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

DEVELOPING QUALITY MEASURES FOR MEDICAID BENEFICIARIES WITH SCHIZOPHRENIA:

FINAL REPORT

January 2012

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 divisions. 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 #HHSP23320095642WC between HHS's ASPE/DALTCP and Mathematica Policy Research. For additional information about this subject, you can visit the DALTCP home page at http://aspe.hhs.gov/office_specific/daltcp.cfm or contact the ASPE Project Officer, Kirsten Beronio, at HHS/ASPE/DALTCP, Room 424E, H.H. Humphrey Building, 200 Independence Avenue, S.W., Washington, D.C. 20201. Her e-mail address is: Kirsten.Beronio@hhs.gov.

DEVELOPING QUALITY MEASURES FOR MEDICAID BENEFICIARIES WITH SCHIZOPHRENIA: Final Report

Sam Simon
Thomas Croghan
Robert C. Saunders
Sarah Hudson Scholle
Milesh M. Patel
Jeremy Gottlich

Mathematica Policy Research

January 30, 2012

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

LIST OF TABLES

iii

ACRONYMS

ACT	assertive community treatment
APA	American Psychiatric Association
ASPE	Office of the Assistant Secretary for Planning and Evaluation
BHO	behavioral healthcare organization
BMI	body mass index
CMS	Centers for Medicare and Medicaid Services
DHHS	New Hampshire Department of Health and Human Services
ED	emergency department
EHR	electronic health record
FFS	fee-for-service
FUH	follow-up after hospitalization
HCPCS	Healthcare Common Procedure Coding System
HbA1c	Hemoglobin A1c
HMO	health maintenance organization
ICSI	Institute for Clinical Systems Improvement
IQR	interquartile range
LAI	long-acting injectable
MAX	Medicaid Analytic eXtract
MBHO	managed behavioral healthcare organization
MMDLN	Medicaid Medical Directors Learning Network
NACBHDD	National Association of County Behavioral Health and Developmental Disability Directors
NAMI	National Alliance on Mental Illness
NCQA	National Committee for Quality Assurance
NICE	National Institute for Health and Clinical Excellence
NQF	National Quality Forum
PCP	primary care provider
PDC	proportion of days covered
PORT	Patient Outcomes Research Team

RCT	randomized controlled trial
SMI	serious mental illness
SPMI	serious and persistent mental illness
TAG	Technical Advisory Group
WFBH	Wake Forest Baptist Health

EXECUTIVE SUMMARY

In August 2010, the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) contracted with Mathematica Policy Research and its subcontractor--the National Committee for Quality Assurance-- to develop evidence-based quality measures to assess the quality of care provided to Medicaid enrollees diagnosed with schizophrenia. The goal of the project was to create a set of claims-based ambulatory care measures that meet National Quality Forum (NQF) criteria for importance, scientific acceptability, usability, and feasibility and would thus be suitable for submission to the NQF for endorsement consideration.

The project began with a review of existing literature and other evidence describing evidence-based practices for people with schizophrenia. Assisted by expert consultants, this effort emphasized the findings of the Schizophrenia Patient Outcomes Research Team and allowed the team to create concepts for new measures that assess the quality of medication management, underuse of evidence-based psychosocial treatments, and access to primary care and preventive health services. Once the measure concepts were vetted by a Technical Advisory Group (TAG), we developed draft specifications and sought comment from measure stakeholders, including representatives from managed behavioral healthcare organizations (MBHOs), Medicaid medical directors, and state mental health directors to assess their perspectives on the importance, scientific acceptability, usability, and feasibility of the proposed measures. After these key stakeholders gave their input, measure specifications were posted for public comment, and they were pilot-tested using Medicaid Analytic eXtract (MAX) data from 2007 and 2008 to further assess their feasibility, reliability, and validity. Throughout the project, the project team received valuable advice and guidance from ASPE, members of the TAG, and our project consultants.

The project team sought to develop measures in three domains, pharmacology, psychosocial care, and physical health, as well as cross-cutting measures that span several of these domains. Based on the review of the literature and feedback from the TAG and ASPE, we developed detailed specifications for an initial set of 17 measure concepts before settling on a final set of ten to be submitted to NQF for endorsement.

Focus groups with state Medicaid and mental health leaders, as well as with MBHO staff, yielded remarkably consistent results. Key points included: (1) claims data are unreliable for identifying some behavioral health services, particularly evidence-based psychosocial treatments; (2) variation in financing of services for people with serious mental illness (SMI) limits the ability to consistently measure the quality of care across Medicaid programs; and (3) some candidate measures address problems that are not unique to patients with schizophrenia--measures could be broadened to include patients with bipolar disorder, schizophrenia, and severe forms of depression. The

feedback from public comment was positive, with 87 percent of the comments either supporting the measures or supporting them with modifications.

Overall, 9.7 percent of Medicaid recipients in our 22-state 2007 MAX dataset had schizophrenia and 12.8 percent had SMI (bipolar disorder and/or schizophrenia). The objective of pilot-testing was to determine the scientific acceptability of each measure to the extent practicable through the use of Medicaid claims data. Five of the ten proposed measures demonstrated significant variability in state-level performance, indicating general utility of the measures. Seven of the ten proposed measures demonstrated evidence of either construct or convergent validity. Construct validity was assessed by examining the association between measure performance and outcomes (schizophrenia-related (1) hospitalization, and (2) emergency department [ED] visits). We reported the percentage of people who were either hospitalized or visited the ED for schizophrenia, comparing the worst and best-performing quartiles of state performance for each measure. Seven measures demonstrated evidence of construct validity, indicated by the association between (higher) measure performance and (lower) rates of adverse events. Convergent validity was determined through enrollee-level measure correlations. Three of the ten measures demonstrated evidence of convergent validity. Nine of the ten measures demonstrated evidence of reliability, assessed between measures calculated during calendar year 2007 and 2008, either through test-retest correlations or relative performance stability over this time period.

Although some of these results are encouraging, important limitations of our findings warrant consideration. First, use of Medicaid claims data as a source to implement and test schizophrenia quality measures limited the number of evidence-based practices that could be implemented as measures. This limitation prevented our ability to develop psychosocial measures. In addition, several topics could not be developed because the evidence base, tools, and methods for tracking these measures are immature. We also found that variation in the financing of services for people with SMI limited our ability to generalize measurement of the care provided by Medicaid programs. For example, the provision of services through state mental health systems, the coverage of mental health services through Medicare for dual-eligible beneficiaries, the prohibition of same-day billing of medical and behavioral health services, and interstate variation in Medicaid and disability standards all underscore the limitations of claims data to measure quality for enrollees with schizophrenia. Finally, the distinction between enrollees with schizophrenia and other SMI conditions is, in many cases, artificial. The project team, ASPE, and measure stakeholders all expressed the belief that conceptually, many issues related to schizophrenia also apply broadly to people with any SMI. Further work is needed to consider whether measures similar to the ones developed and tested under this contract would be relevant for people with bipolar disorder and other SMI.

I. OVERVIEW OF THE PROJECT

Despite enormous expenditures and remarkable breakthroughs in medical treatment, the United States behavioral health care system does not consistently deliver safe and effective treatment to those with serious and persistent mental illness (SPMI), many of whom go untreated or inadequately treated. Now, as the nation stands at the doorstep of fundamental reforms that offer insurance benefits for those without them, remove inequitable treatment limits and financial barriers to mental health treatments, and promote integrated primary and behavioral health care, we have an enormous opportunity to close the gap between the availability of effective treatments and providing them in a manner that promotes recovery. By enhancing transparency, new quality measures that promote feedback to providers and enable value-based purchasing represent an essential tool to achieve the full promise of these reforms.

In August 2010, the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) contracted with Mathematica Policy Research and its subcontractor--the National Committee for Quality Assurance (NCQA)--to develop evidence-based quality measures to assess the quality of care provided to Medicaid enrollees diagnosed with schizophrenia. The goal of the project was to create a set of claims-based ambulatory care measures that meet National Quality Forum (NQF) criteria for importance, scientific acceptability, usability, and feasibility and would thus be suitable for submission to the NQF for endorsement consideration.

The project began with a review of existing literature and other evidence describing evidence-based practices for people with schizophrenia. Assisted by expert consultants, this effort emphasized the findings of the Schizophrenia Patient Outcomes Research Team (PORT) and allowed the team to create concepts for new measures that assess the quality of medication management, underuse of evidence-based psychosocial treatments, and access to primary care and preventive health services. Once the measure concepts were vetted by a Technical Advisory Group (TAG), we developed draft specifications and sought comment from measure stakeholders, including representatives from managed behavioral healthcare organizations (MBHOs), Medicaid medical directors, and state mental health directors to assess their perspectives on the importance, scientific acceptability, usability, and feasibility of the proposed measures. After these key stakeholders gave their input, measure specifications were posted for public comment, and they were pilot-tested using Medicaid Analytic eXtract (MAX) data from 2007 and 2008 to further assess their feasibility, reliability, and validity. Throughout the project, the project team received valuable advice and guidance from ASPE, members of the TAG, and our project consultants.

This report presents a chronology of the process, key findings, and lessons learned during our project to develop claims-based measures of services provided to

Medicaid enrollees with schizophrenia that meet key NQF criteria. Chapter II reviews that process and describes how several findings in our data collection changed the course of measure development. Chapter III summarizes key findings from our field and pilot-testing efforts, and Chapter IV discusses lessons learned that we hope will improve the process of measure development and the quality of the resulting measures. The appendices contain all key documents produced throughout the project, including material presented at each TAG meeting, pilot-testing results, and the candidate measure summary information.

II. THE DEVELOPMENT OF SCHIZOPHRENIA QUALITY MEASURES: A CHRONOLOGY

In developing new quality measures to assess the quality and appropriateness of care for Medicaid enrollees with schizophrenia, Mathematica and NCQA carried out the following tasks under guidance from ASPE:

1. Identified appropriate measure topics and concepts through an environmental scan and a review of the literature.

2. Defined and developed measure specifications.

3. Convened meetings of the project TAG.

4. Field-tested measures with key stakeholders.

5. Posted the measures for public comment.

6. Pilot-tested measures and evaluated the reliability and validity of measures using Medicaid claims data.

1. Environmental Scan: Identify Appropriate Measure Topics and Concepts

The process for identifying the measure concepts included a review of the clinical literature prepared by ASPE, an environmental scan of treatment measure guidelines and existing measures by NCQA, and consultation with experts. We focused on measure concepts in three treatment domains specified by ASPE: pharmacotherapy, psychosocial treatment, and physical health. Drs. Julie Kreyenbuhl and Lisa Dixon, leaders of the Schizophrenia PORT at the University of Maryland School of Medicine, served as content experts and consultants to the project. Their role was to identify potential errors of interpretation, emphasis, inclusion, or omission prior to developing a report that summarized the scientific literature, clinical guidelines, and existing measures that are focused on the population of interest.

The environmental scan identified systematic reviews (e.g., the Schizophrenia PORT reviews), measure specifications, and treatment guidelines and standards developed by professional societies and measurement organizations that relate to care for people with schizophrenia (Buchanan et al. 2010; Dixon et al. 2010). ASPE also conducted a supplemental review of the clinical literature restricted to human adult clinical trials, and in the case of pharmacologic agents, those that have advanced

beyond preliminary safety and efficacy testing (Sherry 2010). Because the PORT recommendations include only studies published through March 2008, the ASPE literature review identified more recent studies. In addition, we consulted with a multistakeholder TAG. To identify existing measures assessing care for people with schizophrenia, we searched measure databases from the NQF, the National Quality Measures Clearinghouse, the National Registry of Evidence-Based Programs and Practices through the Substance Abuse and Mental Health Services Administration, and the Center for Quality Assessment and Improvement in Mental Health. Measures were organized by the measure steward, name, description, numerator, denominator, data source, and measurement domain (that is, physical health, pharmacotherapy, and psychosocial interventions). The final measure concepts are presented in Chapter III.

2. Define and Develop Initial Measure Specifications

Based on the review of the literature and feedback from the TAG and ASPE, we developed detailed specifications for an initial set of 17 measure concepts before settling on a final set of ten to be submitted to NQF for endorsement. Initial measure specifications included codes likely to be found on claims and that define populations eligible to be in the denominator, codes that adequately defined the nature of the processes or outcomes to be assessed (the numerator), and the appropriate time frames for assessment. We used the input of the TAG and our understanding of the MAX data to guide drafting measure specifications. Appendix A lists the original 17 measure concepts.

3. Convene Meetings of the Project Technical Advisory Group

To guide the measure development process and provide the perspectives of all stakeholders, we convened three meetings of a multistakeholder TAG. This group included 16 members representing expertise in clinical care, research, state and federal policy, consumers, managed behavioral health care, and quality measurement. The TAG met three times by teleconference through the course of the project. During the first teleconference, we asked TAG members to review proposed measure concepts, identify potential gaps in these concepts, assess measure development priorities, and recommend measures to be specified and tested. Measure specifications and the testing plan for the selected concepts were then reviewed during the second TAG meeting. The third meeting consisted of reviewing the preliminary results of the field and pilot-testing. In addition, the TAG evaluated and provided further feedback on the specifications and recommended measures for NQF submission. Appendix B lists the TAG members and includes material presented at each TAG meeting.

4. Field-Test Measure Specifications with Key Stakeholders

To inform our understanding of feasibility and usability, we conducted focus groups with: (1) State Medicaid Medical Directors; (2) representatives from MBHOs; and (3) State Mental Health Commissioners and Medical Directors (or their designees). The goal was to obtain feedback on attributes that are reviewed by NQF during the endorsement process, including the importance, usability, and feasibility of the measures. We asked focus group participants about their understanding of the measure specifications; the feasibility of implementing quality data for the measures through a claims-based system, including anticipated operational challenges in collecting and reporting the data; the relevance and importance of the measures to their program or organization; their interest in collecting information and receiving feedback on the measures; and any suggestions for refining the measures.

Focus group testing with the State Medicaid Medical Directors occurred in conjunction with the Medicaid Medical Directors Learning Network meeting in Washington, DC, and 28 states were represented. Representatives of MBHOs were recruited from industry lists; individuals representing commercial and Medicaid plans in six states (Florida, Oklahoma, Pennsylvania, Illinois, Missouri, and Iowa) participated. We later added a focus group of state mental health commissioners and medical directors in response to suggestions from ASPE; officials from five states (California, Michigan, Missouri, Georgia, and Florida) participated. A memo summarizing our conversations with the focus groups is in Appendix C.

5. Post Measure Specifications for Public Comment

For this task, NCQA developed and managed a dedicated web page to receive public comments. Candidate measures (excluding the HIV screening and psychosocial treatment measures) were posted September 15, 2011, through October 15, 2011, and included draft technical specifications, instructions, and supporting information for the public-comment period. We collated the public comments and reviewed them to identify themes and areas of concern. We then prepared a document summarizing the comments and action taken (Appendix D). Twenty-two organizations, including academic institutions, health plans, pharmaceutical companies, universities, and other health care associations, submitted a total of 67 comments.

6. Pilot-Test Measures to Assess Usability, Validity, and Reliability

To assess the usability and scientific acceptability of the measures, we examined the distribution, content and convergent validity, and test-retest reliability of the candidate measures using MAX data from 2007 and 2008. Use of MAX data permits real-world assessment of measure usability for state Medicaid officials. At the same time, operationalization of quality measures in Medicaid claims data provides an opportunity to retrospectively assess measure validity by correlating measure

performance with outcomes such as schizophrenia-related hospitalization and emergency department (ED) use. The MAX data are standardized eligibility and claims files for each state that include person-level on every beneficiary enrolled in Medicaid during the calendar year. The MAX files are created from claims data that each state submits to the Centers for Medicare and Medicaid Services (CMS).

Defining the Population

Diagnosis of schizophrenia was inferred by either a single primary inpatient diagnosis or two outpatient primary diagnoses of schizophrenia.[1,2] In response to comments from Medicaid medical directors, we modified and tested some measures to include persons with *serious mental illness* (SMI) defined by a single primary inpatient diagnosis or two outpatient primary diagnoses of either schizophrenia or bipolar disorder.

In addition, we required that enrollees have 10 months of Medicaid eligibility, non-dual status, and qualification for Medicaid on the basis of a disability, which resulted in 1,019,123 Medicaid recipients who met our inclusion criteria.[3]

Overall, 9.7 percent of Medicaid recipients in our dataset had schizophrenia and 12.8 percent had SMI (bipolar disorder and/or schizophrenia) in 2007. Both of these populations were demographically diverse (Appendix Table E.2). About one in five enrollees with schizophrenia were diagnosed with diabetes (17 percent).

Pilot-Test Methodology: Usability, Validity, and Reliability

Pilot-testing the measures using MAX data took several forms. First, we evaluated measure importance (gaps in quality) and scientific acceptability (meaningful differences in performance) by assessing the distributional properties of each measure. This was accomplished by tabulating the minimum, maximum, median, mean, and interquartile range (IQR) for each measure at the state level. The IQR is demarcated by the values at the 25th and 75th percentiles of a distribution. Generally speaking, measures with a broader IQR are preferable to measures with a narrowly distributed IQR or those with an IQR at the very low or very high end of the distribution. For example, a measure with a narrow IQR may not be sufficiently sensitive to detect differences in quality. Measures with an IQR of at least 10 percentage points were considered to have the strongest evidence of usability for quality measurement purposes.

---

[1] An ICD-9 code of 295.xx was used to flag schizophrenia.

[2] Outpatient diagnoses were observed on different days.

[3] We used MAX data from the following states in 2007: Alabama, Alaska, California, Connecticut, Georgia, Idaho, Illinois, Indiana, Iowa, Louisiana, Maryland, Missouri, Mississippi, New Hampshire, North Carolina, North Dakota, Nevada, Oklahoma, South Dakota, Washington DC, West Virginia, and Wyoming. These states were noted to have complete enrollment, fee-for-service (FFS) claims and encounter records. Although the sample was primarily enrolled in FFS plans, some states with complete encounter data were included in our analytic sample.

Validity and reliability are important characteristics of measure scientific acceptability. Construct validity was evaluated by examining enrollee outcomes with results displayed by quartile of state-level performance for each measure. We compared rates of schizophrenia-related hospitalization and ED utilization, for beneficiaries in the highest and lowest performing quartile for each quality measure. The difference between the outcomes among enrollees in the best and worst quartiles of state performance for each measure was tested using a one-way analysis of variance; an F-test significance level of $p < 0.01$ was used to determine statistically different outcomes. For a given measure, construct validity was inferred when rates for adverse events among enrollees in high performing states were significantly better (i.e., lower) than the rates of adverse events among enrollees in low performing states.

Convergent validity was examined through between-measure correlation coefficients. For example, we hypothesized that adherence to antipsychotics, as measured by a high rate of antipsychotic medication possession ratio, would be negatively associated with measures of mental health ED use and positively correlated with the measures of 30-day outpatient follow-up after a mental health related discharge. We identify measures with a Pearson correlation of at least 0.15 with two or more measures.

We assessed measure reliability using state-level test-retest correlations with data from 2007 and 2008 MAX data.[4] We identify measures with a year-to-year correlation of ≥0.30. We also examined the stability of relative performance quartiles between 2007 and 2008, with the expectation that at the state level, performance measures should not exhibit any discernible pattern of performance instability over time. In other words, measure stability would be demonstrated if a state was in the top quartile of performance for a given measure in 2007, the same state should demonstrate similar relative performance in 2008. Results from the pilot and field-testing efforts are summarized in the next section.

---

[4] 2008 data were available for a subset (N=16) of the 2007 states: Alabama, Alaska, Connecticut, Georgia, Idaho, Indiana, Iowa, Louisiana, Maryland, Mississippi, New Hampshire, North Carolina, Oklahoma, South Dakota, West Virginia, and Wyoming.

III. SUMMARY OF KEY FINDINGS

The purpose of this measure development project was to identify, specify, and test at least three measures that address pharmacological treatment, psychosocial treatment, and physical health needs for patients with schizophrenia that can be calculated solely from Medicaid claims data. Ten measures met our rigorous criteria for measure development, including evidence review, consultation with the TAG, focus groups with key stakeholders, public comment, and pilot-testing using the MAX data.

Tables III.1-III.4 list the measure concepts that we considered based on the environmental scan and initial input from the TAG; these concepts addressed the domains requested by ASPE (pharmacology, psychosocial treatment, and physical health) as well as a set of cross-cutting issues identified through the scan. We did not further pursue some of these topics because we did not believe that they could be assessed in claims; these measure concepts were not presented to the TAG (see Appendix B).

Based on TAG recommendations, 13 measures were specified. Two (use of any psychosocial treatment and HIV screening) were dropped before testing in the MAX files. The psychosocial treatment measure was dropped because procedure codes used in claims data are ambiguous and thus do not provide sufficient detail to reflect the actual service provided, and because these codes are not used consistently in different states and programs. The HIV screening measure was dropped because of the lack of strong evidence suggesting a gap in care for people with schizophrenia. Based on the input received from the public comment period, we dropped the measure of general ED utilization due to provider attribution concerns, which resulted in ten measures that were later pilot-tested in the MAX data.

1. Measure Concepts Considered, Specified, and Tested, and Submitted for Endorsement

The project team sought to develop measures in three domains, pharmacology, psychosocial care, and physical health, as well as cross-cutting measures that span several of these domains. Tables III.1-III.4 list the proposed measure concepts, the measures that were specified and tested in focus groups, the measures that were tested in the MAX data, and the measures submitted for NQF endorsement. The final ten measures submitted to NQF for endorsement consideration are listed in the last column. Appendix F consists of the proposed measures' numerator, denominator, and exclusions.

TABLE III.1. Pharmacological Concepts Considered, Specified, Tested, and Submitted			
Proposed Measure Concepts	Measures Specified & Tested in Focus Groups	Measures Tested in MAX Files	Measures Submitted for NQF Endorsement
1. Use of antipsychotic medications for treatment of schizophrenia. 2. Antipsychotic medication possession ratio. 3. Use of clozapine in treatment-resistant patients. 4. Polypharmacy treatment.	1. Use of antipsychotic medications for treatment of schizophrenia. 2. Antipsychotic medication possession ratio.	1. Use of antipsychotic medications. 2. Antipsychotic medication possession ratio.	1. Use of antipsychotic medications. 2. Antipsychotic medication possession ratio.

Use of clozapine in treatment-resistant patients was dropped due to difficulty with identifying treatment-resistant patients from claims data and concerns about small denominator size. The polypharmacy treatment measure concept was dropped because there is insufficient evidence to define a polypharmacy threshold (e.g., two versus three antipsychotics) and lack of evidence regarding the impact of polypharmacy on quality of care. The TAG also was uncertain whether to broaden the concept to encompass other psychiatric medications (e.g., antidepressants).

TABLE III.2. Psychosocial Concepts Considered, Specified, Tested, and Submitted			
Proposed Measure Concepts	Measures Specified & Tested in Focus Groups	Measures Tested in MAX Files	Measures Submitted for NQF Endorsement
1. Use of Assertive Community Treatment (ACT) post-hospitalization. 2. Use of case management. 3. Use of family therapy. 4. Use of supported employment. 5. Use of cognitive behavioral therapy. 6. Use of social education. 7. Use of any psychosocial treatment. 8. Availability of psychosocial treatment. 9. Presence or duration of waiting list for psychosocial treatment.	1. Use of any psychosocial treatment.	(None)	(None)

Use of ACT post-hospitalization, case management, family therapy, supported employment, cognitive behavioral therapy, and social education were dropped as a result of the inconsistent availability of these services across state Medicaid programs

and, where those services are available, unreliable coding and uncertain fidelity to the evidence-based models. Use of any psychosocial treatment was specified and tested in focus groups, but was dropped because of the fidelity and reliability concerns. Availability of and the presence or duration of a waitlist for psychosocial treatment are structural measures not suited to claims data measurement.

TABLE III.3. Physical Health Concepts Considered, Specified, Tested, and Submitted			
Proposed Measure Concepts	**Measures Specified & Tested in Focus Groups**	**Measures Tested in MAX Files**	**Measures Submitted for NQF Endorsement**
1. Monitoring of metabolic conditions among patients taking antipsychotic medications. 2. Weight assessment and counseling among patients who are taking antipsychotics. 3. Appropriate health maintenance and prevention. 4. Appropriate infectious disease screenings. 5. Screening and counseling of substance use disorders. 6. Tobacco counseling.	1. Cervical cancer screening for women. 2. HIV screening. 3. Diabetes screening (schizophrenia or bipolar disorder). 4. Cardiovascular health screening (schizophrenia or bipolar disorder). 5. Diabetes monitoring. 6. Cardiovascular health monitoring.	1. Cervical cancer screening for women. 2. Diabetes screening (schizophrenia or bipolar disorder). 3. Cardiovascular health screening (schizophrenia or bipolar disorder). 4. Diabetes monitoring. 5. Cardiovascular health monitoring.	1. Cervical cancer screening for women. 2. Cardiovascular health screening (schizophrenia or bipolar disorder). 3. Diabetes screening (schizophrenia or bipolar disorder). 4. Diabetes monitoring. 5. Cardiovascular health monitoring.

Weight assessment and counseling among patients on antipsychotics was deemed identifiable only from chart data, which were out of scope for this project. Concerns about reliable documentation of tobacco and substance use screening and counseling in claims data resulted in removing these concepts from further consideration. HIV screening was dropped because of the lack of strong evidence suggesting a gap in care for people with schizophrenia.

TABLE III.4. Cross-Cutting Concepts Considered, Specified, Tested, and Submitted			
Proposed Measure Concepts	**Measures Specified & Tested in Focus Groups**	**Measures Tested in MAX Files**	**Measures Submitted for NQF Endorsement**
1. Use of combination antipsychotic medication and psychosocial treatment. 2. Outpatient follow-up visit after hospitalization. 3. ED use. 4. Continuous Medicaid enrollment.	1. 7-day follow-up visit after mental health hospital discharge. 2. 30-day follow-up after mental health hospital discharge. 3. Any mental health ED use. 4. Any ED use.	1. 7-day follow-up visit after mental health hospital discharge. 2. 30-day follow-up after mental health hospital discharge. 3. Any mental health ED use.	1. 7-day and 30-day follow-up visit after mental health hospital discharge. 2. Any mental health ED use.

The use of combination antipsychotic medication and psychosocial treatment measure concept was dropped due to the inability to capture psychosocial treatments reliably through claims data.

2. Field-Testing

The focus groups with state Medicaid and mental health leaders, as well as with MBHO staff, yielded remarkably consistent results. Key points included:

- Claims data are unreliable for identifying some behavioral health services, particularly evidence-based psychosocial treatments.

- Variation in financing of services for people with SMI limits the ability to consistently measure the quality of care across Medicaid programs. For example, while some states reimburse for a bundled set of services collectively known as assertive community treatment (ACT), other states reimburse individual services that resemble services included in the ACT model. In other states, some of these services are provided outside of the Medicaid program, such as through the state mental health authority.

- Some candidate measures address problems that are not unique to patients with schizophrenia; measures could be broadened to include patients with bipolar disorder, schizophrenia, and severe forms of depression (SPMI).

While focus group participants generally viewed the proposed measure concepts as important and relevant topics, they noted some gaps. In particular, Medicaid officials raised concerns about the lack of candidate measures addressing perceived problems of overuse of care for people with schizophrenia (for example, polypharmacy or hospital readmissions).

The panels offered specific advice on technical specifications and testing. In particular, they recommended that the measures apply to patients not included in MAX files, specifically TANF enrollees and people with dual Medicare beneficiaries, who receive treatment through Medicaid programs.

3. Public Comment

The feedback from public comment was positive, with 87 percent of the comments either supporting the measures or supporting them with modifications (Appendix D). The majority of the comments touched on issues that had been discussed by the project team and the TAG during the measure development process, such as expanding the denominator in the physical health screening measures to include anyone with SMI, including measures evaluating psychosocial care, and lowering the age of eligibility for the measures.

Some comments raised concerns about the accountability for measures; for example, several commenters expressed concern that offering cervical cancer screening was out of scope for psychiatrists and psychologists. The project team believes this is a misunderstanding on the part of providers. The state, not the provider, is the unit of accountability for these measures. Further, given the push toward integrated care, states may be held accountable for the coordination of care between medical and mental health settings. This may include encouraging mental health professionals, including psychiatrists, to inquire about these services and potentially refer for such services. This is no different from the expectation that psychiatrists address the metabolic condition of patients in their care. Therefore, we propose retaining screening measures.

We received technical comments concerning coding of medication lists, including HbA1c tests as part of the diabetes screening measure, and methods to determine use of injectable antipsychotic medications. The project team carefully considered these concerns when finalizing measure specifications.

The measure that received the least support from public comment was Emergency Department Utilization for People with Schizophrenia. Feedback centered on the measure being non-action-oriented because it included non-mental health admissions. Comments also focused on the measure possibly encouraging overuse of emergency servces. Based on this feedback, the broad measure of Emergency Department Utilization was not submitted for NQF endorsement.

4. Pilot-Testing

The objective of pilot-testing was to determine the scientific acceptability of the measures based on NQF criteria. Table III.5, summarizes the evidence found for each measure through our pilot-testing activities using our 22-state MAX dataset (2007) and our 16-state MAX dataset (2008). Cells containing an 'X' indicate that a measure met predetermined criteria, summarized in Chapter II, which we used to assess differences in performance across states, validity, or reliability. An empty cell indicates that a measure did not meet the criterion in the corresponding column; however, as we discuss in the paragraphs that follow, this does not indicate a measure is without merit or should not be considered useful. In general, as we described below in further detail, caution is warranted in interpreting our pilot-testing findings, as testing results using Medicaid claims should not be used as the sole criteria for judging the merit of the measures.

TABLE III.5. Summary of Pilot-Testing Results: Evidence of Measure Usability, Validity, and Reliability					
Measure	**Detection of Meaningful Differences**	**Validity**		**Reliability**	
	IQR Dispersion[a]	**Construct Validity[b]**	**Convergent Validity[c]**	**Test-Retest Correlation[d]**	**Performance Stability[e]**
Use of Antipsychotic Medication		X			
Antipsychotic Possession Ratio (≥80%)				X	
Diabetes Screening (SMI)[f]	X	X	X	X	X
Diabetes Monitoring	X	X	X	X	X
Cardiovascular Health Screening (SMI)[f]		X		X	
Cardiovascular Health Monitoring	X	X		X	X
Cervical Cancer Screening				X	X
ED Utilization for Mental Health Conditions		N/A		X	
Follow-up after Mental Health Hospital Discharge (7-day)	X	X			X
Follow-up after Mental Health Hospital Discharge (30-day)	X	X	X		X
a. Dispersion indicated by an IQR of at least 10 percentage points (Appendix Table E.13). b. Construct validity indicated by significant performance differences between top and bottom quartile of measure performance for either schizophrenia-related hospitalization or ED utilization (Appendix Table E.14). c. Convergent validity indicated by Pearson r≥0.15 in hypothesized direction with at least 2 other measures (Appendix Table E.15). d. Reliability indicated by state-level test-retest correlation (2007-2008) Pearson r≥0.30 (Appendix Table E.16). e. Stability indicated by no more than 1 performance quartile change for any state between 2007 and 2008. For some measures, states had denominators <100 in 2008; these measure/state combinations were excluded from this analysis. f. Measure calculated among enrollees with schizophrenia or bipolar disorder.					

1. **Five of the ten proposed measures demonstrated significant variability in state-level performance.** A key indicator of a quality measure's utility is its ability to capture a wide range of performance. Appendix Table E.13 lists each measure and its distribution across the 22-state dataset. Table III.5 identifies the four measures with an IQR of at least 10 percentage points and those where the lower and upper bounds of the IQR did not encompass the tails of performance (either low or high), indicating measures with the greatest utility for quality measurement purposes.

   The measure "Use of Antipsychotic Medication" had the most restricted performance range (an IQR of 3 percentage points). For example, a state

performing at the lower end of the IQR (that is, the 25th percentile), reported 92 percent of recipients received an antipsychotic, while a state at the top end of the IQR (the 75th percentile) reported 95 percent of recipients received an antipsychotic. Therefore, we believe that this measure has limited value from a quality improvement perspective, since the performance range is restricted and is already near the top, thus limiting the potential for improvement. However, because antipsychotic use is a fundamental issue for this population and the measure was widely endorsed by our consultants (the TAG and stakeholder groups), "use of antipsychotic medication" has considerable utility as a monitoring measure.

2. **Seven of the ten proposed measures demonstrated evidence of validity.** We assessed validity using two approaches. To assess construct validity we examined the association between measure performance and outcomes (schizophrenia-related hospitalization and ED visits). We compared the percentage of people who hospitalized or visited the ED for schizophrenia, comparing the worst and best-performing quartiles of state performance for each measure. For example, we found enrollees in states with the highest rates of antipsychotic use had significantly lower rates of hospitalization for schizophrenia compared with enrollees in states with the lowest rates of antipsychotic use (Appendix Table E.14). Seven measures demonstrated evidence of construct validity.

   Convergent validity was determined through examination of recipient-level measure correlations (Appendix Table E.15). We considered measures with a correlation coefficient of 0.15 or greater with at least two other measures to demonstrate evidence of convergent validity. Three of the ten measures met this criterion.

   Although some of these results are encouraging, some important limitations of these measures warrant consideration. Our measures of schizophrenia-related hospitalization and schizophrenia-related ED visits assess adverse outcomes at one extreme of care and thus do not reflect the full spectrum of care. Further, measures that assess preventive care processes were not anticipated to have a significant effect on schizophrenia-related hospitalization or ED use, therefore this relationship warrants further investigation to understand this finding.

3. **Nine of the ten measures demonstrated evidence of reliability.** Reliability was assessed through correlation of state-level 2007 and 2008 performance. Seven of the ten measures demonstrated 2007-2008 correlation of 0.30 or higher at the state level (Appendix Table E.16). In addition, we compared each state's performance quartile in 2007 with its performance quartile in 2008 to understand the stability of each measure. We defined *stability* as no more than a one-quartile performance difference between 2007 and 2008; six measures met this criterion (Table III.5). Only "Use of Antipsychotic Medications" failed to show a strong state-level year-to-year correlation (r=0.25) and showed a large performance

difference (a three-quartile change) between 2007 and 2008, although this difference was observed in a single, small state.

In summary, we began with a list of 23 measure concepts to assess the care provided to Medicaid enrollees with schizophrenia, and arrived at a final list of ten measures for submission to NQF. These measures fall into three domains, pharmacological, physical health measures and cross-cutting measures. Current evidence and limitations of claims data prevented us from developing robust measures of psychosocial treatments. Appendix F details the numerator, denominator and exclusions for each of the ten proposed measures.

IV. LESSONS LEARNED

While we successfully developed and tested ten quality measures, development of several additional measures was not feasible given the constraints of Medicaid claims data and Medicaid payment policies. The following discussion of our experience and lessons learned is designed to be instructive for future efforts in the development of quality measures for people with SPMI.

1. Use of Medicaid claims data as a source to implement and test schizophrenia quality measures presented several noteworthy limitations. Because of the limitations of the claims data, several evidence-based practices could not be implemented as measures. These limitations were particularly conspicuous when attempting to operationalize evidence-based guidelines for psychosocial treatments such as those recommended in the Schizophrenia PORT. In analyses using MAX data, we found psychosocial treatments are either inconsistently coded in claims data or not available at all. For example, claims for smoking cessation programs were not observed in the MAX data; therefore, this measure was not developed because it could not be assessed in claims data. Consequently, no psychosocial measures emerged from our measure development process, despite the strength of evidence for these practices. Specific evidence-based recommendations that could not be accurately identified in the claims data, and thus were not field or pilot-tested, included:

    - Supported employment;
    - Family psychoeducation;
    - Assertive community-based treatment;
    - Cognitive behavioral therapy;
    - Social skills training.

    Claims-only assessment presents other challenges for measure development. Because mental health problems are difficult to diagnose, claims often contain incorrect information that present challenges to accurate case finding. We attempted to minimize this problem by requiring either an inpatient claim with a primary diagnosis of schizophrenia or two outpatient claims on different days with a primary diagnosis of schizophrenia, adapting definitions used by others (Busch, Frank & Lehman 2004). However, we acknowledge that claims are not an ideal source to identify this population and may provide an undercount of the target population as diagnosis fields are not required for payment of services. Although current guidelines specify follow-up with a mental health provider following hospitalization, performance on our candidate measure is assessed by follow-up with any provider because mental health providers cannot be identified in Medicaid claims.

Finally, use of MAX data to test the measures limits the external validity of our results. Our MAX analytic study population was purposely limited to Medicaid recipients with claims data so that we could reliably identify patients with schizophrenia and the services they received. As a result, our study population included primarily disabled, non-dual-eligible enrollees in FFS plans. However, this group represents only a minority of the universe of people with SMI who receive mental health treatment through Medicaid programs. In particular, because drugs treatments are reimbursed by Medicare Part D for dually-eligible enrollees we are unable to include them, thus eliminating about 40 percent of all disabled Medicaid recipients from performance assessment.

2. Several topics were of interest to ASPE, the development team, and stakeholders, but the evidence base, tools, and methods for tracking these measures are immature. For example, evaluating receipt of evidence-based psychosocial services may require measures that address the structures of care (e.g., availability of trained providers, supervision). State officials in particular were interested in measures addressing potential overuse of pharmacological treatments, which is challenging to document in the absence of tools for risk adjustment and symptom measurement. In addition, the evidence to support overuse measures is inconsistent. Patient-reported outcomes were also of interest to stakeholders, but they cannot be ascertained using claims data.

There was considerable interest in focus groups and TAG on addressing the physical health needs of people with schizophrenia; however, there was not always evidence to provide a rationale for a particular focus on such people for a given test. Some highly important preventive services, in particular tobacco cessation counseling and assistance, are not feasible in claims data. While there was evidence of low rates of cervical cancer screening among women with schizophrenia, there was no such evidence of a gap in care for HIV screening. Continuity of Medicaid enrollment was proposed to assess whether people with schizophrenia have consistent access to services; however, some lapses in coverage may be related to desirable outcomes (such as employment), and it would not be possible to determine the reason for loss of coverage. As the evidence base grows and use of electronic medical records and other electronic data repositories (for example, registries) also grows, so too will the ability to implement evidence-based measures.

3. Quality measurement for Medicaid recipients with schizophrenia presents implementation issues. During the development process, and in particular during the field-testing process, we became aware of several issues related to measure implementation. Key implementation issues included measure attribution, variations in care financing, and the need for long look-back periods for several measures. For example, although the TAG and several stakeholders endorsed the inclusion of a general measure tracking ED use, some providers voiced concerns about attribution for this measure. Specifically, during the field-testing process, mental health providers felt they should not be held accountable for ED

visits for accidents or other non-mental health reasons. Consequently, we dropped the measure of general ED use from our pilot-testing. However, attribution of care processes and outcomes will likely prove controversial, though implementation of the proposed measures at the state (rather than the provider level) will help to minimize concerns over attribution.

We found that variation in the financing of services for people with SMI limited our ability to measure the care provided by Medicaid programs. For example, the provision of services through state mental health systems, the coverage of mental health services through Medicare for dual-eligible beneficiaries, the prohibition of same-day billing of medical and behavioral health services, and interstate variation in Medicaid and disability standards all underscore the limitations of claims data to measure quality for enrollees with schizophrenia.

Finally, we found that reliance on Medicaid claims to produce rates of health screening can require a large volume of data to address issues of "look-back" for selected conditions. For example, some health conditions have a screening recommendation of every five years. Therefore, to compute a health screening measure for these conditions, information systems require the capacity to look back over a five-year claims history, which for some states could be a daunting task.

4. The distinction between enrollees with schizophrenia and other SMI conditions is, in many cases, artificial. The project team, ASPE, and measure stakeholders all expressed the belief that conceptually, many issues related to schizophrenia also apply broadly to people with any SMI. It was outside the scope of this project to conduct the full evidence review and testing necessary for this work. Further work is needed to consider whether measures similar to the ones developed and tested under this contract would be relevant for people with bipolar disorder and other SMI.

REFERENCES

Buchanan, Robert W. et al. (2010). "The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements." *Schizophrenia Bulletin*, 36(1): 71-93.

Busch, Alisa B., Richard Frank and Anthony Lehman. "The effect of a managed behavioral health carve-out on quality of care for Medicaid patients diagnosed as having schizophrenia." *Archives of General Psychiatry*, 61: 442-448.

Dixon, Lisa B., Faith Dickerson, Alan S. Bellack, et al. (2010). "The 2009 PORT psychosocial treatment recommendations and summary statements." *Schizophrenia Bulletin*, 36(1): 48-70.

Sherry, Tisamarie (2010). "Guidelines for the Treatment of Schizophrenia: A Review of the Literature." Unpublished draft report.