation of a measure, whether a measure is susceptible to inaccuracies, and the level of effort involved in collecting and calculating the measure.

This chapter describes the methods used to test each of these criteria. We briefly summarize the overarching testing questions and then describe the specific methods.

A. Testing Questions

The testing questions vary somewhat according to the measure. Because the validity of most of the screening and monitoring measures are already established for the general population, the testing of these measures had a stronger focus on assessing the availability of data to calculate the measure for the SMI/AOD populations (feasibility), disparities in screening, follow-up, or monitoring among the SMI/AOD populations when compared with the general population (importance), whether the measures could be consistently implemented across health plans and chart abstractors

(reliability), and whether health plans and other stakeholders find value in the measure results (usability). The testing of the follow-up after emergency department measure had a stronger focus in gathering feedback on the validity of the measure, in addition to examining its importance, reliability, usability, and feasibility.

The following overarching questions guided the testing:

- Are the measures appropriate for assessing quality of care and do they address a priority condition? Is there room for improvement, and are there gaps in care? (*importance*)
- Are measure exceptions or exclusions necessary and appropriate? (*validity*)
- As specified, can the data elements and measures be calculated consistently (*reliability*) and capture the intended information? (*validity*)
- Can stakeholders use performance results for quality improvement and decision making? (*usability*)
- Can the measures be calculated accurately and without undue burden? (*feasibility*)

We collected quantitative and qualitative data to test the measures. The quantitative data collection involved gathering data from health plans or using claims data to calculate measure scores and examine various attributes of performance. The qualitative data collection involved gathering feedback from a TEP, multistakeholder focus groups, and public comment. We first describe the approach to quantitative testing of the screening and monitoring the measures and then describe the quantitative testing of the follow-up after emergency department measure. Finally, we describe our approach to collecting feedback on the specifications and measure performance using qualitative methods.

B. Quantitative Testing of Screening and Monitoring Measures

For the screening and monitoring measures, the quantitative testing was designed to answer the questions in Table IV.1. We piloted the screening and monitoring measures at three health plans, which allowed us to examine the performance of the measures if they were implemented following the typical HEDIS reporting processes for measures that use hybrid data sources (that is, using administrative claims data along with medical record review). This allowed us to observe whether health plans could reliably understand the measure specifications and access the necessary data sources and data elements to calculate the measures. Here we describe the characteristics of the health plans and data collection process.

TABLE IV.1. Quantitative Testing and Analysis of Screening and Monitoring Measures			
Criterion	Testing Question(s)	Data Source	Data Analysis
Importance/Performance Gap	<p>Is performance lower for the SMI and/or AOD population compared with the general population?</p> <p>Are there differences in performance across plans?</p> <p>Are there differences related to diagnosis or other patient characteristics?</p>	<p>Performance results for each subpopulation and the performance results for the general population</p> <p>Measure performance across health plans</p> <p>Measure performance by diagnosis and patient demographics</p>	<p>Descriptive analysis (mean, range, outliers) of performance by diagnosis, plan, and populations (SMI/AOD versus the general population)</p>
Feasibility	<p>Are the data needed to define the eligible population available?</p> <p>How large is the eligible population?</p> <p>Where are the data needed to assess the numerator (for example, primary care versus mental health records)?</p>	<p>The size of the denominator in the plans</p> <p>Data on whether the numerator was found in medical/physical health or behavioral health records</p>	<p>Descriptive analysis of the size of the eligible population by diagnosis</p>
Reliability			
Specifications	<p>Can the denominator definitions be implemented consistently across plans?</p>	<p>Data on denominator prevalence and size</p>	<p>Sensitivity analyses to explore the impact of different definitions on prevalence and sample size</p>
Inter-rater Reliability	<p>Are the data required for data element and measure calculation comparable when collected by 2 different chart abstractors?</p>	<p>Data abstracted by 2 abstractors</p>	<p>Agreement using kappa statistic</p>
Validity			
Content validity	<p>Do the definitions for the SMI and AOD denominators capture the intended populations?</p> <p>Are measure exclusions appropriate?</p>	<p>Data on denominator prevalence and size</p> <p>Performance results with and without measure exclusions</p>	<p>Descriptive analyses to explore size of denominator using different specifications</p> <p>Sensitivity analyses to explore the impact of measure exclusions on measure performance</p>

C. Characteristics of Health Plans that Participated in Testing

We sought to recruit three Medicaid health plans that were geographically diverse. We first announced the project via NCQA’s HEDIS Users Group listserv, which reached 146 health plans that reported HEDIS measures in 2013. We also sent the announcement to various stakeholder groups, including the Association for Community Affiliated Plans and the Medicaid Health Plans of America. We then conducted informational meetings with health plans that expressed interest. During the meetings, we provided additional information regarding the specifics of the measures and testing plans, and requested that each health plan submit information on its enrollment, product lines, coverage for mental health and substance use services, and accessibility to general medical and behavioral health records.

We then assessed whether the interested health plans met the following desired requirements:

- Enrolled Medicaid population (including only Medicaid beneficiaries or those eligible for both Medicaid and Medicare [dual eligibles]).
- Sufficient number of patients with SMI and AOD.
- Responsible for MH/SA benefits.
- Access to general medical and behavioral health records for their patients.
- At least two experienced medical record abstractors available for the testing.

We then conducted follow-up interviews with candidate health plans to confirm that they had the capacity to participate in the testing and discuss potential data access challenges before selecting the final three health plans. We established a memorandum of understanding with each health plan to govern the secure use of the data submitted by the health plans. We provided each health plan with a modest honorarium to offset the costs of data collection.

The final three health plans included a Dual Special Needs Plan (D-SNP) for dual eligibles, a plan that enrolled primarily disabled Medicaid beneficiaries, and a plan that enrolled only adult non-disabled Medicaid beneficiaries. These plans differed in their enrollment size and geographic location/coverage (Table IV.2). For the D-SNP, some community MH/SA services were carved out to a separate Medicaid managed behavioral health organization (MBHO) but the MBHO allowed the D-SNP full access to their data systems and patient records as part of their existing relationship and collaborated with the D-SNP as part of the testing. The other two plans were fully responsible for both medical and behavioral health benefits, and therefore had access to the records for all the selected patients.

Health Plan	D-SNP	Medicaid Disabled	Medicaid Adult
Location	Multicounty in Mid-Atlantic region	Single county in Mid-west	Single state in West
Medicare/Medicaid eligibility	Dually enrolled in Medicaid and Medicare	Enrolled in Medicaid due to disability	Enrolled in Medicaid due to poverty
Plan type	HMO and MBHO	HMO	HMO
Covered population	12,755	13,431	131,033
Covered benefits	HMO: Medical, Medicare covered mental health and AOD, and pharmacy MBHO: Medicaid community mental health services	Medicaid medical, pharmacy, mental health, and AOD	Medicaid medical, pharmacy, mental health, and AOD

D. Health Plan Data Collection

Following the recruitment of the health plans, the primary quantitative data collection consisted of three major components: (1) identification of the denominator samples; (2) submission of administrative/claims data; and (3) abstraction of medical and behavioral health records.

Identification of denominator samples. We asked each health plan to use their administrative/claims data to identify the following random samples of patients:

- **Individuals with SMI (schizophrenia, bipolar I disorder, and major depression):** Plans attempted to identify at least 100 patients who had a claim with an SMI diagnosis in calendar year 2012, defined as at least one acute inpatient admission with a principal diagnosis of schizophrenia, bipolar I disorder, or major depression **or** at least two outpatient or non-acute inpatient encounters on different dates with a principal diagnosis of schizophrenia or bipolar I disorder. We aligned the definitions of schizophrenia and bipolar disorder with the existing HEDIS measure, “Diabetes Screening for People with Schizophrenia or Bipolar Disorder who are Using Antipsychotic Medications” (NQF #1932), while the major depression definition was consistent with the HEDIS “Antidepressant Medication Management” measure (NQF #0105). Patients in the SMI sample could *not* have a diagnosis of diabetes or hypertension in 2012. Health plans attempted to identify an equal number of patients with a diagnosis of major depression, schizophrenia, and bipolar disorder to allow for testing of the measure across these subpopulations.
- **Individuals with SMI and diabetes:** Plans attempted to identify at least 100 patients who had a claim with an SMI diagnosis *and* evidence of diabetes in 2012. We used the same SMI criteria as described above but also required that the patient have a diagnosis for diabetes or received medications for diabetes. We provided plans with a detailed list of diabetes medications, which aligned with the HEDIS “Comprehensive Diabetes Care” measure (NQF #0055, #0057, #0059, #0061, #0062, #0575).
- **Individuals with SMI and hypertension:** Plans attempted to identify 100 patients who had a claim for an SMI diagnosis *and* a claim for hypertension. The denominator specification to identify patients with hypertension was consistent with the HEDIS Controlling Blood Pressure measure (NQF #0018).
- **Individuals with AOD:** Plans attempted to identify 100 patients with an AOD claim in calendar year 2012. AOD claims included outpatient visits, intensive outpatient encounters or partial hospitalizations, detoxification visits, emergency department visits, and inpatient discharges with an AOD diagnosis. The AOD definition was consistent with the denominator specification of IET measure (NQF #0004). Plans sought to identify a sample where half of the patients had an

alcohol diagnosis and the other half had a drug use diagnosis to allow for testing among these two groups.

Each of the above samples included patients who were at least 18 years old as of January 1, 2011 and who were continuously enrolled in the health plan from January 1, 2012 - December 31, 2012 (with no more than one enrollment gap of up to 45 days during the measurement year). We provided the plans with detailed tables (in a format similar to HEDIS specifications) that included the diagnosis codes and acceptable place of service codes to identify the denominator samples.

Submission of administrative/claims data. We used administrative/claims data from health plans to examine service utilization among the denominator samples and to calculate exclusions and numerators, when they could be identified using claims data. Each plan generated a data file that contained the demographics, diagnoses, encounter/service utilization counts and other data elements for every patient selected for the denominator. We provided plans with detailed instructions that contained the necessary claims and diagnosis codes to create the administrative file. Plans assigned patients a random identifier that was not linked to any other patient characteristics. Mathematica/NCQA did not have access to patient names, dates of birth, or actual patient enrollment identifiers to protect their confidentiality.

Medical and behavioral health record abstraction. Professional abstractors employed by the health plans abstracted records of patients selected for the denominator samples. Abstractors accessed both paper and electronic records, when available. To ensure that data were collected consistently across the three plans, we provided them with data collection manuals, a Microsoft Access-based electronic data collection tool, training, and ongoing assistance. The manual included the narrative measure specifications and instructions on how to collect and submit the abstracted data. Health plan staff participated in two trainings via webinar to review the data collection process and ask questions about the measure specifications. We worked closely with health plans via phone and e-mail to address questions as they arose. Plans submitted de-identified data to Mathematica/NCQA on an ongoing basis via an encrypted website. We conducted data quality checks on each submission to identify missing data or clarify unexpected data patterns. Plans then found the missing information and resubmitted their data.

In general, plans followed the procedures they would typically use for HEDIS to request records from providers and follow-up with providers. Plans most often asked providers to make charts available for review via faxed requests. If a provider did not initially respond, the plan followed-up with multiple phone calls and faxes. Two plans were able to remotely review EHRs (which often contained records from multiple providers) and one plan conducted on-site reviews when necessary. As described later in this report, some patients did not have a record available for review because he or she did not have an ambulatory visit during the measurement period.

E. Approach to Quantitative Testing of Follow-up after Emergency Department Measure

The quantitative testing of the follow-up after emergency department measure was intended to answer the following questions, aligned with NQF endorsement criteria:

- Does performance on the measure vary across states? Is there room for improvement on the measure? Are there disparities in performance by patient characteristics? (*importance*)
- Are the elements of the measure specification appropriate, such as the denominator exclusions and numerator definition? Is state performance on the measure associated with state-level rates of inpatient hospitalization? (*validity*)
- How precise is the measure at distinguishing the performance of states? (*reliability*)

Data source for testing follow-up after emergency department measure.

Because this measure relies solely on claims data, we tested it using fee-for-service Medicaid Analytic eXtract (MAX) data from calendar year 2008 (the latest year available at the beginning of this project). MAX data are created from eligibility and claims files submitted by states to CMS and then standardized into variables that can be used to create comparable measures of service use across states. These variables include information such as demographic characteristics (for example, race, ethnicity, gender, age), diagnoses, and procedures performed for each beneficiary enrolled in Medicaid at any point during the year. The data also provide the opportunity to retrospectively assess measure validity by correlating measure performance with other outcomes, such as mental health and substance use-related hospitalization. A Data Use Agreement with CMS governed our use of the data.

We limited our analysis to fee-for-service (FFS) claims because although MAX includes some encounters submitted by health maintenance organizations (HMOs) and MBHOs, encounter data do not undergo the data validation process applied to MAX FFS data, and are generally considered to be of lower quality (Byrd et al. 2013; Nysenbaum et al. 2013).

Starting with all 50 states and the District of Columbia, we excluded states where FFS data were not representative of the state Medicaid population due to high rates of HMO or MBHO enrollment (23 states), or where the eligibility information or FFS claims were unreliable or missing (four states). Within the remaining states, we included beneficiaries age 18 and older, who had full Medicaid benefits and were enrolled for the full calendar year. We did not test the measure among dual eligible beneficiaries or those who also had private insurance because we did not have access to the full claims for these populations (therefore, the testing results only represent the non-dual Medicaid FFS population). Finally, after identifying the denominator (described below), we excluded eight states that had less than 150 relevant emergency department

discharges. The final analytic file included 16 states to calculate the rate of follow-up rate after emergency department discharges for mental health diagnoses, and 15 states to calculate the rate of follow-up after emergency department discharges for AOD diagnoses. The 16 states included: Alabama, Alaska, Connecticut, the District of Columbia, Georgia, Illinois, Indiana, Kentucky, Louisiana, Minnesota, Mississippi, New Hampshire, North Carolina, Oklahoma, West Virginia, and Wisconsin. Due to a small denominator size, the District of Columbia was not included among the 15 states whose data were analyzed to calculate the rate of follow-up for AOD ED visits.

Claims from emergency departments were identified in MAX using revenue codes representing facility or professional fees. We first identified all emergency department claims and then narrowed the denominator to emergency department claims that had a primary mental health or AOD diagnosis.

Quantitative testing of exclusions. We examined four denominator exclusions that align with the parent measure (NQF #0576) to decrease reporting burden and confusion for health plans that may report both measures in the future: (1) emergency department discharges after December 1 of the measurement year (because these discharges do not allow enough time for follow-up within 30 days); (2) emergency department discharges followed by death during the 30-day follow-up period (again, because these discharges do not allow enough time for follow-up within 30 days); (3) emergency department discharges that are followed by at least one other emergency department discharge within 30-days (to count only the last emergency department visit within a 30 day period); and (4) emergency department discharges followed by an inpatient or other residential stay during the 30-day follow-up period (because inpatient or other institutional stays, such as residential care, may interfere with the ability to receive follow-up care, and these individuals would be captured in the denominator of NQF #0576).

Quantitative testing of numerator options, and meaningful differences in performance. We considered the distribution in performance using three numerator options along with input from the TEP to select the final numerator for the measure. We then performed a series of chi-square tests to assess whether performance was statistically different in high-performing versus low-performing states.

Quantitative testing of validity. In addition to gathering feedback from the focus groups and TEP on the validity of the measure (described below), we attempted to examine construct validity by exploring whether states' performance on this measure was related to their rates of inpatient hospitalization for mental health and AOD diagnoses. We hypothesized that states with higher rates of follow-up after discharge from the emergency department might have lower state-level rates of inpatient stays for mental health and AOD. This could be due to the management of behavioral health conditions in the community (or lack thereof) or to differences in access to care across the states. For example, in states with high rates of outpatient follow-up, behavioral health conditions may be effectively managed in the community, avoiding crises that

require inpatient care. In states with low rates of follow-up due to limited access to outpatient behavioral health care, individuals may be more likely to be hospitalized.

Quantitative testing of reliability. The reliability testing was designed to examine how well the measure as specified can distinguish performance between states (the ratio of signal to noise). We used a beta-binomial test to examine reliability (Adams 2009). The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. The beta-binomial approach is appropriate for measures like this one, where each denominator event represents a binary opportunity to pass or fail the measure. The approach assumes that the performance measure score (pass/fail rate) across the states has a flexible beta distribution, characterized by a signal variance. Based on the performance measure score, the observed data (number of passes/failures) for each state has a binomial distribution, which provides the noise (measurement error) variance. From the beta-binomial model, the signal and noise variances are used to calculate reliability as: $\text{Signal variance} / (\text{signal} + \text{noise variance})$.

F. Approach to Gathering Stakeholder Feedback for All Measures

In addition to the quantitative testing, we gathered feedback on the measures through health plan debriefings, stakeholder focus groups, and a public comment period. In addition, we received feedback on the testing results from the TEP. This feedback focused on the face validity of the measures and testing results (that is, whether the specification reflected guidelines or good clinical practice and whether measure performance appeared credible), the feasibility of implementing the measures (including the availability of data and burden of data collection), and the usability of the measure results (whether they would be useful for quality improvement efforts). Here we briefly describe each type of data collection.

1. Health Plan Debriefings

In March 2014, we held debriefing meetings with staff from the three health plans. During these meetings we gathered feedback on whether they were able to understand the measure specifications. We also discussed any unanticipated challenges that arose in the collection of data and the time and effort associated with data collection. Prior to each meeting, we calculated the health plan's scores on the measures and prepared a brief summary for the health plans to review, and then discussed whether they perceived that the results were credible.

2. Focus Groups

In May 2014, we hosted four focus groups to receive feedback on the all the measures. Similar to the debriefing calls, we gathered information from each group on the usability and feasibility of the measures. Participants represented four types of stakeholders:

- **Medicaid.** Focus group participants included representatives from state Medicaid programs, Medicaid plans (not involved in pilot testing), and external quality review organizations.
- **Health care organizations.** Participants included representatives from states involved in the federal dually eligible beneficiaries demonstration projects and health plans that have a D-SNP or responsibility for integrated care for dually eligible beneficiaries.
- **Integrated care.** Representatives came from community mental health centers that are grantees in SAMHSA's primary care integration program, state Medicaid, or mental health agencies that are implementing health homes, and providers in health homes.
- **Patient advocacy.** Representatives include consumers, family representatives, and advocacy organizations.

For each focus group meeting, we drafted specific questions to ask participants and modified the questions to fit the particular expertise of each group.

3. Public Comment

We solicited public comment on the measures for two weeks in June 2014. This process provided an opportunity to receive perspectives from key stakeholders who were unable to participate in the focus groups or testing process.

We used the NCQA website to gather public comment on the measures. The NCQA marketing department distributed an e-mail to 10,000+ people who subscribe to receive NCQA public comment notices. Public comment stakeholders included public and private organizations, providers, health plans, trade associations for behavioral health, and consumer advocates. The e-mail and website included a brief description of the rationale for the measures and the narrative specifications. We also sent the announcement to our TEP and focus group participants.

4. Technical Expert Panel Meeting

On June 13, 2014, we convened the final meeting of our TEP to share results from testing and obtain feedback on the measures to submit for NQF endorsement. The TEP provided input on the face validity of the measures, final measure specifications, and the credibility of the quantitative testing results.

G. Data Security

We implemented security controls and processes routinely used on projects that involve sensitive information. Health plans transmitted data to Mathematica via a secure encrypted SharePoint site that was password-protected. Access to sensitive data (including all health plan data and MAX data) was limited to the immediate team and stored on a secure password-protected network drive. We encrypted data in transit and at rest and will securely destroy any data collected at the end of the project. These safeguards are consistent with the Privacy Act of 1974, the Computer Security Act of 1987, Health Insurance Portability and Accountability Act, and the Federal Information Security Management Act of 2002, Office of Management and Budget Circular A-130, and National Institute of Standards and Technology computer security standards.

V. TESTING RESULTS FOR SCREENING AND MONITORING MEASURES

This chapter summarizes the quantitative and qualitative results of the measure testing for the screening and monitoring measures. We first present the characteristics of the denominator populations for each health plan and describe their service utilization to provide context for the measure performance. We then summarize measure performance, inter-rater reliability, and stakeholder feedback.

A. Characteristics of Denominator Population Selected for Testing of Screening and Monitoring Measures

Per our sampling instructions, the SMI population included nearly an equal number of patients with schizophrenia, bipolar I disorder, and major depression. Notably, the Medicaid disabled plan had fewer patients with bipolar I disorder (15.3 percent) and the Medicaid adult plan had fewer patients with major depression (22.1 percent). Slightly more than half of all patients with SMI had a diagnosis of hypertension (39-63 percent across plans) and about 30 percent of the SMI population had diabetes (18-39 percent across plans).

The SMI population selected for testing was diverse in terms of age, gender, diagnosis, and presence of co-morbid conditions across the three plans (Table V.1). The majority were between the ages of 26-64. Roughly half were female. Data on race and ethnicity are not presented because two of the three health plans were not able to provide these data. Likewise, the AOD population was demographically diverse and divided between patients with alcohol or drug use diagnoses (per our sampling instructions).

B. Number of Patients Included in Denominator for Screening and Monitoring Measures

The number of patients included in the denominator varied by measure and health plan (Table V.2). For example, across all plans, 884 patients with SMI were included in the denominator for the alcohol screening measure, which ranged from 219 at the Medicaid disabled plan to 345 at the Medicaid disabled plan.

SMI denominator for each measure. The alcohol and BMI screening measures used the full denominator of patients with SMI. In contrast, the tobacco screening measure did not include patients with diabetes because the diabetes care measure had a separate tobacco use indicator (and therefore we did not want health plans to collect

tobacco use twice for the diabetic population). The blood pressure screening measure did not include patients with diabetes or hypertension because this measure was intended to identify new cases of hypertension rather than assess hypertension control (which is part of the diabetes and hypertension control measures). Finally, the blood pressure control and diabetes care measures were only calculated among of the subsamples of SMI patients with those conditions.

TABLE V.1. Characteristics of Denominator Populations Selected for Screening and Monitoring Measures by Health Plan								
	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
SMI Denominator Population								
Denominator size	884	100.0	345	100.0	219	100.0	320	100.0
Primary SMI Diagnosis								
Major depressive disorder	295	33.3	119	34.4	105	47.9	71	22.1
Schizophrenia	312	35.2	116	33.6	80	36.5	116	36.2
Bipolar I disorder	277	31.3	110	31.8	34	15.5	133	41.5
Diabetes diagnosis in 2012 ¹	258	29.1	135	39.1	40	18.2	83	25.9
Hypertension diagnosis in 2012 ¹	450	50.9	135	39.1	112	51.1	203	63.4
Age								
18 - 25	48	5.4	4	1.1	8	3.6	36	11.2
26 - 64	764	86.4	270	78.2	210	95.8	284	88.7
65 and older	72	8.1	71	20.5	1	0.4	0	0.0
Gender (% female)	456	51.5	212	61.4	94	42.9	150	46.8
Insurance Coverage								
Medicaid-only	534	60.4	0	0.0	214	97.7	320	100.0
Dually eligible	350	39.5	345	100.0	5	2.2	0	0.0
Continuous Enrollment²								
Enrolled in 2011 and 2012	724	81.9	345	100.0	76	34.7	303	94.6
Enrolled in 2012 only	160	18.1	0	0.0	143	65.3	17	5.3
AOD Denominator Population								
Denominator size	306	100.0	102	100.0	102	100.0	102	100.0
Primary AOD Diagnosis								
Alcohol use disorder	146	47.7	51	50.0	51	50.0	44	43.1
Drug use disorder	160	52.2	51	50.0	51	50.0	58	56.8
Age								
18 - 25	19	6.2	1	0.9	8	7.8	10	9.8
26 - 64	278	90.8	92	90.2	94	92.1	92	90.2
65 and older	9	2.9	9	8.8	0	0.0	0	0.0
Gender (% female)	141	46.0	47	46.0	39	38.2	55	53.9
Insurance Coverage								
Medicaid-only	204	66.6	0	0.0	102	100.0	102	100.0
Dually eligible	102	33.3	102	100.0	0	0.0	0	0.0
Continuous Enrollment²								
Enrolled in 2011 and 2012	214	69.9	102	100.0	10	9.8	102	100.0
Enrolled in 2012 only	92	30.0	0	0.0	92	90.2	0	0.0
NOTES:								
1. Hypertension and diabetes were not mutually exclusive.								
2. Continuous enrollment is defined as no more than 1 gap in enrollment of up to 45 days during the measurement year, or no more than a 1-month gap in coverage for those patients with enrollment verified monthly.								

AOD denominator for each measure. The blood pressure, tobacco, and depression screening measures included all patients who met the AOD denominator criteria. The number of AOD patients did not vary across plans.

TABLE V.2. Denominator Size for Screening and Monitoring Measures before Exclusions by Health Plan								
Measure	SMI Denominator Population by Health Plans Before Exclusions				AOD Denominator Population by Health Plan Before Exclusions			
	All Plans	D-SNP	Medicaid Disabled Plan	Medicaid Adult Plan	All Plans	D-SNP	Medicaid Disabled Plan	Medicaid Adult Plan
BMI screening and follow-up ¹	884	345	219	320	---	---	---	---
Alcohol screening and follow-up ¹	884	345	219	320	---	---	---	---
Blood pressure screening and follow-up ²	306	102	102	102	306	102	102	102
Tobacco screening and follow-up ³	756	237	214	305	306	102	102	102
Depression screening and follow-up	---	---	---	---	306	102	102	102
Diabetes care ⁴	258	135	40	83	---	---	---	---
Hypertension control ⁵	450	135	112	203	---	---	---	---

NOTES:

1. Denominator includes patients with SMI only, SMI and diabetes, or SMI and hypertension.
2. Denominator for this measure includes patients with SMI only; those with SMI and diabetes or SMI and hypertension are not included because this measure assesses the identification of new cases of hypertension.
3. Patients with SMI and diabetes were not included in the denominator for this measure because the diabetes care measure contains an indicator for smoking status.
4. Denominator includes patients with SMI and diabetes.
5. Denominator includes patients with SMI and hypertension.

C. Service Utilization among Denominator Populations for Screening and Monitoring Measures

Service utilization is an important component of measure performance because patients must have an ambulatory visit in order to receive the care assessed by these measures and for the health plan to receive credit for the measure (that is, the health plan automatically fails the measure if the patient did not receive any ambulatory care during the measurement period because the services could not be received). As a result, failure to receive ambulatory care will decrease measure performance.

Service utilization varied widely across the three plans (Table V.3a and Table V.3b). Among the SMI population, across all plans, 70 percent of patients had one ambulatory visit for medical or behavioral health care in 2012. However, this ranged from 26 percent in the Medicaid adult plan to 99 percent in the D-SNP. Only 1 percent of patients with SMI in the Medicaid adult plan had both an ambulatory physical health **and** behavioral health visit during the year compared with 73 percent of the SMI D-SNP patients. The D-SNP also had the lowest average number of emergency department visits for the SMI population (Mean = 3.0 per year) whereas the Medicaid adult plan had the highest (Mean = 4.6 per year). As described in the Section D of this chapter, for most measures, the D-SNP plan consistently had the highest measure scores and the Medicaid adult plan had the lowest, suggesting that many patients in the Medicaid adult plan did not meet the numerator requirements because they did not have a visit during the year.

TABLE V.3a. Health Care Utilization in 2012 for Patients in SMI Denominator of Screening and Monitoring Measures								
SMI Denominator	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
Patients with SMI	884	100.0	345	100.0	219	100.0	320	100.0
At least 1 ambulatory visit	617	69.8	344	99.7	191	87.2	82	25.6
Only mental health or AOD visits ¹	63	7.1	8	2.3	16	7.3	39	12.2
Only medical visits ²	175	19.8	84	24.3	52	23.7	39	12.2
Both behavioral health and medical ³	379	42.9	252	73.0	123	56.1	4	1.3
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Average number of inpatient admissions for any diagnosis	1.2	2.4	1.3	1.8	1.8	4.0	0.6	1.0
Average number of emergency department visit for any diagnosis	2.9	5.3	2.9	4.7	4.2	6.7	2.1	4.6
Average number of ambulatory visits for mental health or AOD diagnoses	5.8	11.5	8.8	11.1	7.9	13.9	1.1	8.2
Average number of ambulatory visits for any diagnosis	15.5	19.9	21.9	14.8	18.3	19.6	6.6	21.8
NOTES:								
1. Had at least 1 outpatient non-emergency department visit with a principal mental health or AOD diagnosis and did not have any other outpatient non-emergency department visits for other diagnoses.								
2. Had at least 1 outpatient non-emergency department visit for physical health/medical diagnosis and did not have any outpatient non-emergency department visits with a principal mental health or AOD diagnosis.								
3. Had at least 1 outpatient non-emergency department visit with a principal mental health or AOD diagnosis and also had at least 1 outpatient non-emergency department visit for physical health/medical diagnosis.								

TABLE V.3b. Health Care Utilization in 2012 for Patients in AOD Denominator of Screening and Monitoring Measures								
AOD Denominator	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
Patients with AOD	306	100.0	102	100.0	102	100.0	102	100.0
At least one ambulatory visit	200	65.4	98	96.1	67	65.7	35	34.3
Only mental health or AOD visits ¹	34	11.1	7	6.9	4	3.9	23	22.6
Only medical visits ²	83	27.1	32	31.4	39	38.2	12	11.8
Both behavioral health and medical ³	83	27.1	59	57.8	24	23.5	0	0.0
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Average number of inpatient admissions for any diagnosis	0.4	1.2	0.7	1.4	0.2	1.1	0.4	1.0
Average number of emergency department visit for any diagnosis	1.9	3.1	2.2	3.5	1.6	2.4	1.8	3.4
Average number of ambulatory visits for mental health or AOD diagnoses	2.9	9.1	3.8	9.3	0.1	0.4	4.8	12.2
Average number of ambulatory visits for any diagnosis	11.3	16.4	21.4	19.1	7.5	11.8	5.0	12.2
NOTES: The mean number of visits is calculated for all patients included in the denominator; not just those who had at least one visit.								
1. Had at least 1 outpatient non-emergency department visit with a principal mental health or AOD diagnosis and did not have any other outpatient non-emergency department visits for other diagnoses.								
2. Had at least 1 outpatient non-emergency department visit for physical health/medical diagnosis and did not have any outpatient non-emergency department visits with a principal mental health or AOD diagnosis.								
3. Had at least 1 outpatient non-emergency department visit with a principal mental health or AOD diagnosis and also had at least 1 outpatient non-emergency department visit for physical health/medical diagnosis.								

Among the AOD population, there was a similar pattern of service utilization. Overall 65 percent had at least one ambulatory visit. Thirty-four percent of patients with AOD in the Medicaid adult plan had an ambulatory visit compared with 65 percent of Medicaid disabled plan patients and 96 percent of D-SNP patients.

The next section describes the variation in performance of each measure across the three health plans, and summarizes stakeholder feedback.

D. Screening and Monitoring Measure Testing Results

We first describe performance of the screening measures (BMI, alcohol, blood pressure, tobacco use, and depression) and then describe the monitoring measures (blood pressure control and diabetes care). Within each section, we present the variation across plans and provide a brief summary of stakeholder feedback that synthesizes input from the TEP, focus groups, and public comment.

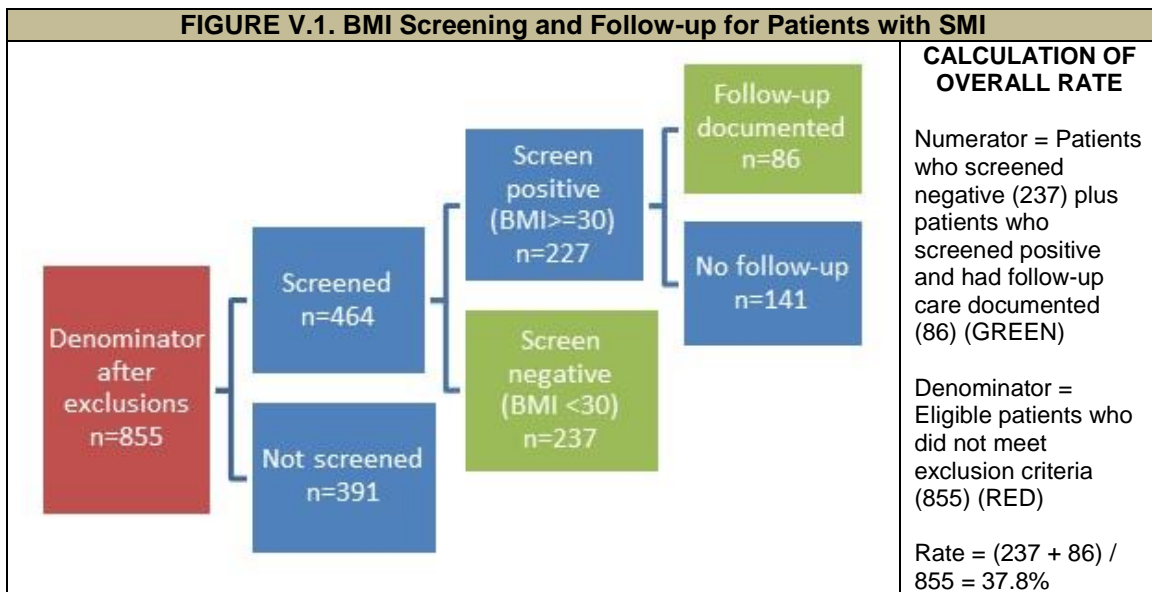
1. Body Mass Index Screening and Follow-up for People with Serious Mental Illness

Performance. Across plans, approximately 3 percent of patients were excluded from the measure due to pregnancy during the current or previous year; this is not a substantial proportion of the denominator population excluded.

There was wide variation in screening and follow-up rates across plans (Table V.4). The proportion of patients who received BMI screening ranged from 19.3 percent to 79.8 percent across plans. Among those screened, 48.9 percent had a BMI \geq 30 (range 39.3-54.5 across plans). Follow-up rates for patients with BMI $>$ 30 ranged from 14.3 percent at the Medicaid Adult plan to 43.1 percent at the D-SNP. Some type of weight management counseling or making a plan to return to the provider for future weight management were the most common forms of follow-up care (results not shown). Very few patients received bariatric weight loss surgery or weight loss medications.

	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
Exclusions								
Denominator prior to exclusions	884	100	345	100	219	100	320	100
Pregnancy ¹	29	3.3	14	4.1	6	2.7	9	2.8
Denominator after exclusions	855	96.7	331	95.9	213	97.3	311	97.2
Screening and Follow-up Results								
Documentation of BMI screening (% of denominator after exclusions)	464	54.3	264	79.8	140	65.7	60	19.3
Screened positive (BMI \geq 30) (% of screened)	227	48.9	144	54.5	55	39.3	28	46.7
Screened positive and received follow-up (% of positive)	86	37.8	62	43.1	20	36.4	4	14.3
Calculation of Performance Rate (% of denominator after exclusions)								
Screened negative (BMI $<$ 30)	237		120		85		32	53.3
Screened positive and received follow-up	86		62	43.1	20	36.4	4	14.3
Overall Rate²	323	37.8	182	55.0	105	49.3	36	11.6
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.								
NOTES:								
1. Pregnant during measurement year or previous year.								
2. Screened with BMI $<$ 30 or had BMI \geq 30 AND received at least 1 event of follow-up care during the measurement year or year prior to measurement year.								

Figure V.1 depicts the pathway to calculate the overall measure rate. The numerator of the measure includes patients who screened negative (BMI < 30), or who screened positive (BMI ≥ 30) AND received follow-up care. Across plans, 237 (27.8 percent of the denominator after exclusions) were screened and had a BMI < 30 documented. Of the 227 patients who screened positive (BMI ≥ 30), only 86 had documentation of follow-up care. Thus, across plans, the overall rate is calculated by adding the numerator components (237 with negative screen plus 86 with positive screen and follow-up) for a numerator of 323. Thus, the overall rate = (237 + 86)/855 or 37.8 percent. The overall rate ranged from 11.6 percent to 55.0 percent across plans.



Stakeholder feedback. Participants in the TEP and focus groups supported moving forward with this measure. Out of 14 total comments that were received during public comment, eight (57 percent) supported the measure as is or with modifications. Other commenters did not support the measure due to concerns about the effort required for record review but did not express concern about validity of the measure. There were concerns that some plans may not cover certain services such as nutrition counseling unless the patient is morbidly obese or has diabetes. The TEP noted these concerns but felt the importance, usability, and validity of the measure outweighed the concerns about burden.

The TEP also recommended changing the numerator requirement to include two follow-up events (for example, counseling). This change raised the intensity of service to address the high-risk status of the SMI population and to take advantage of health plans' opportunity/responsibility for follow-up care beyond the visit.

2. Alcohol Screening and Follow-up for People with Serious Mental Illness

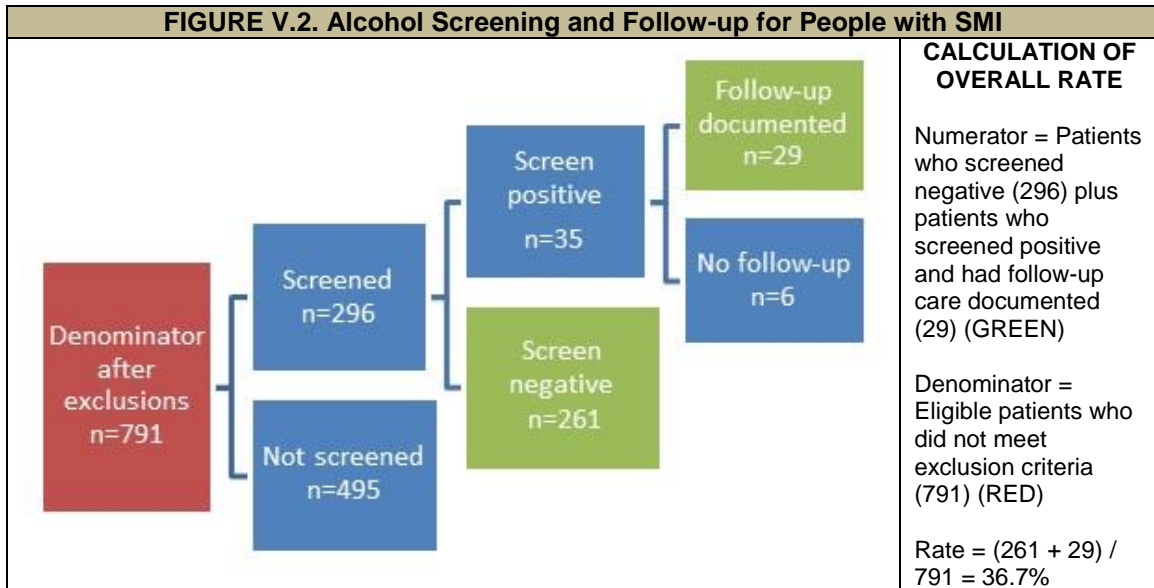
Performance. Patients identified as having an alcohol use disorder in the previous year are excluded from the denominator. Nearly 11 percent of patients

selected for the denominator were excluded because they had a visit with an alcohol diagnosis during the previous year. The Medicaid disabled plan was unable to access claims/administrative data on their patients for the previous year.

There was wide variation in screening and follow-up rates across plans (Table V.5). The proportion of patients who had documentation of alcohol screening ranged from 1.5 percent to 61.2 percent across plans. Among those screened, 11.8 percent were identified as unhealthy alcohol users (range 0-25 percent across plans). The Medicaid disabled plan has the highest rate of identification and follow-up, which may reflect their inability to identify exclusions using data from the previous year.

TABLE V.5. Alcohol Screening and Follow-up for People with SMI by Health Plan								
	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
Exclusions								
Denominator prior to exclusions	884		345		219		320	
Diagnosis of alcohol use disorder ¹	93	10.5	44	12.8	0	0.0	49	15.3
Denominator after exclusions	791	89.5	301	87.2	219	100.0	271	84.7
Screening and Follow-up Results								
Documentation of alcohol screening (% of denominator after exclusions)	296	37.4	158	52.5	134	61.2	4	1.5
Screened positive (% of screened)	35	11.8	1	0.6	34	25.4	0	0.0
Screened positive and received follow-up (% of positive)	29	82.9	1	100	28	82.4	0	n/a
Calculation of Performance Rate (% of denominator after exclusions)								
Screened negative	261		157		100		4	
Screened positive and received follow-up	29		1		28		0	
Overall Rate²	290	36.7	158	52.5	128	58.4	4	1.5
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.								
NOTES:								
1. During year prior to measurement year (identified using claims only for testing; final specification submitted to NQF also allows for exclusions found in medical record).								
2. Screened negative for unhealthy alcohol use or screened positive AND received at least 1 event of follow-up care during measurement year.								

Figure V.2 illustrates the calculation of overall measure rate across the three plans, which ranged from 1.5 percent to 58.4 percent.



Stakeholder feedback. The TEP and focus groups supported moving forward with this measure. Out of 14 total comments that were received during public comment on this measure, nine (64 percent) supported the measure as specified or with modifications. Commenters who did not support the measure cited concerns about the burden of record review but did not express concern about validity of the measure. Stakeholders were concerned that the field test showed lower rates of unhealthy alcohol use than would be expected for the SMI population. Despite these concerns, the TEP recommended moving forward with the measure due to the importance of addressing co-morbid alcohol use problems among people with SMI.

The TEP also recommended changing the numerator requirement to include two events of counseling. This change raises the intensity of service to address the high-risk status of the SMI population and to take advantage of health plans' opportunity/responsibility for follow-up care beyond the visit. In addition, the specifications were amended to allow self-help services documented in the clinical record to meet the numerator requirements. These adaptations strengthened the face validity of this measure for people with SMI and for health plan reporting.

3. High Blood Pressure Screening and Follow-up for People with Serious Mental Illness or Alcohol and Other Drug Dependency

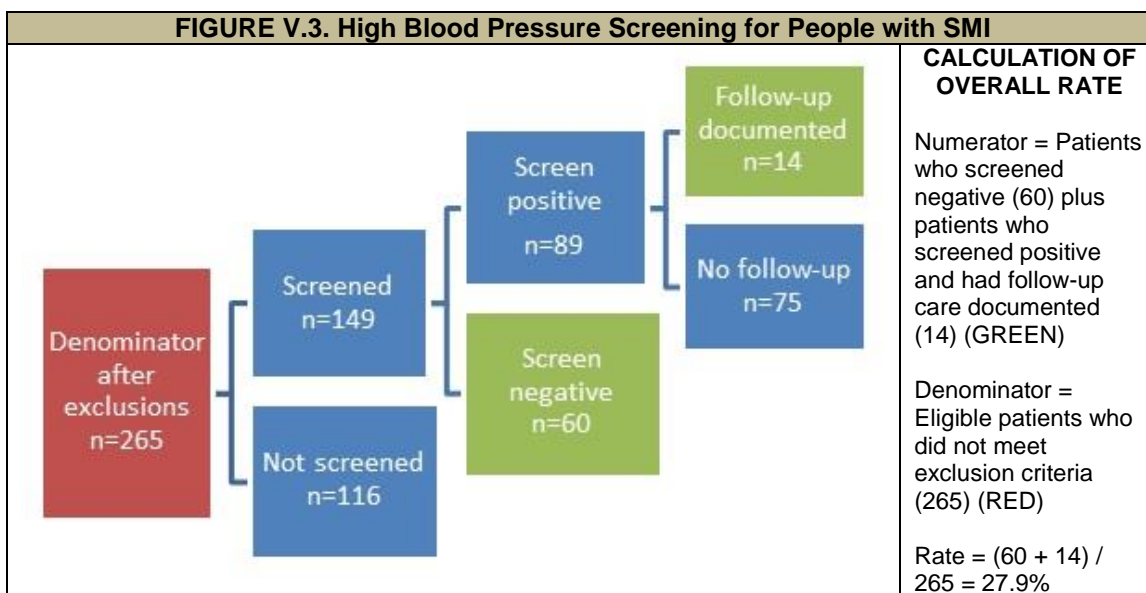
Performance. Fourteen percent of the SMI population and 39.5 percent of the AOD population was excluded (Table V.6) due to a hypertension diagnosis or medication in the previous year. There was considerable variation across plans in the impact of exclusions on final denominator sizes. For the SMI population, the D-SNP excluded considerably more patients than the other two plans (30 percent of the denominator was excluded). For the AOD population, the Medicaid disabled plan excluded considerably more patients than the other two plans (78 percent of the denominator was excluded). Overall, there were a substantial number of patients excluded from the AOD population, suggesting that roughly 40 percent of the AOD

population already had their hypertension identified in the previous year. It is possible that a lower proportion of individuals with SMI were excluded because the plans included those patients in the denominator for the controlling blood pressure measure, described later.

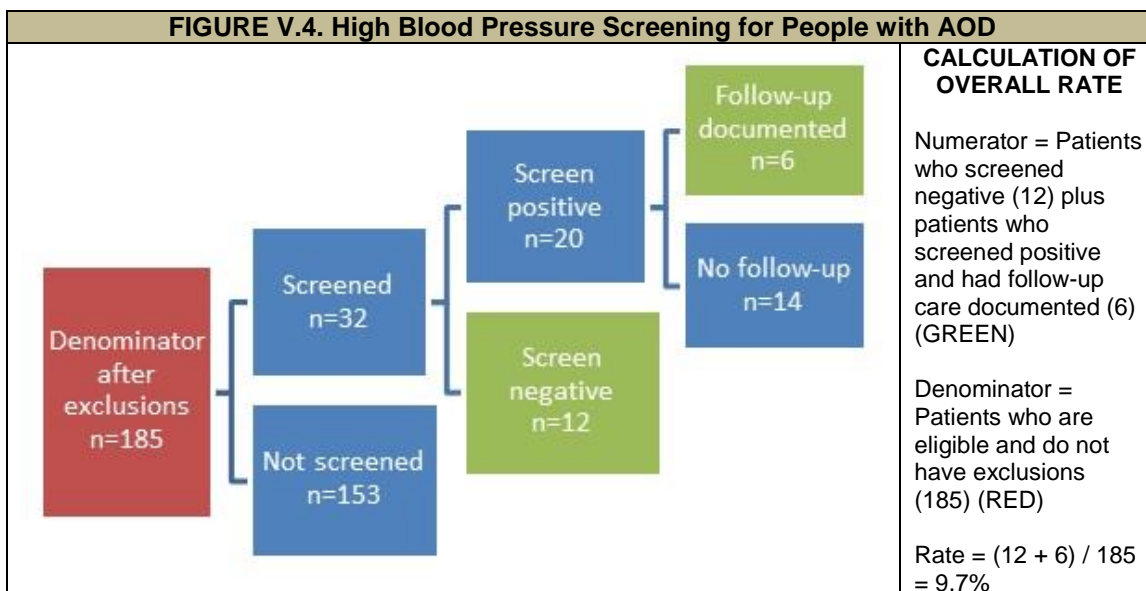
TABLE V.6. High Blood Pressure Screening and Follow-up for People with SMI or AOD by Health Plan								
	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
SMI Population								
Exclusions								
Denominator prior to exclusions	306		102		102		102	
Diagnosis of hypertension ¹	14	4.6	10	9.8	1	1.0	3	2.9
Medication for hypertension ¹	35	11.4	26	25.5	2	2.0	7	6.9
Denominator after exclusions	265	86.6	71	69.9	100	98.0	94	92.2
Screening and Follow-up Results								
Documentation of blood pressure screening (% of denominator after exclusions)	149	56.2	56	78.9	72	72.0	21	22.3
Screened positive (elevated blood pressure) (% of screened)	89	59.7	38	67.9	40	55.6	11	52.4
Screened positive and received follow-up (% of positive)	14	15.7	6	15.8	6	15.0	2	18.2
Calculation of Performance Rate (% of denominator after exclusions)								
Screened negative	60		18		32		10	
Screened positive and received follow-up	14		6		6		2	
Overall Rate²	74	27.9	24	33.8	38	38.0	12	12.8
AOD Population								
Exclusions								
Denominator prior to exclusions	306		102		102		102	
Diagnosis of hypertension ¹	98	32.0	14	13.7	79	77.5	5	4.9
Medication for hypertension ¹	40	13.1	35	34.3	2	2.0	3	2.9
Denominator after exclusions	185	60.5	66	64.7	22	21.6	97	95.1
Screening and Follow-up Results								
Documentation of blood pressure screening (% of denominator after exclusions)	32	17.3	16	24.2	2	9.1	14	14.4
Screened positive (elevated blood pressure) (% of screened)	20	62.5	13	81.3	1	50.0	6	42.9
Screened positive and received follow-up (% of positive)	6	30.0	5	38.5	1	100.0	0	0.0
Calculation of Performance Rate (% of denominator after exclusions)								
Screened negative	12		3		1		8	
Screened positive and received follow-up	6		5		1		0	
Overall Rate²	18	9.7	8	12.1	2	9.1	8	8.2
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.								
NOTES:								
1. During the year prior to measurement year, first visit of measurement year, or same day as first blood pressure reading during measurement year.								
2. Screened negative for elevated blood pressure or screened positive AND received at least 1 event of follow-up care during measurement year.								

There was variation in screening rates across plans for the SMI population (ranged from 22.3 percent to 78.9 percent). Among the SMI population, half or more were screened for hypertension across all plans and there was little variation in the proportion who received follow-up care (ranged from 15.0 percent to 18.2 percent). Figure V.3 illustrates the pathway to calculate overall measure rate, which ranged from 12.8 percent at the Medicaid adult plan to 38.0 percent at the Medicaid disabled plan (Table

V.6). Medication was the most frequent form of follow-up for patients with elevated blood pressure (results not shown). Lifestyle modifications and follow-up visit with the same provider were also commonly provided.



There was less variation in screening rates across plans for the AOD population, and the numbers were too small to compare the proportions of those who had an elevated blood pressure or received follow-up care. Figure V.4 illustrates the pathway to calculate overall measure rate. There was very little variation in overall measure rate across plans (range 8.2-12.1 percent) for the AOD population.



Stakeholder feedback. The TEP and stakeholders raised a number of concerns about this measure. The TEP and focus group participants noted that this measure would be difficult to implement and audit because the specification (following the PQRS

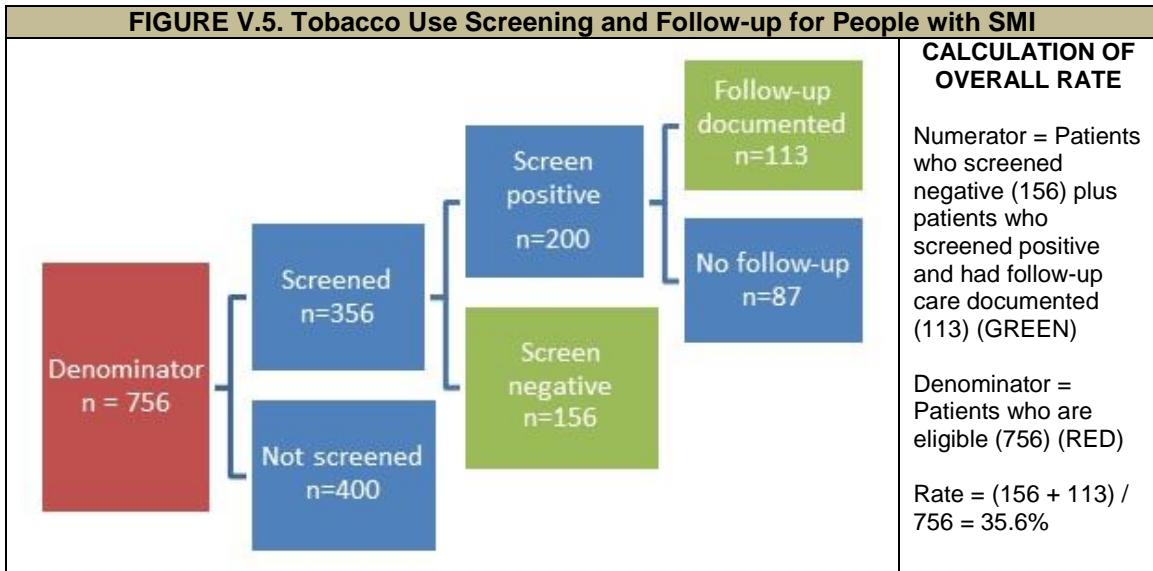
measure) requires a different type of follow-up care depending on the level of hypertension and history of hypertension which was difficult to identify from retrospective review of medical records. They also viewed hypertension control among individuals with SMI as a more important focus of quality measurement compared to screening and identification of new cases of hypertension. For the AOD population, the TEP felt that there was insufficient evidence to suggest that individuals with AOD are at greater risk for hypertension. Public comments (n = 31) were also divided; 50 percent supported the measure for AOD and 47 percent supported the measure for SMI. For these reasons, the TEP recommended not moving forward with this measure.

4. Tobacco Use Screening and Follow-up for People with Serious Mental Illness or Alcohol and Other Drug Dependency

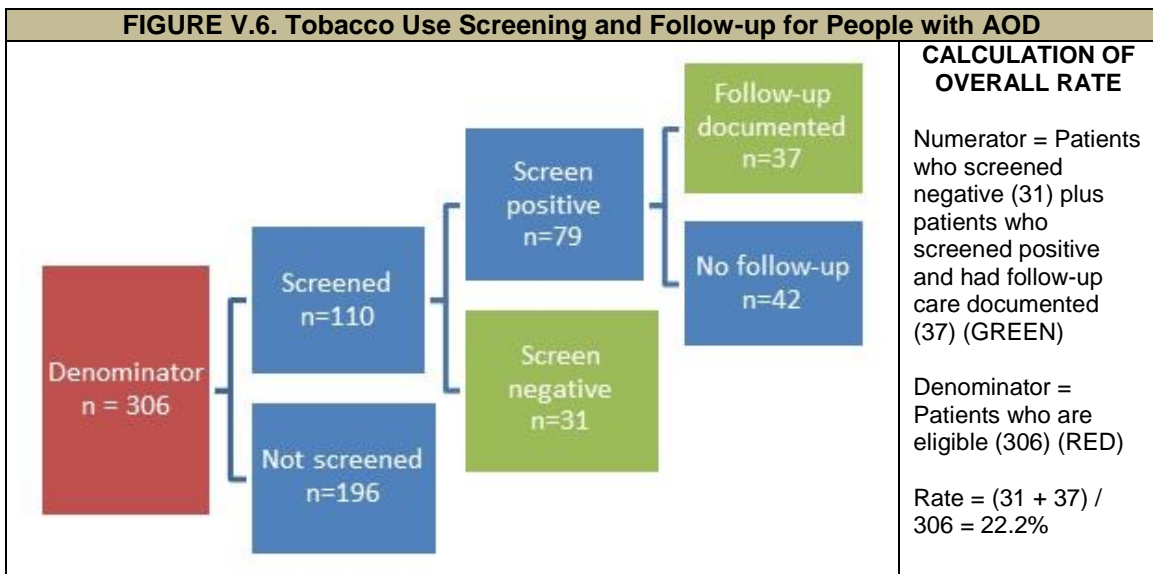
Performance. There are no exclusions for this measure.

TABLE V.7. Tobacco Use Screening and Follow-up for People with SMI or AOD by Health Plan								
	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
SMI Population								
Denominator	756		237		214		305	
Documentation of tobacco screening (% of denominator)	356	47.1	180	75.9	131	61.2	45	14.8
Screened positive (% of screened)	200	56.2	82	45.6	89	67.9	29	64.4
Screened positive and received follow-up (% of positive)	113	56.5	54	66.0	45	50.6	14	48.3
Calculation of performance rate (% of denominator after exclusions)								
Screened negative	156		98		42		16	
Screened positive and received follow-up	113		54		45		14	
Overall rate¹	269	35.6	152	64.1	87	40.7	30	9.8
AOD Population								
Denominator	306		102		102		102	
Documentation of tobacco screening (% of denominator)	110	35.9	35	34.3	61	59.8	14	13.7
Screened positive (% of screened)	79	71.8	25	71.4	44	72.1	10	71.4
Screened positive and received follow-up (% of positive)	37	46.8	18	72.0	14	31.8	5	50.0
Calculation of Performance Rate (% of denominator after exclusions)								
Screened negative	31		10		17		4	
Screened positive and received follow-up	37		18		14		5	
Overall rate¹	68	22.2	28	27.5	31	30.4	9	8.8
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.								
NOTE:								
1. Screened negative for tobacco use or screened positive AND received at least 1 event of follow-up care during measurement year or year prior to measurement year.								

There was wide variation in screening and follow-up rates across plans (Table V.7). Between 14.8-75.9 percent of patients with SMI received tobacco screening across plans; between 45.6-67.9 percent across plans were tobacco users. Among those who used tobacco, between 48.3-66.0 percent received follow-up care. Figure V.5 illustrates overall measure rate, which ranged from 9.8 percent to 64.1 percent across the plans.



Among the AOD population, between 13.7-59.8 percent of patients received tobacco screening across plans. Although each plan identified about 70 percent of those screened as tobacco users, rates of follow-up ranged from 31.8 percent to 72.0 percent. Figure V.6 illustrates overall measure rate, which ranged from 8.8 percent to 30.4 percent across plans.



For both the SMI and AOD populations, counseling was the most common type of follow-up (results not shown).

Stakeholder feedback. Stakeholder support for this measure was mixed. The TEP supported the measure because of the high rate of tobacco use among SMI and AOD populations and the concern about equitable access to cessation treatment. However, there was less support among focus groups participants and public comment. Out of 24 total comments that were received from public comment on this measure for

the SMI population, only eight (33 percent) supported or supported the measure with modifications. Of 15 comments received for the AOD population, only five (33 percent) supported or supported the measure with modifications. The burden of record review was noted for both SMI and AOD groups. For the AOD group in particular, stakeholders and public comment noted concerns about the importance of this measure and feasibility of access to records because tobacco cessation is likely to be addressed in substance use treatment. Although the TEP acknowledged concern over the burden of record reviews, they concluded that the importance, usability, and validity of the measure outweighed the concerns.

The TEP also recommended changing the numerator requirement to include two events of counseling or counseling with medication fill. These changes raise the intensity of service to address the high-risk status of the SMI and AOD populations and take advantage of health plans' opportunity/responsibility for follow-up care beyond the visit. In addition, the specifications were amended to allow new procedure codes for screening and brief intervention and community-based services documented in the clinical record to meet the numerator requirements. These adaptations strengthened the face validity of this for people with SMI or AOD and for health plan reporting.

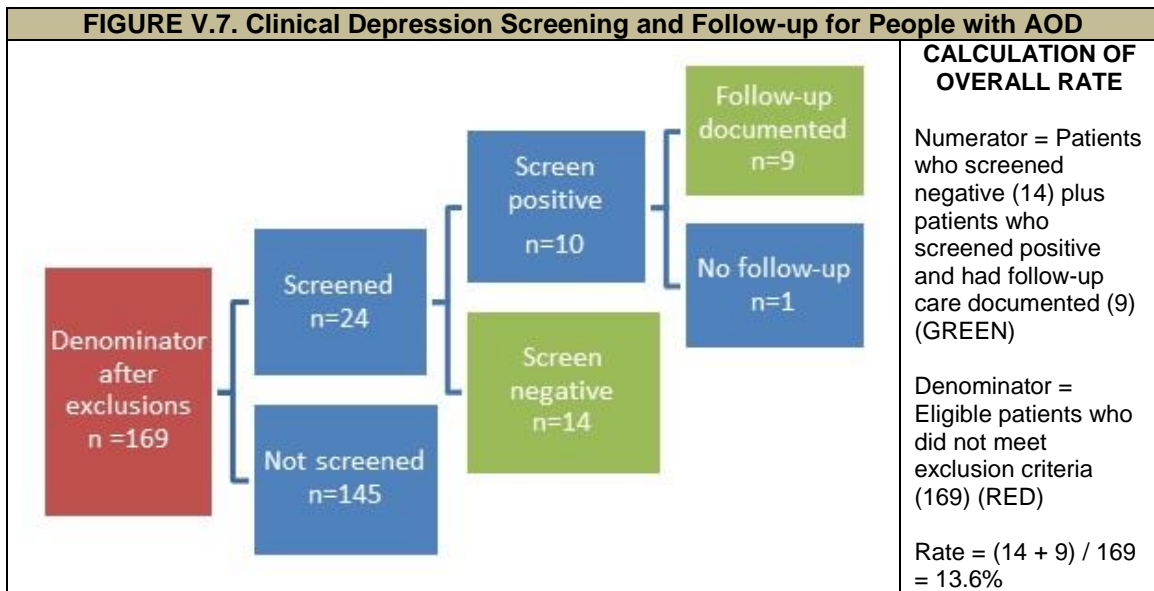
5. Clinical Depression Screening and Follow-up for People with Alcohol and Other Drug Dependency

Performance. Across the plans, 44.8 percent of patients were excluded from the denominator due to a diagnosis of depression or bipolar disorder in the previous year. This consistently excluded about half of patients from the denominator across all plans (Table V.8).

TABLE V.8. Clinical Depression Screening and Follow-up for People with AOD by Health Plan								
	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	N	%	N	%	N	%	N	%
Exclusions								
Denominator prior to exclusions	306		102		102		102	
Diagnosis of depression ¹	128	41.8	50	49.0	38	37.3	40	39.2
Diagnosis of bipolar disorder ¹	30	9.8	22	21.6	4	3.9	4	3.9
Denominator after exclusions	169	55.2	46	45.1	63	61.8	60	58.8
Screening and Follow-up Results								
Documentation of depression screening	24	14.2	9	19.6	14	22.2	1	1.7
Screened positive (% of screened)	10	41.7	1	11.1	9	64.3	0	0.0
Screened positive and received follow-up (% of positive)	9	90.0	1	100.0	8	88.9	n/a	n/a
Calculation of Performance Rate (% of denominator after exclusions)								
Screened negative	14		8		5		1	
Screened positive and received follow-up	9		1		8		n/a	
Overall rate²	23	13.6	9	19.6	13	20.6	1	1.7
NOTES:								
1. Received diagnosis during year prior to measurement year.								
2. Screened negative for depression or screened positive AND received at least 1 event of follow-up care during measurement year.								

Of those remaining in the denominator, very few patients received screening for depression using a standardized tool (1.7-22.2 percent across plans). For this reason,

the numbers are too small to report valid proportions for the screening and follow-up results. Figure V.7 depicts the overall measure rate, which ranged from 1.7 percent to 20.6 percent with the low screening rates accounting for the low overall rates.



Stakeholder feedback. Stakeholder feedback on this measure was mixed. The TEP and focus group participants were divided on the importance of this measure, while public comment was generally supportive (with 67 percent of 12 comments offering support for the measure as defined or with modifications).

The TEP and focus groups raised some concerns about the measure: (1) similar to other measures that focus on the AOD population in that they were concerned that depression screening occurs in specialty behavioral health care settings, for which health plans may have difficulty accessing data; (2) the exclusions may result in a small denominator for some health plans; and (3) the findings from the measure were not useful for quality improvement because so many patients were excluded from the denominator. They perceived that a measure focused on monitoring of depression treatment among people with AOD was needed more than a measure focused on screening to identify new cases of depression, given that the health plans had already identified so many patients with depression even in the absence of a screening measure.

6. Screening and Follow-up Measure Results by Patient Characteristics

There were not many notable differences in overall measure rate by patient age, diagnosis, and co-morbid conditions (Table V.9 and Table V.10).

With the exception of the tobacco screening measure, there were very small differences in overall measure performance between male and female patients.

	BMI Screening and Follow-up	Alcohol Screening and Follow-up	Blood Pressure Screening and Follow-up	Tobacco Screening and Follow-up
Age				
18 - 50 years	30.3	30.9	27.5	28.6
Greater than 50 years	48.7	45.4	29.3	48.0
Gender				
Male	37.6	35.4	25.7	31.2
Female	37.9	37.8	30.6	39.8
SMI Diagnosis				
Schizophrenia	37.6	37.3	23.8	33.7
Bipolar I disorder	31.6	27.2	21.3	34.0
Major depression	43.8	44.7	41.3	39.0
Co-morbid Conditions				
SMI only	41.6	39.1	27.9	39.2
SMI and hypertension	32.1	31.3	n/a	33.1
SMI and diabetes	38.7	41.9	n/a	n/a
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.				

Overall measure rates for patients over age 50 were higher than rates for adults age 18-50 for the alcohol, BMI, and tobacco screening measures among the SMI population. These appear to reflect differences in the patient population across the plans rather than actually reflecting age because most people over age 50 were in the D-SNP, which had the highest performance on these measures.

For all the screening measures, the performance was higher among peoples in the major depression group compared with those in the schizophrenia or bipolar disorder groups.

	Blood Pressure Screening and Follow-up	Tobacco Screening and Follow-up	Depression Screening and Follow-up
Age			
18 - 50 years (inclusive)	11.5	20.9	11.4
Greater than 50 years	5.6	24.3	17.2
Gender			
Male	6.8	16.4	12.4
Female	13.4	29.1	15.6
AOD Diagnosis			
Alcohol use disorder	9.3	22.6	7.2
Substance use disorder	10.1	21.9	19.8
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.			

Although there were some differences in performance of the screening measures among SMI subpopulations with different co-morbid conditions, these were not substantial.

7. Comparison of Health Plan Screening and Follow-up Measures with Physician Quality Reporting System

Because the specifications for the screening measures are based on existing provider-level measures used in the PQRS program, we compared the performance of the three health plans with the performance of Accountable Care Organizations (ACOs)

that report these measures through PQRS under the Medicare Shared Savings Program and Pioneer ACO Model. This is not a perfect comparison given the differences between the patients enrolled in Medicare ACOs and the plans in our field test. For example, those in Medicare are likely older, on average, than the patients in our field test. Nonetheless, this was the only publicly available data we could find for these measures.

The average performance of the three health plans that participated in testing was substantially lower on the tobacco screening, BMI screening, high blood pressure screening, and depression screening measures compared to ACOs reporting on similar provider-level measures in the same year (Table V.11). Although it is difficult to identify the source of such differences, they may point to disparities in care for the SMI and/or AOD population in our testing compared with the population enrolled in the ACOs.

TABLE V.11. Comparison of Health Plan Screening Measure Results with ACOs that Report Through PQRS			
Measure	Average Overall Measure Rate Across 3 Health Plans Participating in Testing (% of patients who met measure requirement)		Average Performance for ACOs Reporting to PQRS¹ under Medicare Shared Savings Program and Pioneer ACO Model (% of patients who met measure requirement)
	SMI Population	AOD Population	
BMI screening and follow-up	37.8	NA	54.3
Alcohol screening and follow-up	36.7	NA	Not reported in PQRS
Blood pressure screening and follow-up	27.9	9.7	72.1
Tobacco screening and follow-up	35.6	22.2	80.7
Depression screening and follow-up	NA	13.6	22.6
SOURCE: Performance rate for health plans were calculated using medical and behavioral health records from 3 health plans. ACO performance rates are published in "2012 Experience Including Trends (2007-2013)," PQRS and Electronic Prescribing Incentive Program, March 14, 2014 Appendix.			
NOTE:			
1. See Table III.1 for PQRS measure names and numbers.			

In addition to comparing health plan results to PQRS rates, for the BMI screening measure we compared the screening rate only (without follow-up) to the BMI screening rates for the overall population enrolled in Medicaid health plans that reported to NCQA in 2012. The rate of BMI screening was 67.6 percent for Medicaid plans, with the 10th percentile at 48.7 percent and 90th percentile of 84.4 percent compared to screening rate of 54.3 percent across the three health plans that participated in testing (ranging from 19.3 percent to 79.8 percent) -- suggesting a disparity for the SMI population compared with the overall Medicaid managed care population.

8. Comprehensive Diabetes Care for People with Serious Mental Illness

Performance. Although there are optional exclusions in the HEDIS specification of the diabetes care measure, we did not test these exclusions. The mean performance of the diabetes indicators varied from 13 percent (eye exam) to 65.4 percent (HbA1c testing) (Table V.12). There was high variation across the plans on each indicator. The variation across three plans was largest on medication attention to nephropathy (56

percentage point difference) and smallest on eye exam (15 percentage point difference).

Measure Component	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan		Average HEDIS Medicaid Rate in 2012*
	N	%	N	%	N	%	N	%	
HbA1c Testing	250	48.0	127	65.4	40	60.0	83	15.7	80.3
HbA1c Control (<8%)	250	32.8	127	48.8	40	37.5	83	6.0	46.5
HbA1c Poor Control (>9%)	250	62.8	127	44.9	40	57.5	83	92.8	44.7
Eye Exams	250	13.2	127	16.5	40	27.5	83	1.2	53.2
Med Attention for Diabetic Nephropathy	250	40.0	127	61.4	40	42.5	83	6.0	78.4
Blood Press Cont (<140/90mm/Hg)	250	42.4	127	61.4	40	45.0	83	12.0	58.9

* NCQA, 2013.

Stakeholder feedback. There was strong stakeholder support for all the diabetes indicators from the TEP, focus group participants, and public comment. The TEP and focus group participants cited clear evidence of disparities in diabetes care for the SMI population compared to national Medicaid managed care performance rates. Additionally, 83 percent of public comments supported or supported the measures with modifications. The TEP and focus group participants noted that this measure would be ready for implementation because health plans are already familiar with the parent measure, and that it would be informative for quality improvement. Health plans expressed concerns about the data collection effort for measures that require data from patient records. Overall, however, most stakeholders recognized that high-quality diabetes care cannot be adequately measured using only claims data.

9. Controlling High Blood Pressure for People with Serious Mental Illness

Performance. Very few patients were excluded from the denominator. Roughly 41 percent of people with SMI and hypertension had their blood pressure controlled, which ranged from 12.5 percent to 60.3 percent across plans (Table V.13).

Measure Component	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan		Average HEDIS Medicaid Rate in 2012*
	N	%	N	%	N	%	N	%	
Denominator prior to exclusions	195		80		75		40		NA
Pregnancy	3		2		1		0		NA
Denominator after exclusions	192	98.5	78	97.5	74	98.7	40	100.0	NA
Overall measure rate: Blood pressure adequately controlled	78	40.6	47	60.3	26	35.1	5	12.5	56.3

* NCQA, 2013.

Stakeholder feedback. There was strong support from all stakeholders for this measure based on the evidence of disparities in blood pressure control among the SMI population compared to the overall Medicaid managed care population. The majority of the public comments (80 percent) supported the measure as specified or with

modifications. However, health plans expressed concerns about the burden of data collection. Still, most stakeholders viewed these measures as feasible given the existing measure.

10. Comparison of Diabetes Care and Controlling High Blood Pressure Measure Results for Serious Mental Illness Population Versus 2012 Medicaid HEDIS Rates

The mean performance across the three plans on each of the diabetes indicators and the blood pressure control measure were 13-40 percentage points lower than the performance rates for HEDIS Medicaid plans, thus suggesting disparities in care when comparing the overall Medicaid rate to the SMI population. On average, 56 percent of patients with hypertension in Medicaid health plans had their blood pressure adequately controlled as reported by 2012 HEDIS Medicaid plans (ranging from 10th percentile of 44.8 to 90th percentile of 69.6 percent) compared with 41 percent of individuals with SMI and hypertension in the three plans that participated in our testing.

11. Variation in Diabetes Care Indicators and Controlling High Blood Pressure Measure by Patient Characteristics

The performance rate of these measures varied by age, gender, and mental health diagnosis (Table V.14). Males tended to be less likely to receive HbA1c testing and meet blood pressure goals than females. Men were more likely to have A1c (>9 percent) poor control but slightly more likely to receive an eye exam and medical attention for nephropathy. When data were combined across plans, there were differences in performance between the two age groups (18-50 years versus 51 and older). However, this result appeared to be related to variations in the age of patients across plans because most of the patients under age 50 were in the Medicaid Adult plan, which had the lowest performance on this measure.

TABLE V.14. Comprehensive Diabetes Care Indicators and Controlling Blood Pressure Rates by Patient Characteristics							
	Comprehensive Diabetes Care Indicators (proportion who met numerator for each indicator)						Controlling Blood Pressure Measure
	A1c Test	A1c <8%	A1c >9%	Eye Exam	Medical Attention to Nephropathy	BP <140/90	BP Adequately Controlled
Age							
18 - 50	37.4	20.6	74.8	14.0	74.8	30.8	31.9
51 and older	55.9	41.3	54.6	12.6	54.6	51.8	45.5
Gender							
Male	43.0	32.5	64.9	15.8	64.9	32.5	38.2
Female	52.2	32.4	61.8	11.0	61.8	51.5	43.3
Diagnosis							
Schizophrenia	52.7	33.3	62.4	19.4	62.4	36.6	35.7
Bipolar I disorder	37.7	26.0	71.4	5.2	71.4	36.4	42.5
Major depression	52.5	37.5	56.3	13.8	56.3	56.3	43.9

12. Failure to Receive Ambulatory Care Contributes to Poor Measure Rates

All the screening and monitoring measures require that patients receive some type of ambulatory service to receive screening and follow-up care or monitoring of hypertension and diabetes. The low rates of service utilization among the SMI and AOD population contributes to low rates of screening and follow-up care on the measures.

For each of the screening measures, approximately 40 percent of patients with SMI across the three plans who did not meet the measure requirement did not have an ambulatory visit in 2012 (Table V.15). The remaining patients had a visit during the year but the health plan could not find evidence that the patient received the screening and/or follow-up care. Likewise, 40-50 percent of the AOD population who did not meet the measure requirement did not have an ambulatory visit (Table V.16). These findings suggest that health plans may have an opportunity to improve performance on these measures through efforts to connect patients with SMI and AOD to ambulatory care.

	BMI Screening and Follow-up	Alcohol Screening and Follow-up	Blood Pressure Screening and Follow-up	Tobacco Screening and Follow-up
Denominator after exclusions	855	791	265	756
Number of patients who did not meet measure requirement	532	501	191	487
Of those, the number who did not have an any ambulatory visit in 2012 (% of failed)	224 (42.1)	225 (44.9)	75 (39.3)	226 (46.4)

	Blood Pressure Screening and Follow-up	Tobacco Screening and Follow-up	Depression Screening and Follow-up
Denominator after exclusions	185	306	169
Number of patients who did not meet measure requirement	167	238	146
Of those, the number who did not have an any ambulatory visit in 2012 (% of failed)	75 (44.9)	94 (39.5)	71 (48.6)

E. Variation in Screening Measure Rate by Medical and Behavioral Health Data Sources

Because health plans attempted to abstract both a medical/primary care and behavioral health record, we assessed whether using data from both records versus only the medical record improved the overall measure rate. With the exception of the alcohol and tobacco screening measures, the overall measure rate did not substantially improve when data was included from behavioral health records (Table V.17). That is, a very small number of patients met the measure requirement based solely on data from behavioral health records; the services were not frequently found in behavioral health

records. This may be either because these services are not being delivered in behavioral health settings or they are not being documented, or because the health plan was unable to access the behavioral health record.

TABLE V.17. Overall Measure Rate by Medical or Behavioral health Data Sources								
	All Plans		D-SNP		Medicaid Disabled Plan		Medicaid Adult Plan	
	Medical Only	Medical + BH	Medical Only	Medical + BH	Medical Only	Medical + BH	Medical Only	Medical + BH
Overall Measure Rate for SMI Population								
BMI screening and follow-up	35.6	37.8	52.4	55.0	45.5	49.3	10.9	11.6
Alcohol screening and follow-up	28.8	36.7	45.2	52.5	41.1	58.4	0.7	1.5
Tobacco screening and follow-up	32.9	35.6	62.4	64.1	33.6	40.7	9.5	9.8
Blood pressure screening and follow-up	26.7	27.9	35.2	33.8	33.7	38.0	12.8	12.8
Overall Measure Rate for AOD Population								
Blood pressure screening and follow-up	8.0	9.7	7.6	12.1	4.3	9.1	9.2	8.2
Tobacco screening and follow-up	19.9	22.2	22.5	27.5	28.4	30.4	8.8	8.8
Depression screening and follow-up	7.7	13.6	10.6	19.6	10.8	20.6	1.6	1.7
SOURCE: Administrative data and medical and behavioral health records from 3 health plans.								
NOTE: Medical = medical/primary care record; BH = behavioral health record.								

The rates for the alcohol and tobacco screening measures at the Medicaid disabled plan increased most substantially when the behavioral health data was included in the measure calculation. The performance rate for alcohol screening among the SMI population increased from 41.1 percent to 58.9 percent when information from the behavioral health record was incorporated into the measure calculation. The performance rate for the tobacco screening measure among the SMI population increased from 33.6 percent to 41.1 percent when data from the behavioral health data was included. For the remaining measures, incorporating data from behavioral health records yielded only a modest increase in measure performance; medical/physical health records contributed most of the data that provided credit for the numerator. We do not report this comparison for the diabetes or controlling blood pressure measure because almost no data for these measures came from behavioral health records, so differences in performance were even more negligible than the screening measures.

F. Inter-rater Agreement for Screening and Monitoring Measures

Inter-rater reliability assesses whether two chart abstractors, independently reviewing data from the same record, agreed on whether the patient met the requirements for the numerator, denominator, and/or exclusions for the measure. In order to assess inter-rater reliability each plan had two abstractors independently abstract the same record for a sample of charts. We used Cohen's kappa statistic, a measure of agreement adjusted for chance, to quantify agreement.

Inter-rater reliability was moderate to high for all measures. Percent agreement ranged from 74-100 percent across measure components (Table V.18). With the exception of the diabetes HbA1c control indicators/measures, the kappa coefficients for measure components ranged from 0.65-1.00 across measures, which is considered substantial or almost perfect agreement (Landis and Koch, 1977).

TABLE V.18. Inter-rater Reliability for Screening and Monitoring Measures				
	Number of Patient Charts Double Abstracted	Percent Agreement	Kappa Coefficient	95% Confidence Interval
BMI Screening and Follow-up				
Exclusions	223	92.8	0.85	0.78, 0.92
Numerator	223	92.6	0.84	0.77, 0.92
Blood Pressure Screening and Follow-up				
Exclusions	122	96.7	0.91	0.83, 1.00
Numerator	122	93.4	0.86	0.77, 0.95
Alcohol Screening and Follow-up				
Exclusions	NA	NA	NA	NA
Numerator	223	90.0	0.79	0.72, 0.88
Tobacco Screening and Follow-up				
Exclusions	NA	NA	NA	NA
Numerator	230	87.0	0.74	0.65, 0.82
Depression Screening and Follow-up				
Exclusions	46	95.7	0.91	0.79, 1.00
Numerator	46	89.1	0.77	0.58, 0.95
Comprehensive Diabetes Care				
Exclusions	NA	NA	NA	NA
HbA1c Testing	69	85.5	0.65	0.46, 0.85
HbA1c Poor Control (>9%)	69	73.9	0.49	0.29, 0.68
HbA1c Control (<8%)	69	75.4	0.51	0.31, 0.71
Eye Exam	69	89.9	0.74	0.56, 0.92
Medical Attention for Nephropathy	69	88.4	0.76	0.61, 0.92
Blood Pressure Control (<140/90mm/Hg)	69	88.4	0.75	0.59, 0.91
Controlling High Blood Pressure				
Denominator	67	91.0	0.68	0.44, 0.91
Exclusions	53	100.0	1.00	1.00, 1.00
Numerator	53	94.3	0.88	0.74, 1.00
SOURCE: Abstracted medical and behavioral health records at 3 health plans; records were independently abstracted by 2 abstractors at each health plan to calculate inter-rater reliability.				

The kappa coefficient for the diabetes HbA1c control numerators were the lowest with 0.49 for HbA1c poor control (>9 percent) and 0.51 for HbA1c control (<8 percent). One reason for this was that at one site, one abstractor did not systematically capture the numeric result of the HbA1c test. The diabetes and high blood pressure control kappa coefficients also have the widest confidence intervals due to small sample size. The denominators of the screening and diabetes measures and the exclusions for the alcohol screening measure were identified using administrative data and therefore did not require chart-abstraction or assessment of inter-rater reliability. The tobacco measure does not have exclusions, and we did not collect optional exclusions for the diabetes measure.

Overall, the results indicate that abstractors independently interpreted the measure specifications consistently. The high inter-rater reliability at the critical data element level also provides confidence in the validity of the measure.

G. Conclusion from Testing of Screening and Monitoring Measures: Revisions to Measure Specification and National Quality Forum

Overall, measures for monitoring diabetes and hypertension had strong stakeholder support and good testing results. For the screening measures, the testing results and stakeholder support varied depending on the focus of the measure and the target population. Table V.19 provides a high-level summary of the testing results and final NQF submission decisions, which were made in collaboration with ASPE and SAMHSA.

For the SMI population, the comprehensive diabetes care, controlling high blood pressure, BMI screening, and tobacco screening measures had the strongest testing results and stakeholder support. Health plans were able to reliably understand and implement the measures. Although we and some TEP members questioned the credibility of the alcohol screening measure results for the SMI population (due to the low screening rates), the TEP, and many stakeholders perceived that this measure was important enough to move forward for NQF consideration. The TEP, health plans, and some stakeholders were less enthusiastic about the blood pressure screening measure because they found that it was overly complicated to implement and perceived that blood pressure control was more important than screening for new cases of hypertension.

The measures that were focused on the AOD population faced more implementation challenges compared with the measures for the SMI population and, in general, received less stakeholder support. The primary challenge involved the feasibility of health plans accessing data for individuals with AOD due to providers' and health plans' interpretations of federal and state privacy rules. The tobacco screening measure was the most promising measure for the AOD population. Stakeholders did not view blood pressure screening as a high priority focus of quality measurement for the AOD population and were concerned about the complexity of the measure. The high rate of prior depression diagnoses among the AOD population led to concerns that quality measurement would best be focused on examining the adequacy of treatment for depression rather than identifying new cases.

Based on the findings from our testing and stakeholder feedback, we revised the final specifications of the screening measures prior to NQF submission to strengthen the numerator requirement. An overarching concern from the TEP and some stakeholders was that the screening measures should require more intensive follow-up for individuals with SMI and AOD (that is, that one follow-up event was insufficient). In response, we returned to the literature and clinical guidelines to identify follow-up steps that would strengthen the numerator requirements. Our analysis determined that more intensive follow-up was justified based on existing literature and guidelines. The final specifications require two follow-up events rather than one (see Appendix B for final measure specifications). In addition, the specifications take into account the need to allow time for follow-up time within the measurement period.

We aligned the diabetes and controlling blood pressure measure specifications with 2015 HEDIS reporting requirements prior to submission to NQF. During the period of our testing, new guidelines for cholesterol management among people with diabetes were published. As a result, NCQA retired two indicators, LDL screening and LDL control. For this reason, we removed these indicators from consideration for this measure set. We did not submit the HbA1c control (<7.0 percent) indicator of the diabetes care measure because it does not have NQF endorsement for the general population. The TEP and stakeholders did not suggest any changes to either the diabetes or hypertension measures.

There were consistent concerns expressed about the burden of multiple new measures requiring chart review. Health plans could coordinate the denominator sampling to decrease the burden of data collection. This could be done by coordinating sampling for existing measures -- for example, if a plan is currently reporting the Comprehensive Diabetes Care measure for the general population, it could draw that sample and then oversample to obtain a sufficient number of patients for reporting on the SMI population separately. Alternatively, samples for people with SMI and diabetes or hypertension could be drawn for reporting the suite of prevention and monitoring measures (as we did in our field test). In this way, a single record review could provide data on multiple measures.

TABLE V.19. Summary of Testing Results and Stakeholder Feedback for Screening and Monitoring Measures

Measure	Testing Results	Concerns from Stakeholders (focus groups, TEP, and public comments)	Revisions to Specification Following Testing	NQF Submission Decision (submitted July 2014) ¹
BMI screening and follow-up for people with SMI	Good variation Credible results Strong stakeholder support	1 follow-up event insufficient for SMI population Effort required for data collection	Require 2 follow-up events rather than 1	Submitted
Alcohol screening and follow-up for people with SMI	Good variation Not credible positive screening results Strong stakeholder support	Unrealistically low number of individuals screened positive; some services may be provided in settings not captured in health plan data sources 1 follow-up event insufficient for SMI population Effort required for data collection	Require 2 follow-up events rather than 1	Submitted

TABLE V.19 (continued)

Measure	Testing Results	Concerns from Stakeholders (focus groups, TEP, and public comments)	Revisions to Specification Following Testing	NQF Submission Decision (submitted July 2014) ¹
Blood pressure screening and follow-up for people with SMI or AOD	<p>For SMI: Good variation Credible results Divided stakeholder support</p> <p>For AOD: Poor variation Large number excluded Divided stakeholder support</p>	<p>Already have blood pressure control measure for SMI</p> <p>Overly complicated numerator; effort required for data collection</p> <p>Insufficient evidence to support measure for AOD population</p> <p>Difficulty accessing records for patients with AOD</p>	No revisions	Not submitted
Tobacco screening and follow-up for SMI or AOD	<p>For SMI: Good variation Credible results Divided stakeholder support</p> <p>For AOD: Less variation than SMI Less credible results than SMI Divided stakeholder support</p>	<p>Credible performance for SMI population but not credible rate of tobacco use among AOD population may reflect data access challenges for AOD</p> <p>1 follow-up event insufficient for SMI and AOD populations</p> <p>Effort required for data collection</p>	Require 2 follow-up events rather than 1	Submitted
Depression screening and follow-up for people with AOD	<p>Large number excluded</p> <p>Divided stakeholder support</p> <p>Did not yield useful information</p>	<p>Under-identification of depression appears to be less of a quality problem; focus should be on adequacy of treatment</p> <p>Difficulty accessing records for patients with AOD</p>	No revisions	Not submitted
Comprehensive diabetes care for people with SMI (6 measures)	<p>Good variation</p> <p>Credible results</p> <p>Strong stakeholder support</p>	<p>Some health plans thought this measure was most applicable to plans responsible for both medical and behavioral health benefits</p> <p>Effort required for data collection</p>	Aligned with HEDIS 2015 specifications	Submitted (6 separate measures)
Controlling high blood pressure for people with SMI	<p>Good variation</p> <p>Credible results</p> <p>Strong stakeholder support</p>	<p>Some health plans thought this measure was most applicable to plans responsible for both medical and behavioral health benefits</p> <p>Effort required for data collection</p>	Aligned with HEDIS 2015 specifications	Submitted

NOTE:

1. All of the measures submitted to NQF in Table V.19 were endorsed on March 6, 2015.

VI. TESTING RESULTS FOR FOLLOW-UP AFTER EMERGENCY DEPARTMENT MEASURE

This chapter summarizes the quantitative results and stakeholder feedback for the follow-up after emergency department measure. We first present the characteristics of denominator population. We then present the measure performance, reliability, validity, and stakeholder feedback.

A. Characteristics of the Denominator Populations

Across the 16 states included in this analysis, nearly 12,000 Medicaid beneficiaries had an emergency department discharge with a primary AOD diagnosis and 27,000 had an emergency department discharge with a primary mental health diagnosis in calendar year 2008 that met our denominator criteria after exclusions (Table VI.1). The mental health denominator included a larger proportion of female beneficiaries and was younger compared with the AOD denominator. Among beneficiaries with MH ED discharges, roughly 60 percent were female and between the ages of 21-44. Among beneficiaries with an AOD ED discharge, 52 percent were male, and most were between the ages of 21-44 and 45-64 (46 percent and 48 percent, respectively). The majority of beneficiaries included in both denominator groups were Caucasian (56 percent with MH ED visits and 59 percent with AOD ED visits) and just over 40 percent lived in metropolitan areas.

The denominator for the measure is based on discharges, rather than individual patients/beneficiaries; the number of beneficiaries was smaller than the total number of emergency department discharges, indicating that some individuals were discharged from the emergency department to the community more than once during the measurement year. The final denominator of MH ED discharges included 31,952 total discharges across 16 states, with state denominator sizes ranging from 181 (District of Columbia) to 5,681 (Illinois). The final denominator of AOD ED discharges was smaller; 13,337 visits across 15 states, which ranged across states from 212 visits (Alaska) to 2,412 visits (North Carolina).

B. Measure Exclusions

To understand the impact of exclusions on the denominator size, we tested them in two ways. First, we reviewed the proportion of the denominator removed due to each exclusion. Next, we examined how the exclusions affected measure performance.

TABLE VI.1. Characteristics of Beneficiaries in the Follow-up after Emergency Department Measure Denominator after Exclusions				
Characteristic	Beneficiaries with MH ED Discharges (states = 16)		Beneficiaries with AOD ED Discharges (states = 15)	
	N	%	N	%
Total Individuals	26,982	100	11,743	100
Gender				
Male	10,744	39.8	6,068	51.7
Female	16,238	60.2	5,675	48.3
Unknown	0	0	0	0
Age				
18 - 20	2,015	7.5	550	4.7
21 - 44	15,602	57.8	5,447	46.4
45 - 64	9,214	34.1	5,656	48.2
65 - 74	132	0.5	84	0.7
75 - 84	17	0.1	6	0.1
85+	2	0	0	0
Race/Ethnicity				
African American	8,920	33.1	3,324	28.3
Caucasian	15,144	56.1	6,934	59.0
Hispanic	883	3.3	326	2.8
Other	485	1.8	377	3.2
Unknown	1,550	5.7	782	6.7
Medicaid Eligibility Category				
Adult	3,877	14.4	1,876	16.0
Disabled	22,439	83.2	9,575	81.5
Children 18+ ¹	666	2.5	292	2.5
Geography				
Metropolitan	11,146	41.3	5,021	42.8
Micropolitan	7,887	29.2	3,315	28.2
Other	7,845	29.1	3,383	28.8
Unknown	104	0.4	24	0.2
SOURCE: MAX 2008.				
NOTES: Counts in this table are Medicaid FFS beneficiaries age 18+ with full Medicaid benefits discharged from the emergency department to the community after all exclusions in Table VI.2 were applied. Dually eligible beneficiaries and those with private insurance are not included.				
1. Includes beneficiaries that remain in the "child" eligibility category after their 18th birthday.				

Less than 10 percent of emergency department discharges occurred after December 1 of the measurement year (exclusion 1) and less than 1 percent resulted in death within 30 days (exclusion 2) (Table VI.2). Sixteen percent of MH ED discharges are followed by another MH ED discharge within 30 days and are therefore excluded from the denominator (exclusion 3). Likewise, 18 percent of AOD ED discharges are followed by another AOD ED discharge and are excluded. For both the mental health and AOD denominators, more than one-third of discharges are followed by an inpatient or other residential stay within 30 days (exclusion 4). The rationale behind this exclusion is that an inpatient or institutional stay may interfere with the ability of the beneficiary to receive ambulatory follow-up care after an emergency department discharge. However, the difference in performance due to this exclusion is only 1-2 percentage points for all four variations of the measure (Table VI.3).

TABLE VI.2. Proportion of Eligible Discharges Excluded from the Follow-up after Emergency Department Measure			
Exclusion	Rationale	Proportion of MH ED Discharges Excluded	Proportion of AOD ED Discharges Excluded
1. Emergency department discharges after December 1	If an emergency department discharge is after December 1, then the full 30-day follow-up period is not available for patient to receive follow-up care during the measurement year.	7.5	6.9
2. Death within 30 days of emergency department discharge	Death prevents follow-up care.	Less than 1	Less than 1
3. For an emergency department discharge where the patient also visited the emergency department in the previous 30 days, exclude those previous emergency department discharges	<ul style="list-style-type: none"> – Including these emergency department discharges would influence the number of discharges in the denominator and measure performance. – This exclusion aligns with the NQF-endorsed (#0576) Follow-up after Hospitalization for Mental Illness measure to reduce the burden and confusion for health plans implementing both measures. 	16.2	17.3
4. Emergency department discharges with an inpatient or other residential stay during follow-up period	<ul style="list-style-type: none"> – An inpatient or otherwise residential stay may interfere with the receipt of outpatient follow-up care. – This exclusion aligns with the NQF-endorsed (#0576) follow-up after hospitalization for mental illness measure to reduce the burden and confusion for health plans implementing both measures. 	34.2	40.8
All exclusions (proportion excluded for any of the exclusions above)		34.7	35.6
<p>SOURCE: MAX 2008.</p> <p>NOTE: The exclusions presented in this table are not mutually exclusive. For example, exclusions 1 and 4 may apply to the same discharge and would be counted toward the proportion reported for each exclusion.</p>			

TABLE VI.3. Follow-up after Emergency Department Rates after Applying Denominator Exclusions		
Measure	Average Measure Performance Applying Exclusions 1-3	Average Measure Performance Applying Exclusions 1-4
Mental Health: 7-day follow-up	64.6	66.0
Mental Health: 30-day follow-up	75.6	76.1
AOD: 7-day follow-up	61.9	64.3
AOD: 30-day follow-up	65.9	66.7
SOURCE: MAX 2008.		
NOTES: The overall performance rates presented here are pooled across states.		
The 4 exclusions are: (1) emergency department discharges after December 1 of the measurement year (because these discharges do not allow enough time for follow-up within 30 days); (2) emergency department discharges followed by death during the 30-day follow-up period (again, because these discharges do not allow enough time for follow-up within 30 days); (3) emergency department discharges that are followed by at least one other emergency department discharge within 30 days (to count only the last emergency department visit within a 30 day period to prevent incentivizing emergency department visits); and (4) emergency department discharges followed by an inpatient or other institutional stay during the 30-day follow-up period.		

After reviewing these findings with the TEP, we determined that all four exclusions are necessary for the face validity of the measure and to maintain consistency with the follow-up after hospitalization for mental illness measure (NQF #0567) and other measures. Further, eliminating any one of these exclusions from the measure specification does not substantially change average measure performance across states. At the state level, the exclusion rate ranges from 26.6 percent of eligible MH ED discharges in Alaska to 48.6 percent in Illinois, and from 23.1 percent of eligible AOD ED discharges in Alabama to 60.7 percent in Illinois. Across all states, when all exclusions are applied, 35 percent of eligible MH ED discharges and 36 percent of eligible AOD ED discharges are excluded (Table VI.4), leaving roughly two-thirds of the eligible population in the denominator.

State	MH ED Discharges			AOD ED Discharges		
	Number of MH ED Discharges Before Exclusions	Number of MH ED Discharges After Exclusions (final denominator)	Percent of Discharges Remaining After Exclusions	Number of AOD ED Discharges Before Exclusions	Number of AOD ED Discharges After Exclusions (final denominator)	Percent of Discharges Remaining After Exclusions
AK	297	221	74.4	294	212	72.1
AL	3,244	2,294	70.7	1,135	873	76.9
CT	2,800	1,608	57.4	2,081	1,135	54.5
DC ¹	311	181	58.2	302	N/A	N/A
GA	5,009	3,506	70.0	1,796	1,273	70.9
IL	11,057	5,681	51.4	3,179	1,248	39.3
IN	1,405	990	70.5	765	563	73.6
KY	4,762	3,520	73.9	1,879	1,403	74.7
LA	3,738	2,447	65.5	1,451	1,081	74.5
MN	3,192	2,149	67.3	1,100	747	67.9
MS	1,198	842	70.3	524	392	74.8
NC	6,755	4,907	72.6	3,372	2,416	71.6
NH	800	574	71.8	292	188	64.4
OK	1,183	813	68.7	717	514	71.7
WI	1,491	1,041	69.8	895	588	65.7
WV	1,699	1,178	69.3	934	704	75.4
Total	48,941	31,952	65.3	20,716	13,337	64.4

SOURCE: MAX 2008.
NOTE:
1. The District of Columbia was not included in the AOD ED denominator due to small sample size (less than 150 AOD ED discharges).

C. State Variation in Follow-up after Emergency Department Performance using Different Numerator Options

We examined the average performance and distribution of state-level performance on the measure: the minimum, maximum, median, and the interquartile range (IQR). The IQR is the difference between the values at the 25th and 75th percentiles of a distribution. A larger IQR indicates greater variation in performance. Measures with low variability/low IQR (for example, less than 10 percentage points) may be less useful for comparing entities when performance on the measure is uniformly high.

For each denominator population, we report two follow-up rates within 7 days of emergency department discharge and within 30 days of emergency department discharge for a total of four rates:

1. 7-day follow-up after MH ED discharges.
2. 30-day follow-up after MH ED discharges.
3. 7-day follow-up after AOD ED discharges.
4. 30-day follow-up after AOD ED discharges.

As an early step of testing we assessed three numerator options to define follow-up care including: (1) an outpatient follow-up visit for any diagnosis; (2) an outpatient follow-up visit for either a primary mental health or AOD diagnosis; or (3) an outpatient follow-up visit that required a primary mental health diagnosis for MH ED discharges or a primary AOD diagnosis for AOD ED discharges. After selecting a numerator, we

conducted a chi-square test to assess the statistical significance of differences in performance between states in the lowest performance quartile versus states in the highest performance quartile.

Average performance for numerator options. Comparing the numerator options, allowing for any diagnosis to count toward the numerator yielded the highest average performance across states.

Seven-day and 30-day follow-up rates for the mental health denominator were roughly 10 percentage points higher when any diagnosis counted toward the numerator compared with when only a primary mental health diagnosis counted toward the numerator (Table VI.5). On average across states, nearly 90 percent of individuals with a MH ED discharge had a follow-up visit with any diagnosis within 30 days whereas 78 percent had a follow-up visit with a primary mental health diagnosis within 30 days.

TABLE VI.5. Performance of Follow-up for MH ED Measure by Numerator Options										
Measure	# States Included	Min	Max	10 th Percentile	25 th Percentile	Median	Mean	75 th Percentile	90 th Percentile	IQR
Numerator 1: Follow-up Visit with Any Diagnosis										
7-day follow-up	16	52.3	91.1	57.5	71.2	81.3	77.1	84.8	88.2	13.6
30-day follow-up	16	77.2	95.9	79.6	87.2	91.8	89.7	92.9	95.2	5.7
Numerator 2: Follow-up Visit with a Primary Diagnosis of Mental Health or AOD										
7-day follow-up	16	35.7	89.7	43.5	61.2	74.3	69.7	80.3	81.2	19.1
30-day follow-up	16	54.4	92.7	62.5	76.7	83.0	79.5	85.6	87.1	8.9
Numerator 3: Follow-up Visit with a Primary Diagnosis of Mental Health (final specification)										
7-day follow-up	16	35.4	89.4	42.2	59.5	73.8	68.9	79.5	81.0	20.0
30-day follow-up	16	53.8	92.4	59.9	75.3	81.8	78.3	84.8	86.0	9.5
SOURCE: MAX 2008.										
NOTES: Follow-up defined as outpatient visit, intensive outpatient encounter, or partial hospitalization for their mental disorder after discharge. The mean presented here is the simple average of performance at the state level.										

Likewise, 7-day follow-up rates were about 10 percentage points higher for the AOD denominator when any diagnosis counted toward the numerator compared with the other two numerator options; 30-day follow-up rates for the AOD denominator were on average 20 points higher when the numerator allowed for any diagnosis versus a primary AOD diagnosis (Table VI.6). That is, on average across states, 86 percent of AOD ED visits had a follow-up visit with any diagnosis within 30 days whereas only 65 percent had a follow-up visit with a primary AOD diagnosis within 30 days.

TABLE VI.6. Performance of Follow-up for AOD ED Measure by Numerator Options										
Measure	# States Included	Min	Max	10 th Percentile	25 th Percentile	Median	Mean	75 th Percentile	90 th Percentile	IQR
Numerator 1: Follow-up Visit with Any Diagnosis										
7-day follow-up	15	35.7	93.2	53.1	63.1	79.2	74.1	85.8	87.8	22.7
30-day follow-up	15	65.5	96.0	72.6	77.8	89.9	85.7	92.8	94.4	15.0
Numerator 2: Follow-up Visit with a Diagnosis of AOD or Mental Health										
7-day follow-up	15	20.4	91.2	33.5	54.1	73.4	66.1	81.3	83.9	27.2
30-day follow-up	15	35.2	92.4	43.1	59.9	79.5	71.8	84.0	87.5	24.1
Numerator 3: Follow-up Visit with a Primary Diagnosis of AOD (final specification)										
7-day follow-up	15	15.5	90.3	21.8	49.5	68.5	62.2	80.3	83.2	30.8
30-day follow-up	15	26.8	90.3	28.7	52.2	70.6	64.7	80.8	83.9	28.5
SOURCE: MAX 2008.										
NOTES: Follow-up defined as outpatient visit, intensive outpatient encounter, or partial hospitalization for their mental disorder after discharge. The mean presented here is the simple average of performance at the state level.										

Distribution of performance for numerator options. With some exceptions, all the numerator options demonstrated wide variation across states; requiring a primary mental health diagnosis at follow-up for MH ED visits or a primary AOD diagnosis at follow-up for AOD ED visits -- numerator 3 -- consistently demonstrated the largest IQR or distribution across states relative to the other numerator options. The distribution of the measure performance across states suggests gaps in performance and room for improvement on the measure.

After selecting numerator 3, we identified groups of low-performing and high-performing states based on measure performance for each of the four follow-up rates: low-performing states scored in the bottom 25th percentile and high-performing states scored in the top 75th percentile. For each follow-up rate, there were four high-performing states and four low-performing states, with the exception of one high-performing group (for 30-day follow-up after AOD ED discharges) which included three states. We then conducted a series of chi-square tests to assess the statistical significance of differences between every possible pair of states in the low-performing versus high-performing groups. For all four follow-up rates associated with numerator 3, the variation across states is statistically significant (chi-square < 0.05).

TABLE VI.7. Follow-up after Emergency Department Performance Rates by State				
State	Follow-up After MH ED Discharges¹		Follow-up After AOD ED Discharges²	
	7-day Follow-up	30-day Follow-up	7-day Follow-up	30-day Follow-up
All States	66.0	76.1	64.3	66.7
AK*	80.5	86.0	53.3	55.2
AL*	74.4	81.3	80.3	80.8
CT	70.9	80.4	68.5	71.7
DC**	56.9	66.3	---	---
GA	89.4	92.4	90.3	90.3
IL*	42.2	59.9	15.5	26.8
IN	78.5	85.5	67.7	69.6
KY*	35.4	53.8	32.8	34.1
LA*	81.0	84.0	82.4	82.5
MN	73.2	84.1	66.5	69.1
MS*	80.9	85.9	83.2	83.9
NC*	77.2	83.6	78.0	79.5
NH*	58.9	77.0	21.8	28.7
OK*	75.0	82.3	74.3	75.1
WI*	60.1	73.9	49.5	52.2
WV*	67.3	76.7	69.5	70.6

SOURCE: MAX 2008.

NOTES: The all-state rate is a pooled average across discharges from all states. Performance in this table was calculated using numerator option 3: follow-up visits with a primary diagnosis of mental health (for MH ED discharges) or with a primary diagnosis of AOD (for AOD ED discharges).

1. Numerator is primary mental health diagnosis at follow-up.
2. Numerator is primary AOD diagnosis at follow-up.
* Includes beneficiaries in the following Medicaid eligibility categories: Disabled, Adult, non-disabled and those age 18 and older in the Child eligibility categories. (One exception: Kentucky includes Disabled and Adult, non-disabled eligibility categories only.) States without asterisks include only adults in the Disabled eligibility category. Included eligibility categories are based on the completeness of data related to HMO and MBHO enrollment, described earlier in this report. The inclusion of a broader group of Medicaid eligibility categories was not systematically associated with state-level measure performance rates.
** The District of Columbia was not included in the analysis of follow-up after AOD ED discharges due to a sample size of less than 150.

In sum, we found that requiring a primary mental health diagnosis at follow-up for MH ED visits or primary AOD diagnosis at follow-up for AOD ED visits best differentiated states and left the most room for improvement relative to the other numerator options. Therefore, we used this numerator specification (numerator 3) for subsequent testing of reliability and validity.

State follow-up rates. The average performance rates for all four variations of the measure suggest that there are opportunities to improve on the measure (Table VI.7). When beneficiaries are pooled across states, an average of 66 percent receive follow-up care within 7 days following MH ED discharges (ranging 35.4-89.4 percent across states) while 76.1 percent received follow-up within 30 days (ranging 53.8-92.4 percent across states). The average follow-up rate within 7 days after emergency department discharge among the AOD denominator was 66.6 percent (range = 15.5-90.3 percent across states) and 30-day follow-up rates averaged 68.7 percent (range = 26.8-90.3 percent across states).

D. Follow-up after Emergency Department by Beneficiary Characteristics

The bivariate results suggest that performance varies by age. For both MH ED and AOD ED discharges, 7-day follow-up rates were lower for adults 21-44 years relative to other age groups (Table VI.9). This pattern was also true of 30-day AOD follow-up rates. For 30-day mental health follow-up rates there are variations by age, but the lowest follow-up rates are among those 65-74 years old.

The bivariate results also suggest that performance varies by diagnosis. Among MH ED discharges, beneficiaries with major depression had the lowest 7-day follow-up rates (59.1 percent) and 30-day rates (72.0 percent). Among AOD ED discharges, beneficiaries with drug use disorders had the lowest 7-day follow-up rates (60.7 percent) and 30-day follow-up rates (63.4 percent) compared with beneficiaries with alcohol use disorders. Although performance rates between racial and ethnic groups are statistically different, these differences tend to be small and therefore difficult to interpret meaningfully. Follow-up rates did not appear associated with beneficiary gender.

TABLE VI.8. 7-day and 30-day Follow-up Rates after Mental Health Discharge from the Emergency Department, by Patient Characteristics				
Characteristic	7-day Follow-up Rate		30-day Follow-up Rate	
	Avg Rate	p-value	Avg Rate	p-value
All Eligible Discharges (regardless of characteristics)	66.0	---	76.1	---
Gender	---	0.077	---	0.283
Male	66.5	---	76.4	---
Female	65.6	---	75.9	---
Age	---	<0.001	---	<0.001
18 - 20	71.9	---	79.8	---
21 - 44	64.9	---	75.6	---
45 - 64	66.5	---	76.2	---
65 - 74	69.3	---	74.3	---
75 - 84	82.4	---	88.2	---
Race/Ethnicity	---	<0.001	---	<0.001
African American	68.2	---	76.6	---
Caucasian	64.8	---	75.8	---
Hispanic	62.1	---	73.3	---
Other	73.6	---	83.4	---
Unknown	62.8	---	75.3	---
Eligibility Category	---	<0.001	---	<0.001
Adult	56.4	---	68.3	---
Disabled	67.3	---	77.3	---
Children 18+*	72.3	---	78.6	---
Geography	---	<0.001	---	<0.001
Metropolitan	69.9	---	78.7	---
Micropolitan	62.6	---	73.3	---
Other	63.6	---	75.2	---
Unknown**	83.6	---	86.4	---
Most Common Diagnosis Associated with Emergency Department Discharge	---	<0.001	---	<0.001
Schizophrenia	69.8	---	81.0	---
Other depression	65.8	---	73.7	---
Bipolar	60.7	---	74.5	---
Non-organic psychosis	78.6	---	84.4	---
Major depression	59.1	---	72.0	---
SOURCE: MAX 2008.				
NOTES: The average rate presented here is a pooled average across states. Performance in this table was calculated using numerator option 3: follow-up visits with a primary diagnosis of mental health.				
* Some beneficiaries age 18 and older remain in the “child” eligibility category after their 18th birthday. All beneficiaries included in these analyses were age 18 or older.				
** The unknown geography category makes up less than 1 percent of the denominator.				

Performance for both mental health and AOD discharges varied by the geographic area in which the beneficiary lived. Among MH ED discharges, beneficiaries living in micropolitan areas had the lowest 7-day follow-up rates (62.6 percent) and 30-day follow-up rates (73.3 percent) relative to metropolitan areas, or areas that were neither metropolitan nor micropolitan (“other”). The same was true of AOD discharges living in micropolitan areas but at slightly lower rates of 7-day follow-up (61.3 percent) and 30-day follow-up (64.9 percent). Performance appears to be highest among discharges whose type of geographic setting is “unknown.” However, this group is very small and makes up less than 1 percent of the mental health and AOD denominators.

TABLE VI.9. 7-day and 30-day Follow-up after AOD Discharge from the Emergency Department, by Patient Characteristics				
Characteristic	7-day Follow-up Rate		30-day Follow-up Rate	
	Avg Rate	p-value	Avg Rate	p-value
All Eligible Discharges (regardless of characteristics)	64.3	---	66.7	---
Gender	---	0.955	---	0.649
Male	64.3	---	66.5	---
Female	64.3	---	66.9	---
Age	---	<0.001	---	<0.001
18 - 20	71.4	---	72.3	---
21 - 44	62.3	---	64.8	---
45 - 64	65.5	---	67.9	---
65 - 74	65.2	---	65.2	---
75 - 84	71.4	---	71.4	---
Race/Ethnicity	---	<0.001	---	<0.001
African American	69.5	---	71.8	---
Caucasian	62.4	---	64.9	---
Hispanic	62.0	---	66.6	---
Other	66.7	---	67.4	---
Unknown	57.9	---	60.1	---
Eligibility Category	---	<0.001	---	0.001
Adult	61.2	---	65.2	---
Disabled	64.6	---	66.7	---
Children 18+	75.4	---	76.1	---
Geography	---	<0.001	---	0.001
Metropolitan	66.7	---	68.5	---
Micropolitan	61.3	---	64.9	---
Other	63.6	---	65.7	---
Unknown	73.3	---	73.3	---
AOD Diagnoses Associated with Emergency Department Discharge*	---	<0.001	---	<0.001
Alcohol abuse and dependence	66.9	---	69.0	---
Drug abuse and dependence	60.7	---	63.4	---

NOTES: The average rate presented here is a pooled average across states. Rates here are based on numerator option 3: follow-up visits with a primary diagnosis of AOD.

* Some beneficiaries age 18 and older remain in the “child” eligibility category after their 18th birthday. All beneficiaries included in these analyses were age 18 or older.

** The unknown geography category makes up less than 1 percent of the denominator.

E. Relationship Between Follow-up after Emergency Department and Inpatient Stays

To evaluate the hypothesized relationship between state-level performance on our measure and the state-level rate of inpatient stays, we conducted a random effects logistic regression that modeled the odds of an inpatient stay among all Medicaid beneficiaries as a function of whether that beneficiary resided in a state that scored within the highest quartile of the follow-up after emergency department measure versus the lowest quartile. The logistic regression included a random effect to account for clustering of beneficiaries within states. We then calculated the regression-adjusted hospitalization rate for each group of states using the inverse logit of the logistic regression coefficient (rates are presented in Table VI.10).

TABLE VI.10. Relationship Between Follow-up after Emergency Department Measure Performance and Inpatient Stays				
States Sorted According to Performance on:	Regression-adjusted Percentage of Beneficiaries with Inpatient Stay for Mental Health Diagnosis		Regression-adjusted Percentage of Beneficiaries with Inpatient Stay for AOD Diagnosis	
	States in Bottom 25% of FU ED	States in Top 25% of FU ED	States in Bottom 25% of FU ED	States in Top 25% of FU ED
Mental Health Follow-up				
7-day follow-up	1.95	1.58	0.34	0.28
30-day follow-up	1.87	1.64	0.37	0.30
AOD Follow-up				
7-day follow-up	1.54	1.62	0.32	0.35
30-day follow-up	1.54	1.69	0.32	0.33
SOURCE: MAX 2008.				
NOTE: Rates of inpatient hospitalization were not statistically different between high and low-performing states.				

We found that the state-level hospitalization rates were not statistically different between states that perform well on this measure versus states that perform poorly. For example, as illustrated in Table V1.10, 1.95 percent of Medicaid beneficiaries had an inpatient mental health stay in states that scored within the bottom 25th percentile of the 7-day follow-up emergency department (FU ED) measure compared with an inpatient stay rate of 1.58 percent for beneficiaries in states that scored within the top 25th percentile of the measure, after adjusting for clustering of beneficiaries within state -- but this difference was not statistically significant. The rate of AOD inpatient stays was also not significantly different between states that scored within the 25th percentile of the FU ED measure versus the top 25th percentile. The lack of statistical significance may reflect the small variation in inpatient stay rates across these states, when the rate is calculated among the entire Medicaid population.

F. Reliability of Follow-up after Emergency Department Measure

To assess measure precision, or reliability, in the context of the observed variability across states, we calculated signal to noise reliability using the beta-binomial (Adams 2009). To calculate a measure-level reliability score, we first determined state-specific reliability scores and then calculated a single, simple average. Reliability scores fall between 0.0 and 1.0. A score of zero indicates that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 indicates that all variation is attributable to real differences in performance across entities. Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between entities. The testing indicates that all four follow-up rates reported as part of this measure have strong reliability between 0.98 and 0.99, well above the threshold of 0.7.

G. Stakeholder Feedback on Follow-up after Emergency Department Measure

The TEP and stakeholder focus groups strongly supported the measure. Approximately 72 percent of public comments supported the measure (for both AOD and mental health populations) as specified or with modifications. Although some stakeholders raised concerns about the need for and feasibility of follow-up within 7 days of an emergency department discharge, the TEP concluded that the measure is actionable and the performance rates suggested that rapid follow-up was possible. The TEP members considered alternative specifications for the numerator and the consensus was that restricting the numerator to visits for primary diagnosis matching the original emergency department visit but allowing visits in any outpatient setting to count was reasonable.

H. Final Outcome of Testing: Revisions to Measure Specification and National Quality Forum Submission

Our analysis suggested that this measure is useful to monitor timely follow-up after discharge from the emergency department for mental health or AOD conditions. There was good variation in measure performance across states and strong stakeholder support. Based on our testing and stakeholder feedback, the measure specification submitted to NQF required that emergency department visits have a primary mental health or AOD diagnosis to define the denominator. The final numerator of the measure required primary mental health diagnosis at follow-up for MH ED discharges and a primary AOD diagnosis at follow-up for AOD ED discharges. Although our validity test did not demonstrate a strong relationship between FU ED rates and state-level inpatient stays, in the absence of strong literature, the assumptions of this relationship may not be true (Table VI.11). Nonetheless, the TEP and most stakeholders supported the validity of the specifications and measure performance. NQF-endorsed the measure on March 6, 2015.

TABLE VI.11. Summary of Testing and Stakeholder Feedback for Follow-up after Emergency Department Measure				
Measure	Testing Results	Concerns from Testing and/or Stakeholders	Revisions to Specification	NQF Submission Decision
Follow-up after emergency department	Good variation across states Strong stakeholder support	7-day follow-up may be unnecessary Health plans have difficulty communicating with emergency departments	Limited specification to primary diagnosis at emergency department discharge and at follow-up visit; required that follow-up visit have same category of diagnosis as emergency department discharge (for example, mental health follow-up for MH ED discharges)	Submitted for mental health and AOD population

VII. 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 a measure that would be suitable for NQF submission due to insufficient evidence or challenges identifying an entity accountable for 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 MH/SA 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.

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 EHR incentive program and Medicaid quality reporting) and take advantage of existing data sources and the evolving infrastructure for measurement.

REFERENCES

- Agency for Healthcare Research and Quality. "SNAC-Recommended Initial Core Set: September 18, 2009." Rockville, MD: AHRQ, January 2010. Available at <http://www.ahrq.gov/chipra/coreset/coreset.htm>. Accessed September 9, 2014.
- Agency for Healthcare Research and Quality, National Advisory Council Subcommittee. "Identifying Quality Measures for Medicaid-Eligible Adults: Background Report." Rockville, MD: AHRQ, December 2010. Available at <http://www.ahrq.gov/about/nacqm/>. Accessed September 9, 2014.
- Borck, R., A. Dodd, A. Zlatinov, A. Verghese, R. Malsberger, and C. Petroski. "The Medicaid Analytic eXtract 2008 Chartbook." Washington, DC: CMS, 2012.
- Byrd, V., and A. Dodd. "Assessing the Usability of MAX 2008 Encounter Data for Comprehensive Managed Care." *Medicare and Medicaid Research Review*, Vol. 3, No. 1, 2013, pp. E1-E19, doi/10.5600/mmrr.003.01.b01.
- Brown, J., S. Scholle, and M. Azur. "Strategies for Measuring the Quality of Psychotherapy." White paper prepared for the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. Washington, DC: Mathematica Policy Research, February 25, 2013. Available at <https://aspe.hhs.gov/report/strategies-measuring-quality-psychotherapy-white-paper-inform-measure-development-and-implementation>.
- Centers for Medicare & Medicaid Services. "2014 Clinical Quality Measures Adult Recommended Core Measures." 2014. Available at http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/2014_CQM_AdultRecommend_CoreSetTable.pdf. Accessed September 9, 2014.
- Fisher, C., B. Spaeth-Rublee, and H. Pincus for the IIMHL Clinical Leaders Group. "Developing Mental Health-Care Quality Indicators: Toward a Common Framework." *International Journal for Quality in Health Care*, November 21, 2012, pp. 1-6.
- Health IT Policy Committee Quality Measures Work Group. Transmittal letter to Farzad Mostashari, M.D., Sc.M., National Coordinator for Health Information Technology, August 5, 2011.
- Institute of Medicine (IOM), Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders. "Improving the Quality of Health Care for Mental and Substance-Use Conditions." Washington, DC: National Academies Press, 2006.
- Landis J.R., and G.G. Koch. "The Measurement of Observer Agreement for Categorical Data." *Biometrics*, Vol. 33, 1977, pp. 159-174.

- National Committee for Quality Assurance. "Improving Quality and Patient Experience: The State of Health Care Quality 2013." Washington, DC: National Committee for Quality Assurance, 2013. Available at http://www.ncqa.org/Portals/0/Newsroom/SOHC/2013/SOHC-web_version_report.pdf. Accessed September 9, 2014.
- National Committee for Quality Assurance. "State of Health Care Quality 2010." Washington, DC: NCQA, 2010.
- National Quality Forum. "National Voluntary Consensus Standards for Patient Outcomes: A Consensus Report." Washington, DC: NQF, 2011.
- Nysenbaum, J., E. Bouchery, and R. Malsberger.. "Availability and Usability of Behavioral Health Organization Encounter Data in MAX 2009." *Medicare and Medicaid Research Review*, Vol. 4, No. 2, 2014, E1-E12, doi/10.5600/mmrr.004.02.b02
- Pincus, H., B. Spaeth-Rublee, and K. Watkins. "The Case for Measuring Quality in Mental Health and Substance Abuse Care." *Health Affairs*, Vol. 30, No. 4, 2011, pp. 730-736.
- Substance Abuse and Mental Health Services Administration. "Leading Change: A Plan for SAMHSA's Roles and Actions 2011-2014, Executive Summary and Introduction." HHS Publication No. (SMA) 11-4629 Summary. Rockville, MD: SAMHSA, 2011.

APPENDIX A. TECHNICAL EXPERT PANEL MEMBERS

TABLE A.1. TEP Members	
Stakeholder Group	TEP Members
Consumer and Family Representatives	<ul style="list-style-type: none"> • Keris Myrick, Project Return Support Network/NAMI Board of Directors • Jonathan Delman, Transitions Research and Training Center
Medicaid	<ul style="list-style-type: none"> • Jeff Thompson, Washington State Medicaid • David Kelley, Pennsylvania Department of Public Welfare
State Mental Health and Substance Abuse	<ul style="list-style-type: none"> • Michael Hogan, New York State Office of Mental Health • Judy Mohr Peterson, Oregon Health Authority • Kevin Huckshorn, Delaware Department of Health and Social Services • Renata Henry, Maryland Department of Health and Mental Hygiene
Providers	<ul style="list-style-type: none"> • Frank Ghinassi, Western Psychiatric Institute and Clinic • Kathleen McCann, National Association of Psychiatric Health Systems • Neil Korsen, Mental Health Integration Program
Health Plans	<ul style="list-style-type: none"> • Francisca Azocar, OptumHealth Behavioral Solutions • Dan Rome, Beacon Health Strategies • James Schuster, Community Care Behavioral Health
Research and Performance Measurement	<ul style="list-style-type: none"> • Alisa Busch, McLean Hospital
Federal Representatives	<ul style="list-style-type: none"> • Kirsten Beronio, ASPE • D.E.B. Potter, ASPE/AHRQ • Lisa Patton, Marna Hoard, Danyelle Manniz, Alexander Camacho and Nicholas Reuter, SAMHSA • Charlotte Mullican and Nancy Wilson, AHRQ • Alex Ross and Ian Corbridge, HRSA • Shaheen Halim, CMS • Marc Safran, CDC • Daniel Kivlahan and Ira Katz, VA

APPENDIX B. MEASURE SPECIFICATIONS

TABLE B.1. Specifications of Parent Measures and New Measures Submitted to NQF		
	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Preventive Care and Screening: Unhealthy Alcohol Use: Screening and Brief Counseling (2152)	Alcohol Screening and Follow-up for People with SMI (2599)
Level of Reporting	Clinician/Provider	Health Plan
Data Source	Provider report using G-codes or e-measures	Administrative claims, Electronic Clinical Data, Paper Medical Records
Numerator	Patients who were screened for unhealthy alcohol use at least once within 24 months using a systematic screening method AND who received brief counseling if identified as an unhealthy alcohol user.	Patients 18 years and older who are screened for unhealthy alcohol use during the last 3 months of the year prior to the measurement year through the first 9 months of the measurement year and received 2 events of counseling if identified as an unhealthy alcohol user.
Denominator	All patients aged 18 years and older.	All patients 18 years of age or older as of December 31 of the measurement year with at least 1 inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year.
Exclusions	Documentation of medical reasons for not screening for unhealthy alcohol use (for example, limited life expectancy, other medical reasons)	Active diagnosis of alcohol abuse or dependence during the first 9 months of the year prior to the measurement year.
	Preventive Care & Screening: Tobacco Use: Screening & Cessation Intervention (0028)	Tobacco Use Screening and Follow-up for People with SMI or AOD (2600)
Level of Reporting	Clinician/Provider	Health Plan
Data Source	Provider report using G-codes or e-measures	Administrative claims, Electronic Clinical Data, Paper Medical Records
Numerator	Patients who were screened for tobacco use at least once within 24 months AND who received tobacco cessation counseling intervention (brief counseling or pharmacotherapy) if identified as a tobacco user.	SMI: Screening for tobacco use in patients with SMI during the measurement year or year prior to the measurement year and received follow-up care if identified as a current tobacco user. AOD: Screening for tobacco use in patients with AOD during the measurement year or year prior to the measurement year and received follow-up care if identified as a current tobacco user.
Denominator	All patients aged 18 years and older.	SMI: All patients 18 years of age or older as of December 31 of the measurement year with at least 1 inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year. AOD: All patients 18 years of age or older as of December 31 of the measurement year with any diagnosis of AOD during the measurement year.
Exclusions	Documentation of medical reason(s) for not screening for tobacco use (for example, limited life expectancy, other medical reason)	None

TABLE B.1 (continued)

	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Preventive Care and Screening: BMI Screening and Follow-Up (0421)	BMI Screening and Follow-up for People with SMI (2601)
Level of Reporting	Clinician/Provider	Health Plan
Data Source	Provider report using G-codes or e-measures	Administrative claims, Electronic Clinical Data, Paper Medical Records
Numerator	<p>Patients with BMI calculated within the past 6 months or during the current visit, and a follow-up plan documented within the past 6 months or during the current visit if the BMI is outside of normal parameters.</p> <p>Follow-up:</p> <ol style="list-style-type: none"> 1) Documentation of a future appointment 2) Education 3) Referral 4) Pharmacological interventions 5) Dietary supplements for people with low BMI 6) Exercise counseling 7) Nutrition counseling 	<p>Patients 18 years and older with calculated BMI documented during the first 9 months of the measurement year or year prior to the measurement year and follow-up care is provided if a person's BMI is greater than or equal to 30kg/m2.</p> <p>Follow-Up:</p> <p>Follow-up documented within 3 months of screening for patients with a BMI greater than or equal to 30kg/m2:</p> <ul style="list-style-type: none"> - Two events of counseling, on different dates, for weight management (such as nutrition or exercise counseling) with the provider who did the screening or another provider including health plan clinical case managers, or - One event of counseling and 1 fill of medication (Orlistat) for weight management.
Denominator	All patients aged 18 years and older.	All patients 18 years of age or older as of December 31 of the measurement year with at least 1 inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year.
Exclusions	<ul style="list-style-type: none"> - Patient is receiving palliative care, pregnant or refuses BMI measurement. - Other reason documented in the medical record by the provider explaining why BMI measurement or follow-up plan was not appropriate. - Patient is in an urgent or emergent medical situation where time is of the essence and to delay treatment would jeopardize the patient's health status. 	Active diagnosis of pregnancy during the measurement year or the year prior to the measurement year.

TABLE B.1 (continued)

	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Controlling High Blood Pressure (0018)	Controlling High Blood Pressure for People with SMI (2602)
Level of Reporting	Health Plan	Health Plan
Data Source	Hybrid (claims plus medical record)	Administrative claims, Electronic Clinical Data, Paper Medical Records
Numerator	<p>Patients whose most recent blood pressure is adequately controlled during the measurement year (after the diagnosis of hypertension) based on the following criteria:</p> <ul style="list-style-type: none"> - Patients 18-59 years of age as of December 31 of the measurement year whose blood pressure was <140/90mm/Hg. - Patients 60-85 years of age as of December 31 of the measurement year and flagged with a diagnosis of diabetes whose blood pressure was <140/90mm/Hg. - Patients 60-85 years of age as of December 31 of the measurement year and flagged as not having a diagnosis of diabetes whose blood pressure was <150/90mm/Hg. 	The same as the parent measure.
Denominator	Patients 18-85 years of age by the end of the measurement year who had at least 1 outpatient encounter with a diagnosis of hypertension during the first 6 months of the measurement year and continuously enrolled during the measurement year.	All patients 18-85 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND a diagnosis of hypertension on or before June 30 of the measurement year.
Exclusions	<ul style="list-style-type: none"> - Individuals with evidence of end-stage renal disease. - Diagnosis of pregnancy. - Individuals who had an admission to a non-acute inpatient setting during the measurement year. 	<p>All patients who meet 1 or more of the following criteria should be excluded from the measure:</p> <ul style="list-style-type: none"> - Evidence of end-stage renal disease or kidney transplant - A diagnosis of pregnancy

TABLE B.1 (continued)		
	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Comprehensive Diabetes Care: HbA1c testing (0057)	Comprehensive Diabetes Care for People with SMI: HbA1c Testing (2603)
Level of Reporting	Health Plan	Health Plan
Data Source	Hybrid (claims plus medical record)	Same as Parent Measure
Numerator	Patients who had an HbA1c test performed during the measurement year.	Same as Parent Measure
Denominator	Adults aged 18-75 years of age with diabetes and continuously enrolled during the measurement year.	Patients 18-75 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND diabetes (type 1 and type 2) during the measurement year or year before.
Exclusions	(Optional) Patients who do not have a diagnosis of diabetes and meet 1 of the following criteria are excluded from the measure: <ul style="list-style-type: none"> – Patients with a diagnosis of polycystic ovaries who did not have a face-to-face encounter in any setting. – Patients with gestational or steroid-induced diabetes who did not have a face-to-face encounter in any setting. 	Same as parent measure
	Comprehensive Diabetes Care: Medical Attention for Nephropathy (0062)	Comprehensive Diabetes Care for People with SMI: Medical Attention to Nephropathy (2604)
Level of Reporting	Health Plan	Health Plan
Data Source	Hybrid (claims plus medical record)	Same as Parent Measure
Numerator	Patients who received a nephropathy screening test or had evidence of nephropathy during the measurement year.	Same as Parent Measure
Denominator	Adults aged 18-75 years of age with diabetes and continuously enrolled during the measurement year.	Patients 18-75 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND diabetes (type 1 and type 2) during the measurement year or year before.
Exclusions	(Optional) Patients who do not have a diagnosis of diabetes and meet 1 of the following criteria are excluded from the measure: <ul style="list-style-type: none"> – Patients with a diagnosis of polycystic ovaries who did not have a face-to-face encounter in any setting. – Patients with gestational or steroid-induced diabetes who did not have a face-to-face encounter in any setting. 	Same as parent measure

TABLE B.1 (continued)

	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Comprehensive Diabetes Care: Blood Pressure Control (<140/90mm/Hg) (061)	Comprehensive Diabetes Care for People with SMI: Blood Pressure Control (<140/90mm/Hg) (2606)
Level of Reporting	Health Plan (Composite measure NQF #0731 & Individual Measure #0061)	Health Plan
Data Source	"Hybrid" claims plus medical record review	Same as Parent Measure
Numerator	Patient whose most recent blood pressure screening result is <140/90mm/Hg during the measurement year.	Same as Parent Measure
Denominator	Adults aged 18-75 years of age with diabetes and continuously enrolled during the measurement year.	All patients 18-75 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND diabetes (type 1 and type 2) during the measurement year or year prior to the measurement year.
Exclusions	(Optional) Patients who do not have a diagnosis of diabetes and meet 1 of the following criteria are excluded from the measure: <ul style="list-style-type: none"> - Patients with a diagnosis of polycystic ovaries who did not have a face-to-face encounter in any setting. - Patients with gestational or steroid-induced diabetes who did not have a face-to-face encounter in any setting. 	Same as parent measure
	Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%) (0061)	Comprehensive Diabetes Care for People with SMI: HbA1c Poor Control (>9.0%) (2607)
Level of Reporting	Health Plan	Health Plan
Data Source	Hybrid (claims plus medical record)	Same as Parent Measure
Numerator	Patients whose most recent HbA1c level is greater than 9.0% or is missing a result, or for whom an HbA1c test was not done during the measurement year.	Same as Parent Measure
Denominator	Adults aged 18-75 years of age with diabetes and continuously enrolled during the measurement year.	Patients 18-75 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND diabetes (type 1 and type 2) during the measurement year or the year before.
Exclusions	(Optional) Patients who do not have a diagnosis of diabetes and meet 1 of the following criteria are excluded from the measure: <ul style="list-style-type: none"> - Patients with a diagnosis of polycystic ovaries who did not have a face-to-face encounter in any setting. - Patients with gestational or steroid-induced diabetes who did not have a face-to-face encounter in any setting. 	Same as parent measure

TABLE B.1 (continued)		
	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Comprehensive Diabetes Care: HbA1c Control (<8.0%) (0575)	Comprehensive Diabetes Care for People with SMI: HbA1c Control (<8.0%) (2608)
Level of Reporting	Health Plan	Health Plan
Data Source	Hybrid (claims plus medical record)	Same as Parent Measure
Numerator	Patients whose most recent HbA1c level is less than 8.0% during the measurement year.	Same as Parent Measure
Denominator	Adults aged 18-75 years of age with diabetes and continuously enrolled during the measurement year.	Patients 18-75 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND diabetes (type 1 and type 2) during the measurement year or the year before.
Exclusions	(Optional) Patients who do not have a diagnosis of diabetes and meet 1 of the following criteria are excluded from the measure: <ul style="list-style-type: none"> – Patients with a diagnosis of polycystic ovaries who did not have a face-to-face encounter in any setting. – Patients with gestational or steroid-induced diabetes who did not have a face-to-face encounter in any setting. 	Same as parent measure

TABLE B.1 (continued)		
	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Comprehensive Diabetes Care: Eye Exam (0055)	Comprehensive Diabetes Care for People with SMI: Eye Exam (2609)
Level of Reporting	Health Plan	Health Plan
Data Source	"Hybrid" claims plus medical record review	Same as Parent Measure
Numerator	<p>Patients who received an eye screening for diabetic retinal disease.</p> <p>This includes people with diabetes who had the following:</p> <ul style="list-style-type: none"> - A retinal or dilated eye exam by an eye care professional (optometrists or ophthalmologist) in the measurement year; OR - A negative retinal exam or dilated eye exam (negative for retinopathy) by an eye care professional in the year prior to the measurement year. For exams performed in the year prior to the measurement year, a result must be available. 	Patients who received an eye exam during the measurement year.
Denominator	Adults aged 18-75 years of age with diabetes and continuously enrolled during the measurement year.	Patients 18-75 years of age as of December 31 of the measurement year with at least 1 acute inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year AND diabetes (type 1 and type 2) during the measurement year or the year before.
Exclusions	<p>(Optional)</p> <p>Patients who do not have a diagnosis of diabetes and meet 1 of the following criteria are excluded from the measure:</p> <ul style="list-style-type: none"> - Patients with a diagnosis of polycystic ovaries who did not have a face-to-face encounter in any setting. - Patients with gestational or steroid-induced diabetes who did not have a face-to-face encounter in any setting. 	Same as parent measure

TABLE B.1 (continued)		
	Parent Measure Specification (NQF number)	New Measure Specification (NQF number assigned for review; measures are currently under review)
	Follow-Up After Hospitalization for Mental Illness (0576)	Follow-Up After Emergency Department Use for Mental Health Conditions or AOD (2605)
Level of Reporting	Health Plan	Health Plan
Data Source	Administrative Claims	Same as Parent Measure
Numerator	<p>7-day: An outpatient visit, intensive outpatient encounter or partial hospitalization with a mental health practitioner within 7 days of discharge.</p> <p>30-day: An outpatient visit, intensive outpatient encounter or partial hospitalization with a mental health practitioner within 30 days of discharge.</p>	<p>The numerator for each denominator population consists of 2 rates:</p> <p>Mental Health:</p> <ul style="list-style-type: none"> – 7-day: An outpatient visit, intensive outpatient encounter or partial hospitalization with any provider with a primary diagnosis of mental health within 7 days after emergency department discharge – 30-day: An outpatient visit, intensive outpatient encounter or partial hospitalization with any provider with a primary diagnosis of mental health within 30 days after emergency department discharge <p>AOD:</p> <ul style="list-style-type: none"> – 7-day: An outpatient visit, intensive outpatient encounter or partial hospitalization with any provider with a primary diagnosis of AOD within 7 days after emergency department discharge – 30-day: An outpatient visit, intensive outpatient encounter or partial hospitalization with any provider with a primary diagnosis of AOD within 30 days after emergency department discharge
Denominator	Discharged alive from an acute inpatient setting (including acute care psychiatric facilities) with a principal mental health diagnosis on or between January 1 and December 1 of the measurement year.	Patients who were treated and discharged from an emergency department with a primary diagnosis of mental health or AOD on or between January 1 and December 1 of the measurement year.
Exclusions	<p>Exclude discharges followed by re-admission or direct transfer to a non-acute facility within the 30-day follow-up period, regardless of principal diagnosis for the re-admission.</p> <p>Exclude discharges followed by re-admission or direct transfer to an acute facility within the 30-day follow-up period if the principal diagnosis was for non-mental health (any principal diagnosis code other than those included in the Mental Health Diagnosis Value Set).</p> <p>These discharges are excluded from the measure because re-hospitalization or transfer may prevent an outpatient follow-up visit from taking place.</p>	<p>If the discharge is followed by re-admission or direct transfer to an emergency department for a principal diagnosis of mental health or AOD within the 30-day follow-up period, count only the re-admission discharge or the discharge from the emergency department to which the patient was transferred.</p> <p>Exclude discharges followed by admission or direct transfer to an acute or non-acute facility within the 30-day follow-up period, regardless of primary diagnosis for the admission.</p> <p>These discharges are excluded from the measure because hospitalization or transfer may prevent an outpatient follow-up visit from taking place.</p>

TABLE B.2. Specifications of Parent Measures and Measures Tested but not Submitted to NQF		
	Parent Measure Specification (NQF number)	New Measure Specification
	Preventive Care and Screening: Screening for High Blood Pressure and Follow-Up Documented (Not NQF-Endorsed)	High Blood Pressure Screening and Follow-up for People with SMI or AOD
Level of Reporting	Clinician/Provider	Health Plan
Data Source	Claims, Registry	Administrative claims, Electronic Clinical Data, Paper Medical Records
Numerator	Patients who were screened for high blood pressure and a recommended follow-up plan is documented as indicated if the blood pressure is pre-hypertensive or hypertensive.	Blood Pressure Screening: Blood pressure screening documented during the measurement year. Blood Pressure Follow-up: A recommended follow-up plan is documented or follow-up care provided if the blood pressure recording is in the pre-hypertensive or hypertensive range.
Denominator	All patients aged 18 years and older.	SMI: All patients 18 years of age or older as of December 31 of the measurement year with at least 1 inpatient visit or 2 outpatient visits for schizophrenia or bipolar I disorder, or at least 1 inpatient visit for major depression during the measurement year. AOD: All patients 18 years of age or older as of December 31 of the measurement year with any diagnosis of AOD during the measurement year.
Exclusions	<ul style="list-style-type: none"> – Patient has active diagnosis of hypertension – Patient refuses blood pressure measurement – Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient’s health status. This may include but is not limited to severely elevated blood pressure when immediate medical treatment is indicated. 	Individuals are not eligible if they have an active diagnosis of hypertension at the first blood pressure screening during the measurement year.

TABLE B.2 (continued)		
	Parent Measure Specification (NQF number)	New Measure Specification
	Preventive Care and Screening: Screening for Clinical Depression (0418)	Clinical Depression Screening and Follow-up for People with AOD
Level of Reporting	Clinician/Provider	Health Plan
Data Source	Administrative claims, Electronic Clinical Data: EHR, Paper Medical Records	Administrative claims, Electronic Clinical Data, Paper Medical Records
Numerator	Patient's screening for clinical depression using an age appropriate standardized tool AND follow-up plan is documented if screened positive.	Depression Screening: Screening for clinical depression during the measurement year using an age appropriate standardized tool. Follow-up: A follow-up plan is documented or follow-up care is provided if an individual screens positive.
Denominator	All patients aged 12 years and older	All patients 18 years of age or older as of December 31 of the measurement year with any diagnosis of AOD during the measurement year.
Exclusions	<ul style="list-style-type: none"> – Patient refuses to participate – Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status – Situations where the patient's motivation to improve may impact the accuracy of results of nationally recognized standardized depression assessment tools. For example: certain court appointed cases – Patient was referred with a diagnosis of depression – Patient has been participating in ongoing treatment with screening of clinical depression in a preceding reporting period – Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example: cases such as delirium or severe cognitive impairment, where depression cannot be accurately assessed through use of nationally recognized standardized depression assessment tools 	Active diagnosis of depression or bipolar disorder at the first depression screening during the measurement year or during the year prior to the measurement year.

DEVELOPMENT AND TESTING OF BEHAVIORAL HEALTH QUALITY MEASURES

Reports Available

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

HTML <https://aspe.hhs.gov/basic-report/development-and-testing-behavioral-health-quality-measures-health-plans-final-report>

PDF <https://aspe.hhs.gov/pdf-report/development-and-testing-behavioral-health-quality-measures-health-plans-final-report>

Development of Quality Measures for Inpatient Psychiatric Facilities: Final Report

HTML <https://aspe.hhs.gov/basic-report/development-quality-measures-inpatient-psychiatric-facilities-final-report>

PDF <https://aspe.hhs.gov/pdf-report/development-quality-measures-inpatient-psychiatric-facilities-final-report>

Review of Medication-Assisted Treatment Guidelines and Measures for Opioid and
Alcohol Use

HTML <https://aspe.hhs.gov/report/review-medication-assisted-treatment-guidelines-and-measures-opioid-and-alcohol-use>

PDF <https://aspe.hhs.gov/pdf-report/review-medication-assisted-treatment-guidelines-and-measures-opioid-and-alcohol-use>

Strategies for Measuring the Quality of Psychotherapy: A White Paper to Inform
Measure Development and Implementation

HTML <https://aspe.hhs.gov/report/strategies-measuring-quality-psychotherapy-white-paper-inform-measure-development-and-implementation>

PDF <https://aspe.hhs.gov/pdf-report/strategies-measuring-quality-psychotherapy-white-paper-inform-measure-development-and-implementation>

To obtain a printed copy of this report, send the full report title and your mailing information to:

U.S. Department of Health and Human Services
Office of Disability, Aging and Long-Term Care Policy
Room 424E, H.H. Humphrey Building
200 Independence Avenue, S.W.
Washington, D.C. 20201
FAX: 202-401-7733
Email: webmaster.DALTCP@hhs.gov

NOTE: All requests must be in writing.

RETURN TO:

Office of Disability, Aging and Long-Term Care Policy (DALTCP) Home
http://aspe.hhs.gov/office_specific/daltcp.cfm

Assistant Secretary for Planning and Evaluation (ASPE) Home
<http://aspe.hhs.gov>

U.S. Department of Health and Human Services (HHS) Home
<http://www.hhs.gov>