

State of Colorado



Department of Health Care Policy & Financing

FY 06–07 PIP VALIDATION REPORT

Screening for Bipolar Disorder

for

Behavioral HealthCare, Inc.

June 2007



1600 East Northern Avenue, Suite 100 • Phoenix, AZ 85020
Phone 602.264.6382 • Fax 602.241.0757

1. Executive Summary	1-1
Overview.....	1-1
Summary of Study	1-1
Study Topic	1-1
Study Methodology	1-2
Study Results	1-2
Scoring	1-3
Summary of Validation Findings.....	1-4
Conclusions	1-4
Requirements	1-4
Recommendations.....	1-4
Comparison of Years 1 and 2.....	1-4
2. Scoring Methodology	2-1
PIP Scores.....	2-2
3. Validation and Findings Summary.....	3-1
Validations and Findings Summary	3-1
Activity I. Appropriate Study Topic	3-1
Activity II. Clearly Defined, Answerable Study Question.....	3-2
Activity III. Clearly Defined Study Indicator(s)	3-2
Activity IV. Use a Representative and Generalizable Study Population.....	3-3
Activity V. Valid Sampling Techniques.....	3-4
Activity VI. Accurate/Complete Data Collection	3-4
Activity VII. Appropriate Improvement Strategies.....	3-5
Activity VIII. Sufficient Data Analysis and Interpretation.....	3-6
Activity IX. Real Improvement Achieved	3-6
Activity X. Sustained Improvement Achieved.....	3-7
4. FY 06–07 PIP Validation Tool	4-1

Appendices

Introduction	Appendices Cover Page
Appendix A: Summary Form: Behavioral HealthCare, Inc.'s <i>Screening for Bipolar Disorder</i>.....	A-1
Appendix B: CMS Rationale by Activity.....	B-1
Appendix C: Definitions and Explanations by Activity.....	C-1

Overview

The Balanced Budget Act (BBA) of 1997 (Public Law 105-33) requires that states conduct an annual evaluation of their managed care organizations (MCOs) and prepaid inpatient health plans (PIHPs) to determine the MCOs' and PIHPs' compliance with federal regulations and quality improvement standards. According to the BBA, the quality of health care delivered to Medicaid consumers in MCOs and PIHPs must be tracked, analyzed, and reported annually. The Colorado Department of Health Care Policy & Financing (the Department) has contractual requirements with each MCO and behavioral health organization (BHO) to conduct and submit performance improvement projects (PIPs) annually. As one of the mandatory external quality review activities under the BBA, the Department is required to validate the PIPs. To meet this validation requirement, the Department contracted with Health Services Advisory Group, Inc. (HSAG) as an external quality review organization. The primary objective of the PIP validation is to determine the compliance with requirements set forth in 42 CFR 438.240(b)(1), including:

- ◆ Measurement of performance using objective quality indicators.
- ◆ Implementation of system interventions to achieve improvement in quality.
- ◆ Evaluation of the effectiveness of the interventions.
- ◆ Planning and initiation of activities for increasing or sustaining improvement.

The Centers for Medicare & Medicaid Services (CMS) publication, *Validating Performance Improvement Projects: A Protocol for Use in Conducting Medicaid External Quality Review Activities*, Final Protocol, Version 1.0, May 1, 2002, was used in the evaluation and validation of the PIPs.

Summary of Study

The purpose of the study was to evaluate access to and quality of care for Medicaid consumers receiving services from **Behavioral HealthCare, Inc. (BHI)**. Both clinician education and providing screening tools to the clinicians will be assessed to determine if these interventions increase the incidence of screening for mania in consumers diagnosed with a mental health disorder.

Study Topic

The topic addressed CMS' requirements related to access to and quality of care. The topic looked specifically at whether consumers diagnosed with a mental health disorder were being screened for mania. The study topic, *Screening for Bipolar Disorder*, reflected a large number of the plan's Medicaid population.

The study question presented by **BHI** was:

- ◆ “Does clinician education and the provision of screening tools improve the incidence of screening for mania in individuals diagnosed at admission by the intake clinician with the following diagnoses: substance-induced mood disorder, mood disorder due to a general medical condition, schizophrenia, schizoaffective disorder, major depressive disorder, mood disorder NOS, delusional disorder, psychotic disorder NOS, anxiety disorder, generalized and atypical, dysthymic disorder, cyclothymic disorder, borderline personality disorder, depressive disorder NOS, undifferentiated disruptive disorder, oppositional defiant disorder, attention deficit/hyperactive disorder?”

Study Methodology

For baseline, **BHI** collected data for four study indicators. For the first remeasurement, **BHI** discontinued two study indicators that measured adults and children screened for bipolar disorder because items from the screening were not found in one location within the document. During intake, clinicians checked a number of boxes on several different forms and many times did not remember all of the assessment criteria for mania. **BHI** determined that clinicians needed one tool to assess for mania. For the first remeasurement, **BHI** collected data for two study indicators that measured adult and youth screened for bipolar disorder with the Mood Disorder Questionnaire (MDQ) and Young Mania Rating Scale—Parent Version tool (YMRS-P), which were recommended by the American Psychiatric Association. Data were collected twice a year from manual and administrative sources, and statistical significance testing was performed between measurement periods for each mental health center.

The study indicators, as stated in the BHI PIP Summary Form, were:

- ◆ “Adult Screening with the MDQ”
- ◆ “Youth Screening with the MDQ or YMRS-P tool”

Study Results

For this validation cycle, **BHI** collected a baseline and first remeasurement of the study indicators. **BHI** calculated rates for Centers A, B, and C. There was statistically significant improvement in the rates of adults screened for bipolar disorder using the MDQ and in the rates of children and adolescents screened for bipolar disorder using the MDQ or YMRS-P tool for all three centers. See Table 1-1.

Table 1-1— BHI Bipolar Screening PIP Report					
Percent of Adults Screened with MDQ Tool			Percent of Youth Screened with MDQ or YMRS-P Tool		
Center	Baseline	Remeasurement 1	Center	Baseline	Remeasurement 1
A	0	74	A	3	62
B	3	56	B	0	37
C	0	36	C	0	36

As the study progresses, **BHI** will collect a second remeasurement for its two study indicators and will add more study indicators to address the issue of consumers with positive screens receiving medication evaluations.

Scoring

HSAG validates a total of 10 activities for each PIP. The PIP is validated annually. The validation reflects activities that have been completed. A health plan (BHO) may take up to three years to complete all 10 activities. Each activity consists of elements necessary for the successful completion of a valid PIP. Evaluation elements are the key CMS protocol components for each activity that reflect the intent of what is being measured and evaluated. Some of the elements are critical elements and must be scored as *Met* to produce an accurate and reliable PIP. Given the importance of critical elements, any critical element that receives a *Not Met* score results in an overall PIP validation status of *Not Met*. If one or more critical elements are *Partially Met*, but none is *Not Met*, the PIP will be considered valid with low confidence. Revisions and resubmission of the PIP would be required.

Summary of Validation Findings

- ◆ For this review, nine activities with a total of 52 elements were validated. Of this number:
 - 47 evaluation elements were *Met*.
 - 2 evaluation elements were *Partially Met*.
 - 0 evaluation elements were *Not Met*.
 - 3 evaluation elements were *Not Applicable (N/A)*.
- ◆ The total number of critical elements that were evaluated equaled 11. Of this number:
 - 11 critical elements were *Met*.
 - 0 critical elements were *Partially Met*
 - 0 critical elements were *Not Met*.
 - 0 critical elements were *N/A*.

The final validation finding for **BHI**'s PIP showed an overall score of 96 percent, a critical element score of 100 percent, and a *Met* validation status.

Conclusions

The study has successfully addressed access to and quality of care related to screening for mania in the behavioral health population. For this validation cycle, **BHI** collected a baseline and first remeasurement for each study indicator. From baseline to the first remeasurement, there was statistically significant improvement in the rates of adults, children, and adolescents screened for bipolar disorder using the MDQ and YMRS-P tool.

Requirements

There were no requirements for this validation cycle.

Recommendations

A description of the training and qualifications of all manual data collection staff members should be provided in the PIP Summary Form. Clear and concise written instructions for completing the manual data collection tool should also be included.

Comparison of Years 1 and 2

For Year 1, only Activity I, Appropriate Study Topic, through Activity VII, Appropriate Improvement Strategies, were assessed because it was the first year of the PIP and only baseline data analysis was completed at the time of the submission. For Year 2, Activity I, Appropriate Study Topic, through Activity IX, Real Improvement Achieved, were assessed. The study had

completed a baseline and first remeasurement at the time of the submission. **BHI** showed statistically significant improvement in the rates of adults, children, and adolescents screened for bipolar disorder from Year 1 to Year 2.

2. Scoring Methodology

for Behavioral HealthCare, Inc.

Validating PIPs involves a review of the following 10 activities:

- ◆ Activity I. Appropriate Study Topic
- ◆ Activity II. Clearly Defined, Answerable Study Question
- ◆ Activity III. Clearly Defined Study Indicator(s)
- ◆ Activity IV. Use a Representative and Generalizable Study Population
- ◆ Activity V. Valid Sampling Techniques (If Sampling was Used)
- ◆ Activity VI. Accurate/Complete Data Collection
- ◆ Activity VII. Appropriate Improvement Strategies
- ◆ Activity VIII. Sufficient Data Analysis and Interpretation
- ◆ Activity IX. Real Improvement Achieved
- ◆ Activity X. Sustained Improvement Achieved

All PIPs are scored as follows:

<i>Met</i>	(1) All critical elements were <i>Met</i> , and (2) 80 percent to 100 percent of all critical and noncritical elements were <i>Met</i> .
<i>Partially Met</i>	(1) All critical elements were <i>Met</i> , and 60 percent to 79 percent of all critical and noncritical elements were <i>Met</i> , or (2) One critical element or more was <i>Partially Met</i> .
<i>Not Met</i>	(1) All critical elements were <i>Met</i> , and <60 percent of all critical and noncritical elements were <i>Met</i> , or (2) One critical element or more was <i>Not Met</i> .
<i>Not Applicable</i> (N/A)	N/A elements (including critical elements if they were not assessed) were removed from all scoring.

For FY 06-07, the BHOs were provided an opportunity to resubmit additional information and/or documentation. The plans were required to take action for any evaluation element receiving a score of *Partially Met* or *Not Met*. The action could include resubmission of additional PIP documentation prior to final scoring. Future annual PIP submission should include all information pertinent to the PIP study to achieve a *Met* status.

PIP Scores

For this PIP, HSAG reviewed Activities I through IX. Table 2-1 and Table 2-2 show **BHI**'s scores based on HSAG's PIP evaluation of *Screening for Bipolar Disorder*. Each activity has been reviewed and scored according to HSAG's validation methodology.

**Table 2-1—FY 06-07 Performance Improvement Project Scores
for Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.**

Review Activity	Total Possible Evaluation Elements (Including Critical Elements)	Total Met	Total Partially Met	Total Not Met	Total N/A	Total Possible Critical Elements	Total Critical Elements Met	Total Critical Elements Partially Met	Total Critical Elements Not Met	Total Critical Elements N/A
I. Appropriate Study Topic	6	6	0	0	0	1	1	0	0	0
II. Clearly Defined, Answerable Study Question	2	2	0	0	0	1	1	0	0	0
III. Clearly Defined Study Indicator(s)	7	6	0	0	1	3	3	0	0	0
IV. Use a Representative and Generalizable Study Population	3	3	0	0	0	2	2	0	0	0
V. Valid Sampling Techniques	6	6	0	0	0	1	1	0	0	0
VI. Accurate/Complete Data Collection	11	9	2	0	0	1	1	0	0	0
VII. Appropriate Improvement Strategies	4	2	0	0	2	No Critical Elements				
VIII. Sufficient Data Analysis and Interpretation	9	9	0	0	0	2	2	0	0	0
IX. Real Improvement Achieved	4	4	0	0	0	No Critical Elements				
X. Sustained Improvement Achieved	1	Not Assessed				No Critical Elements				
Totals for All Activities	53	47	2	0	3	11	11	0	0	0

**Table 2-2—FY 06-07 Performance Improvement Project Overall Score
for Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.**

Percentage Score of Evaluation Elements Met*	96%
Percentage Score of Critical Elements Met**	100%
Validation Status***	Met

* The percentage score is calculated by dividing the total *Met* by the sum of the total *Met*, *Partially Met*, and *Not Met*.

** The percentage score of critical elements *Met* is calculated by dividing the total critical elements *Met* by the sum of the critical elements *Met*, *Partially Met*, and *Not Met*.

*** *Met* equals confidence/high confidence that the PIP was valid.

Partially Met equals low confidence that the PIP was valid.

Not Met equals reported PIP results that were not valid.

3. Validation and Findings Summary for Behavioral HealthCare, Inc.

Validations and Findings Summary

This section summarizes the evaluation of the activities validated for the PIP. A description of the findings, strengths, requirements, and recommendations is outlined under each activity section. See Appendix B for a complete description of CMS rationale for each activity.

The validation was performed on a PIP submitted by **Behavioral HealthCare, Inc., (BHI)**. The PIP evaluated access to and quality of care and services. **BHI** used two study indicators to collect the data and assess the outcomes for this study. The study indicators measured adults and youth screened for bipolar disorder. **BHI** completed nine activities for this validation cycle.

Activity I. Appropriate Study Topic

Study Topic

BHI continued the study topic, *Screening for Bipolar Disorder*, as its clinical PIP topic for the FY 06–07 validation cycle.

Finding(s)

Six of six evaluation elements, including one critical element, were *Met*.

Strength(s)

The study topic reflected a high-volume and high-risk condition for **BHI** consumers. The topic assessed access to and quality of care and services. The topic had the potential to affect consumer health.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity II. Clearly Defined, Answerable Study Question

Study Question(s)

BHI's study question, as stated in its PIP Summary Form, was:

- ◆ “Does clinician education and the provision of screening tools improve the incidence of screening for mania in individuals diagnosed at admission by the intake clinician with the following diagnoses: substance-induced mood disorder, mood disorder due to a general medical condition, schizophrenia, schizoaffective disorder, major depressive disorder, mood disorder NOS, delusional disorder, psychotic disorder NOS, anxiety disorder, generalized and atypical, dysthymic disorder, cyclothymic disorder, borderline personality disorder, depressive disorder NOS, undifferentiated disruptive disorder, oppositional defiant disorder, attention deficit/hyperactive disorder?”

Finding(s)

Both evaluation elements, including the critical element, were *Met*.

Strength(s)

The study question stated the problem in simple terms and set the focus of the study, which was to improve the incidence of screening for mania in individuals diagnosed at admission by the intake clinician. The goal of the study was to impact access to and the timeliness and quality of health care provided to **BHI** consumers.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity III. Clearly Defined Study Indicator(s)

Study Indicator(s)

The indicators, as stated in **BHI**'s current PIP Summary Form, were:

- ◆ “Adult screening with the MDQ”
- ◆ “Youth screening with MDQ or YMRS-P”

Finding(s)

Six of seven evaluation elements for this activity, including three critical elements, were *Met*. One evaluation element was *Not Applicable* because the study indicators were not nationally recognized measures.

Strength(s)

The study indicators were well-defined, objective, and measurable. There were data available for each indicator, allowing for the study question to be answered. The PIP included the basis on which the indicators were developed.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity IV. Use a Representative and Generalizable Study Population

Study Population

The study population included all **BHI**-eligible consumers with primary psychiatric diagnosis codes as specified in the population definition. All age ranges and enrollment lengths were acceptable.

Finding(s)

All three evaluation elements, including the two critical elements, were *Met*.

Strength(s)

The study population was defined completely and accurately. The enrollment requirements were provided and the study captured all consumers to whom the study question applied.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity V. Valid Sampling Techniques**Sampling Technique(s)**

The sampling technique used by **BHI** for this clinical PIP was specified. A random sample from each mental health center and age group was selected using the National Committee for Quality Assurance (NCQA) sampling calculator. The margin of error was 0.05 and the confidence level was 95 percent.

Finding(s)

Six of the six evaluation elements for this activity were *Met*, including the critical element.

Strength(s)

The sample sizes for each center and age group were identified, and the sampling techniques ensured a representative sample of the eligible population. The sampling techniques were in accordance with generally accepted principles of research design and statistical analysis.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity VI. Accurate/Complete Data Collection**Data Collection**

The hybrid method was used to collect data for this PIP. A combination of medical record abstraction and administrative data was used to collect data for the two study indicators.

Finding(s)

Nine of 11 evaluation elements were *Met*, including the one critical evaluation element. Two evaluation elements were *Partially Met* because the training and qualifications of the manual data collection staff members were not discussed, and clear, concise written instructions for the manual data collection tool were not provided.

Strength(s)

The data elements collected were defined. The sources of data were specified and a timeline for the collection of data was included. The manual data collection tool ensured consistent, accurate

collection of data. The manual data collection interrater reliability process was discussed in the PIP study.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

Training and qualifications of all manual data collection staff members should be provided in the PIP Summary Form. Clear and concise written instructions for completing the manual data collection tool should also be included.

Activity VII. Appropriate Improvement Strategies

Improvement Strategies

The interventions discussed in the PIP study consisted of educating clinicians on screening guidelines from bipolar practice guidelines and conducting a Web-based presentation on screening for bipolar disorder and the use of screening tools.

Finding(s)

Two of four evaluation elements were *Met* for this activity. Two evaluation elements were *Not Applicable*. One evaluation element was *Not Applicable* because there was statistically significant improvement in the study indicators, so the original interventions did not need to be revised. The other evaluation element was *Not Applicable* because the PIP was not to the point of standardizing interventions at the time of the submission.

Strength(s)

The interventions were related to causes/barriers identified through data analysis and quality improvement processes. The interventions were system changes that were likely to induce permanent change.

Requirement(s) (for Critical Elements)

There were no critical elements for this activity.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity VIII. Sufficient Data Analysis and Interpretation**Data Analysis and Interpretation**

BHI completed data analysis and interpretation for two study indicators from baseline to the first remeasurement. **BHI** used chi-square testing to identify statistical differences between measurement periods. **BHI** calculated rates for Centers A, B, and C. There was statistically significant improvement in the rates of adults screened for bipolar disorder using the MDQ and in the rates of children and adolescents screened for bipolar disorder using the MDQ or YMRS-P tool for all three centers from baseline to the first remeasurement.

Finding(s)

Nine of nine evaluation elements for this activity were *Met*, including the two critical elements.

Strength(s)

Data analysis was conducted according to the plan in the study. Factors that affected the internal/external validity of the study and factors that affected the ability to compare measurements were identified. The data were presented in an accurate and easily understood way, and an interpretation of findings was included.

Requirement(s) (for Critical Elements)

There were no requirements identified for this activity during this review.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity IX. Real Improvement Achieved**Real Improvement Achieved**

BHI provided statistical evidence demonstrating that real improvement was achieved for this PIP study.

Finding(s)

All four evaluation elements for this activity were *Met*.

Strength(s)

Remeasurement methodology was the same as baseline methodology. There was documented improvement in processes as demonstrated through the improved use of the screening tools for children, adolescents, and adults. The improvement appeared to be the result of the interventions. Statistically significant improvement was achieved for both study indicators.

Requirement(s) (for Critical Elements)

There were no critical elements for this activity.

Recommendation(s) (for Noncritical Elements)

There were no recommendations identified for this activity during this review.

Activity X. Sustained Improvement Achieved

Activity X was not assessed for the FY 06–07 submission of this PIP report. The PIP completed baseline and one remeasurement period. Sustained improvement cannot be assessed until the PIP has completed two or more remeasurement periods.

The BHO will continue with the PIP process, and Activity X can be assessed and validated at the next annual submission of the PIP.

Section 4: Colorado FY 06-07 PIP Validation Tool:
**Screening for Bipolar Disorder
 for Behavioral HealthCare, Inc.**

DEMOGRAPHIC INFORMATION			
Health Plan Name:	Behavioral HealthCare, Inc.		
Study Leader Name:	Ann Terrill Torrez	Title:	Director, Quality Improvement
Phone Number:	(303) 617-2815	E-mail Address:	ann_torrez@bhiinc.org
Name of Project/Study:	Screening for Bipolar Disorder		
Type of Study:	Clinical		
Date of Study:	7/1/2004 to 12/31/2006		
Type of Delivery System:	BHO	Number of Medicaid Consumers in BHO:	9,869
		Number of Medicaid Consumers in Study:	861
Year 2 Validation	Resubmission		

**Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.**

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
I. Appropriate Study Topic: Topics selected for the study should reflect the Medicaid enrollment in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of the disease. Topics could also address the need for a specific service. The goal of the project should be to improve processes and outcomes of health care. The topic may be specified by the State Medicaid agency or on the basis of Medicaid consumer input.			
1.	Reflects high-volume or high-risk conditions (or was selected by the State). N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study topic reflected a high-volume and high-risk condition.
2.	Is selected following collection and analysis of data (or was selected by the State). N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study topic was selected following the collection and analysis of data.
3.	Addresses a broad spectrum of care and services (or was selected by the State). The scoring for this element will be Met or Not Met.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study topic addressed a broad spectrum of care and services.
4.	Includes all eligible populations that meet the study criteria. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	All eligible populations that met the study criteria were included.
5.	Does not exclude consumers with special health care needs. The scoring for this element will be Met or Not Met.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Consumers with special health care needs were not excluded.
C* 6.	Has the potential to affect consumer health, functional status, or satisfaction. The scoring for this element will be Met or Not Met.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study topic had the potential to affect consumer health and functional status.

Results for Activity I				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
1	6	0	0	0

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
II.	Clearly Defined, Answerable Study Question: Stating the study question(s) helps maintain the focus of the PIP and sets the framework for data collection, analysis, and interpretation.		
	1. States the problem to be studied in simple terms. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study question stated the problem to be studied in simple terms.
C*	2. Is answerable. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study question was answerable.

Results for Activity II				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
1	2	0	0	0

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS	SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation		
III. Clearly Defined Study Indicator(s): A study indicator is a quantitative or qualitative characteristic or variable that reflects a discrete event (e.g., an older adult has not received a flu shot in the last 12 months) or a status (e.g., a consumer's blood pressure is or is not below a specified level) that is to be measured. The selected indicators should track performance or improvement over time. The indicators should be objective, clearly and unambiguously defined, and based on current clinical knowledge or health services research.		
C* 1. Are well-defined, objective, and measurable. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The numerator of Study Indicator 3 included adults who were screened using the Mood Disorders Questionnaire or Young Mania Rating Scale-Parent Version. It appeared the Young Mania Rating Scale-Parent Version tool was only used for children and adolescents; therefore, it should not have been included in the definition of the numerator for Study Indicator 3. Both definitions of the numerators for Study Indicators 3 and 4 referred to "(indicator #1)" within the definition, which was confusing. Study Indicator 1, which was deleted from the PIP, pertained to adults. It was unclear why the definition of the numerator for Study Indicator 4, which was specific to children and adolescents, referenced Study Indicator 1. Rereview April 2007 BHI made the necessary changes to the definitions of the study indicators in the resubmission. This evaluation element was changed from Partially Met to Met.
2. Are based on current, evidence-based practice guidelines, pertinent peer review literature, or consensus expert panels.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study indicators were based on practice guidelines.
C* 3. Allow for the study question to be answered. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study indicators allowed for the study question to be answered.

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

**Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.**

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
III.	Clearly Defined Study Indicator(s): A study indicator is a quantitative or qualitative characteristic or variable that reflects a discrete event (e.g., an older adult has not received a flu shot in the last 12 months) or a status (e.g., a consumer's blood pressure is or is not below a specified level) that is to be measured. The selected indicators should track performance or improvement over time. The indicators should be objective, clearly and unambiguously defined, and based on current clinical knowledge or health services research.		
	4. Measure changes (outcomes) in health or functional status, consumer satisfaction, or valid process alternatives. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study indicators measured changes in valid process alternatives.
C*	5. Have available data that can be collected on each indicator. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	There were available data that were collected on the study indicators.
	6. Are nationally recognized measures such as HEDIS specifications, when appropriate. The scoring for this element will be Met or N/A.	<input type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input checked="" type="checkbox"/> N/A	The study indicators were not nationally recognized measures.
	7. Includes the basis on which the indicator(s) was adopted, if internally developed.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The basis on which each indicator was adopted was included.

Results for Activity III				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
3	6	0	0	1

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
IV.	Use a representative and generalizable study population: The selected topic should represent the entire eligible Medicaid enrollment population with systemwide measurement and improvement efforts to which the PIP study indicators apply.		
C*	1. Is accurately and completely defined. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study population was accurately and completely defined.
	2. Includes requirements for the length of a consumer's enrollment in the BHO.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Requirements for length of enrollment were included.
C*	3. Captures all consumers to whom the study question applies. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The study population captured all consumers to whom the study question applied.

Results for Activity IV				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
2	3	0	0	0

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
V.	Valid Sampling Techniques: (This activity is only scored if sampling was used.) If sampling is to be used to select consumers of the study, proper sampling techniques are necessary to provide valid and reliable information on the quality of care provided. The true prevalence or incidence rate for the event in the population may not be known the first time a topic is studied.		
	1. Consider and specify the true or estimated frequency of occurrence.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The frequency of occurrence was specified.
	2. Identify the sample size.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The sample size was identified.
	3. Specify the confidence level.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The confidence level was 95 percent.
	4. Specify the acceptable margin of error.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The acceptable margin of error was specified.
C*	5. Ensure a representative sample of the eligible population.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The sampling techniques ensured a representative sample of the eligible population.
	6. Are in accordance with generally accepted principles of research design and statistical analysis.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The sampling techniques were in accordance with generally accepted principles of research design and statistical analysis.

Results for Activity V				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
1	6	0	0	0

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS	SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation		
VI. Accurate/Complete Data Collection: Data collection must ensure that the data collected on the PIP indicators are valid and reliable. Validity is an indication of the accuracy of the information obtained. Reliability is an indication of the repeatability or reproducibility of a measurement.		
1. Clearly defined data elements to be collected. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The data elements collected were defined.
2. Clearly identified sources of data. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The sources of data were specified.
3. A clearly defined and systematic process for collecting data that includes how baseline and remeasurement data will be collected. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The process for collecting administrative data was not completely explained. Rereview April 2007 The resubmission included a defined and systematic process for collecting data. This evaluation element was changed from Partially Met to Met.
4. A timeline for the collection of baseline and remeasurement data. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	A timeline for the collection of data was included.

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS	SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation		
VI. Accurate/Complete Data Collection: Data collection must ensure that the data collected on the PIP indicators are valid and reliable. Validity is an indication of the accuracy of the information obtained. Reliability is an indication of the repeatability or reproducibility of a measurement.		
5. Qualified staff and personnel to abstract manual data.	<input type="checkbox"/> Met <input checked="" type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	<p>Training of the manual data collection staff was not discussed. A training summary was provided; however, it appeared to be a training for clinicians on how to screen for mania, and not a training for the staff collecting the manual data.</p> <p>Rereview April 2007 The resubmission included additional information about training for clinicians who complete intakes. The data collection staff members were described as "QI Director, QI Research Coordinator, other QI staff, and contractor." The HSAG PIP review team was looking for a description of the training for manual data collection staff members or the reviewers on how to complete the manual data collection tool. Additionally, the qualifications of manual data collection staff members should be provided. Based on the rereview, this score did not change.</p>
C* 6. A manual data collection tool that ensures consistent and accurate collection of data according to indicator specifications.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The manual data collection tool was included and ensured consistent, accurate collection of data.
7. A manual data collection tool that supports interrater reliability.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The manual data collection tool supported interrater reliability (IRR). The IRR process was discussed in the PIP study report.

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION ELEMENTS	SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation		
VI. Accurate/Complete Data Collection: Data collection must ensure that the data collected on the PIP indicators are valid and reliable. Validity is an indication of the accuracy of the information obtained. Reliability is an indication of the repeatability or reproducibility of a measurement.		
8. Clear and concise written instructions for completing the manual data collection tool.	<input type="checkbox"/> Met <input checked="" type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	<p>There were instructions included with the PIP study report; however, the instructions appeared to be an explanation of changes made to the tool since the last submission and what was considered a "hit." They did not appear to be instructions for the data abstractors completing the manual data collection tool.</p> <p>Rereview April 2007 The HSAG PIP review team was looking for detailed, step-by-step instructions explaining how to complete the manual data collection tool. The instructions should include guidance on how to complete each section of the form and how to handle situations not covered in the instructions. For example, if a reviewer found a partially completed MDQ in the chart, would the reviewer answer "yes" or "no" on the tool? Based on the rereview, this score did not change.</p>
9. An overview of the study in written instructions.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	<p>An overview of the study was not included in the written instructions.</p> <p>Rereview April 2007 The resubmission included an overview of the study that would be included with the written instructions for the manual data collection tool. This evaluation element changed from Not Met to Met.</p>

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
VI. Accurate/Complete Data Collection: Data collection must ensure that the data collected on the PIP indicators are valid and reliable. Validity is an indication of the accuracy of the information obtained. Reliability is an indication of the repeatability or reproducibility of a measurement.			
10. Administrative data collection algorithms/flow charts that show activities in the production of indicators.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A		A partial explanation of the administrative data collection process was provided. An algorithm, flow chart, or narrative description of the complete administrative data collection process should be provided. Rereview April 2007 The resubmission included an algorithm for administrative data collection. This evaluation element was changed from Partially Met to Met.
11. An estimated degree of administrative data completeness. Met = 80 - 100% Partially Met = 50 - 79% Not Met = <50% or not provided	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A		An estimated degree of administrative data completeness was not provided. Rereview April 2007 The resubmission included an estimated degree of administrative data completeness of 99 percent and the process used for that determination. This score was changed from Not Met to Met.

Results for Activity VI				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
1	9	2	0	0

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS	SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation		
VII. Appropriate Improvement Strategies: Real, sustained improvements in care result from a continuous cycle of measuring and analyzing performance, and developing and implementing systemwide improvements in care. Interventions are designed to change behavior at an institutional, practitioner, or consumer level.		
1. Related to causes/barriers identified through data analysis and quality improvement processes. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The interventions were not related to causes/barriers identified through data analysis or quality improvement processes. Rereview April 2007 The resubmission included a fishbone diagram that linked the interventions to causes/barriers. This evaluation element was changed from Not Met to Met.
2. System changes that are likely to induce permanent change.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The interventions were system changes that were likely to induce permanent change.
3. Revised if the original interventions were not successful.	<input type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input checked="" type="checkbox"/> N/A	There were improvements in the rates of adults and children screened.
4. Standardized and monitored if interventions were successful.	<input type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input checked="" type="checkbox"/> N/A	The PIP was not to the point of standardizing interventions at the time of the submission.

Results for Activity VII				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
0	2	0	0	2

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
VIII. Sufficient Data Analysis and Interpretation: Describe the data analysis process on the selected clinical or nonclinical study indicators. Include the statistical analysis techniques used.			
C*	1. Is conducted according to the data analysis plan in the study design. N/A is not applicable to this element for scoring.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Data analysis was conducted according to the data analysis plan.
C*	2. Allows for the generalization of results to the study population if a sample was selected. If no sampling was performed, this element is scored N/A.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The data analysis allowed for generalization of the results to the study population.
	3. Identifies factors that threaten internal or external validity of findings.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Factors that affected the internal or external validity of findings were discussed.
	4. Includes an interpretation of findings.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	An interpretation of the findings was included.
	5. Is presented in a way that provides accurate, clear, and easily understood information.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The data were presented in an accurate and easily understood way.
	6. Identifies initial measurement and remeasurement of study indicators.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Initial measurement and remeasurement of all study indicators were identified.
	7. Identifies statistical differences between initial measurement and remeasurement.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Statistical differences between measurements were identified; however, the chi-square for comparison of Group C youth was incorrect. Rereview April 2007 BHI updated the numerators and denominators for the first remeasurement of Group C adults and youth. The resubmitted p values and chi-square were correct. This evaluation element was changed from Partially Met to Met.

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS	SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation		
VIII. Sufficient Data Analysis and Interpretation: Describe the data analysis process on the selected clinical or nonclinical study indicators. Include the statistical analysis techniques used.		
8. Identifies factors that affect the ability to compare initial measurement with remeasurement.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Factors that affected the ability to compare measurements were not identified. Rereview April 2007 The resubmission included factors that affected the ability to compare measurements. This evaluation element was changed from Not Met to Met.
9. Includes interpretation of the extent to which the study was successful.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	An interpretation of the extent to which the study was successful was included.

Results for Activity VIII				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
2	9	0	0	0

* "C" in this column denotes a critical evaluation element.

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
IX.	Real Improvement Achieved: Describe any meaningful change in performance observed and demonstrated during baseline measurement. Discuss any random year-to-year variation, population changes, and sampling error that may have occurred during the measurement process.		
1.	Remeasurement methodology is the same as baseline methodology.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Remeasurement methodology was the same as baseline methodology.
2.	There is documented improvement in processes or outcomes of care.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	There was documented improvement in processes and adults and youth screened for Bipolar disorder with the MDQ and YMRS-P tools.
3.	The improvement appears to be the result of planned intervention(s).	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	The improvement appeared to be the result of the interventions.
4.	There is statistical evidence that observed improvement is true improvement.	<input checked="" type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Statistical differences between measurements were identified; however, the chi-square for comparison of Group C youth was incorrect. The rate of Group C youth screened did not achieve statistically significant improvement from Baseline to Remeasurement 1. Rereview April 2007 New calculations were included in the resubmission and statistically significant improvement was achieved for Group C youth. This evaluation element was changed from Partially Met to Met.

Results for Activity IX				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
0	4	0	0	0

** This number is a tally of the total number of critical evaluation elements for this review activity.

EVALUATION ELEMENTS		SCORING	COMMENTS
Performance Improvement Project/Health Care Study Evaluation			
X.	Sustained Improvement Achieved: Describe any demonstrated improvement through repeated measurements over comparable time periods. Discuss any random year-to-year variation, population changes, and sampling error that may have occurred during the remeasurement process.		
1.	Repeated measurements over comparable time periods demonstrate sustained improvement, or that a decline in improvement is not statistically significant.	<input type="checkbox"/> Met <input type="checkbox"/> Partially Met <input type="checkbox"/> Not Met <input type="checkbox"/> N/A	Not assessed. BHI had only collected baseline and one remeasurement at the time of the submission.

Results for Activity X				
# of Elements				
Critical Elements**	Met	Partially Met	Not Met	Not Applicable
0	0	0	0	0

** This number is a tally of the total number of critical evaluation elements for this review activity.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

**Table A-1—FY 06-07 PIP Validation Report Scores:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.**

Review Activity		Total Possible Evaluation Elements (Including Critical Elements)	Total Met	Total Partially Met	Total Not Met	Total N/A	Total Possible Critical Elements	Total Critical Elements Met	Total Critical Elements Partially Met	Total Critical Elements Not Met	Total Critical Elements N/A
I.	Appropriate Study Topic	6	6	0	0	0	1	1	0	0	0
II.	Clearly Defined, Answerable Study Question	2	2	0	0	0	1	1	0	0	0
III.	Clearly Defined Study Indicator(s)	7	6	0	0	1	3	3	0	0	0
IV.	Use a representative and generalizable study population	3	3	0	0	0	2	2	0	0	0
V.	Valid Sampling Techniques	6	6	0	0	0	1	1	0	0	0
VI.	Accurate/Complete Data Collection	11	9	2	0	0	1	1	0	0	0
VII.	Appropriate Improvement Strategies	4	2	0	0	2	0	No Critical Elements			
VIII.	Sufficient Data Analysis and Interpretation	9	9	0	0	0	2	2	0	0	0
IX.	Real Improvement Achieved	4	4	0	0	0	0	No Critical Elements			
X.	Sustained Improvement Achieved	1		Not Assessed			0	No Critical Elements			
Totals for All Activities		53	47	2	0	3	11	11	0	0	0

**Table A-2—FY 06-07 PIP Validation Report Overall Scores:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.**

Percentage Score of Evaluation Elements Met*	96%
Percentage Score of Critical Elements Met**	100%
Validation Status***	Met

- * The percentage score is calculated by dividing the total Met by the sum of the total Met, Partially Met, and Not Met.
- ** The percentage score of critical elements Met is calculated by dividing the total critical elements Met by the sum of the critical elements Met, Partially Met, and Not Met.
- *** Met equals confidence/high confidence that the PIP was valid.
Partially Met equals low confidence that the PIP was valid.
Not Met equals reported PIP results that were not credible.

*Section 4: Colorado FY 06-07 PIP Validation Tool:
Screening for Bipolar Disorder
for Behavioral HealthCare, Inc.*

EVALUATION OF THE OVERALL VALIDITY AND RELIABILITY OF PIP/STUDY RESULTS

HSAG assessed the implications of the study's findings on the likely validity and reliability of the results based on CMS protocols. HSAG also assessed whether the State should have confidence in the reported PIP findings. Determining when an accumulation of threats to validity and reliability, and PIP design problems, reach a point at which the PIP findings are no longer credible is always a judgment call.

***Met = Confidence/high confidence in reported PIP results**

****Partially Met = Low confidence in reported PIP results**

*****Not Met = Reported PIP results not credible**

Summary of Aggregate Validation Findings

* **Met**

** **Partially Met**

*** **Not Met**

Summary statement on the validation findings:

Activities I through IX were assessed for this PIP Validation Report. Based on the validation of this PIP study, HSAG's assessment determined high confidence in the results.

Introduction

The appendices consist of documentation supporting the validation process conducted by HSAG using the CMS Protocol for validating PIPs. Appendix A is the study submitted to HSAG for review, Appendix B is CMS rationale for each activity, and Appendix C includes PIP definitions and explanations.

- ◆ Appendix A: **Behavioral HealthCare, Inc.**'s PIP Study: *Screening for Bipolar Disorder*
- ◆ Appendix B: CMS Rationale by Activity
- ◆ Appendix C: Definitions and Explanations by Activity



**Appendix A: PIP Summary Form:
Screening for Bipolar Disorder
for Behavioral Health Care, Inc.**

DEMOGRAPHIC INFORMATION	
BHO Name and ID:	<u>BHI</u>
Study Leader Name:	<u>Ann Terrill Torrez</u> Title: <u>Director, Quality Improvement</u>
Telephone Number:	<u>303-617-2815</u> E-mail Address: <u>ann_torrez@bhiinc.org</u>
Name of Project/Study:	<u>Screening for Bipolar disorder</u>
Type of Study:	<input checked="" type="checkbox"/> Clinical <input type="checkbox"/> Nonclinical
Date of Study Period:	From <u>7/1/04</u> to 6/30/06 <u>12/31/06</u>
<u>9,869 (FY04)</u>	Number of Medicaid Consumers served by BHO
<u>861</u>	Number of Medicaid Consumers in Project/Study
<p align="center">Section to be completed by HSAG</p> <p align="center"> <input type="checkbox"/> Year 1 Validation <input type="checkbox"/> Initial Submission <input type="checkbox"/> Resubmission <input checked="" type="checkbox"/> Year 2 Validation <input type="checkbox"/> Initial Submission <input checked="" type="checkbox"/> Resubmission <input type="checkbox"/> Year 3 Validation <input type="checkbox"/> Initial Submission <input type="checkbox"/> Resubmission </p>	

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

A. Activity I: Choose the Selected Study Topic. Topics selected for study should reflect the Medicaid enrollment in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of the disease. Topics could also address the need for a specific nonclinical service. The goal of the project should be to improve processes and outcomes of health care for the full affected population. The topic may be specified by the State Medicaid agency or on the basis of Medicaid consumer input.

Study Topic:

At this time, there is no cure for bipolar disorder; however, treatment can significantly decrease the associated morbidity and mortality. With early diagnosis and appropriate treatment, Bipolar disorder is a manageable illness like heart disease or diabetes. The person with a bipolar spectrum disorder will require medications as well as psychosocial therapy to obtain the best outcome. As the lag time increases in obtaining accurate diagnosis and treatment, those who have Bipolar disorder report more difficulty with illness management, less confidence about lifelong prognosis, and worry that medications will stop working. (_1_).

Because specific medical treatment for mood stabilization is necessary, failure to provide a medical component or providing inappropriate medical treatment can lead to poor and even fatal outcomes. These include poor response to treatment, unnecessary disability and death by suicide. Suicide risk is extremely high for individuals with Bipolar disorder 20-25% of individuals with bipolar spectrum disorders attempt suicide. At least 19% of deaths among those with bipolar spectrum disorders result from suicide. (_2_) The mortality rate for untreated Bipolar disorder is higher than that for most types of heart disease and some types of cancer. (_3_)

For example, the person with bipolar depression may have suicidal ideation but lack the energy to follow through with his plans. The addition of an antidepressant without first stabilizing mood can lead to the increased energy of mania without the remission of suicidal ideation leaving the person with adequate energy to follow through with plans for suicide.

According to a survey of 600 people conducted by the National Depressive Manic Depressive Association, 35 percent were not correctly diagnosed for 10 or more years. (_1_) Forty-four percent were not correctly diagnosed for between one and 10 years. Sixty percent believed they were misdiagnosed due to a lack of understanding of Bipolar disorder among the professionals they had consulted. Thirty-nine percent felt the professional consulted did not take their symptoms seriously. Thirty-seven percent identified lack of communication between the patient and the professional, and 28 percent had not reported all of their symptoms to the professional. Among this sample of 600 people, the average number of years from onset of illness to seeking help was five to seven years, and the average time from seeking help to accurate diagnosis was four and a half years.

Prevalence of Bipolar Disorder in the general population: --Bipolar spectrum disorders affect 1.1% of the general population and Bipolar II affects 0.6 percent or approximately 2.2 million people in the United States. (_2_)

Bipolar Disorder is a high volume diagnosis for BHI. Prevalence of Bipolar spectrum disorders (Bipolar I, II and NOS) in the BHI consumer served population in 2004 was 984 or 12.6% (741 adults (19.3%), 272 children (6.6%)) out of 8036 with known diagnoses. Bipolar disorder accounted for 16% of children, 30% of adolescents and 32% of adults hospitalized in 2004. Based on data from 2004 and 2005, 59% of consumers evaluated and opened to services at our Mental Health Centers were diagnosed with other disorders in which the differential includes bipolar disorder.

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

A. Activity I: Choose the Selected Study Topic. Topics selected for study should reflect the Medicaid enrollment in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of the disease. Topics could also address the need for a specific nonclinical service. The goal of the project should be to improve processes and outcomes of health care for the full affected population. The topic may be specified by the State Medicaid agency or on the basis of Medicaid consumer input.

There is the potential for significant incidence of missed bipolar diagnosis in the BHI consumer population. In a small study conducted in the fall of 2000, 42 BHI treatment records of individuals identified as diagnosed with major depressive disorder (296.2x, 296.3x) were evaluated. Findings: Seven (17%) records indicated no assessment of history of mania, five (17%) records had documentation of history of mania, mixed symptoms of depression and mania, and two records (5%) had documentation that individual presented with pressured speech during the therapy intake session.

The 2002 APA Practice Guideline for the Treatment of Patients with Bipolar Disorder (Second Edition) states that one way to improve efficiency and increase sensitivity in detecting bipolar disorder is to screen for it, particularly in patients with depression, irritability, or impulsivity. The APA recommends the use of the Mood Disorder Questionnaire, a 13-item, self-report screening instrument for bipolar disorder that has been used successfully in psychiatric clinics and in the general population (_4_).

Beyond depression, there are several disorders that can be misdiagnosed and mistreated if Bipolar is not considered in the differential and screening does not occur. These diagnoses are: Substance induced mood disorder, mood disorder due to a general medical condition, schizophrenia, schizoaffective disorder, major depressive disorder, mood disorder NOS, delusional disorder, psychotic disorder NOS, anxiety disorder, generalized and atypical, dysthymic disorder, cyclothymic disorder, borderline personality disorder, depressive disorder NOS; undifferentiated disruptive disorder, oppositional defiant disorder, and attention deficit/hyperactive disorder. (_5_)

4112 individuals, or 51.2% of BHI consumers in 2004 were diagnosed with one of the above disorders. Because this represents over half of all BHI consumers, the risks for missed or delayed diagnosed are significant. Valid screening tools exist and are recommended by the APA in their treatment guideline, therefore, a focus effort towards improving screening for Bipolar disorder is indicated.

References:

1. Depression and Bipolar Support Alliance, "Living with Bipolar Disorder: How Far Have We Really Come? Constituency Survey", Chicago, IL, 2001.
2. Goodwin FK, and Jamison, KR: Manic-Depressive Illness. New York: Oxford University Press, 1990.
3. Bowden CL: "Update on Bipolar Disorder: Epidemiology, Etiology, Diagnosis, and Prognosis." Medscape Mental Health, 2, no.6 1997. <http://www.medscape.com>.
4. American Psychiatric Association. Practice Guideline for the Treatment of Patients with Bipolar Disorder (Second Edition). Washington, D.C. 2002.
5. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders Fourth Edition. Washington, D.C., 1994

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

B. Activity II: The Study Question. Stating the question(s) helps maintain the focus of the PIP and sets the framework for data collection, analysis, and interpretation.

Study Question:

Does clinician education and provision of screening tools improve incidence of screening for mania in individuals diagnosed at admission by the intake clinician with the following diagnoses: Substance induced mood disorder, mood disorder due to a general medical condition, schizophrenia, schizoaffective disorder, major depressive disorder, mood disorder NOS, delusional disorder, psychotic disorder NOS, anxiety disorder, generalized and atypical, dythymic disorder, cyclothymic disorder, borderline personality disorder, depressive disorder NOS; undifferentiated disruptive disorder, Oppositional defiant disorder, attention deficit/hyperactive disorder.

"Clinician education" is defined as:

1. 4.5 hour class on Bipolar Guidelines
2. Web-based PowerPoint training on Screening for Bipolar disorder and tools

Documentation of adequate screening includes:

Documentation of assessment of the following **seven** DSMIV criteria:

- inflated self -esteem or grandiosity
- decreased need for sleep (feels rested after only 3 hours of sleep)
- more talkative than usual or pressure to keep talking
- flight of ideas or subjective that thoughts are racing
- distractibility
- increase in goal-directed activity or psychomotor agitation
- excessive involvement in pleasurable activities have a high potential for painful consequences

OR

A completed Mood Disorder Questionnaire or Young Mania Rating Scale/Parent's version Screening Tools (**powerpoint slide #13 Screen for Mania Training.ppt**)

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

C. Activity III: Selected Study Indicators. A study indicator is a quantitative or qualitative characteristic or variable that reflects a discrete event (e.g., rates of hospital readmissions within 30 or 90 days), or a status (e.g., percent of consumers reporting that they actively participate in treatment planning) that is to be measured. The selected indicators should be appropriate for the study topic and question as well as track performance or improvement over time. The indicators should be objective, clearly and unambiguously defined, and based on current clinical knowledge or health services research. **Indicators were not solely developed based on practice guidelines, Initially BHI needed to document screening of mania (see Step 2), but each of these items were not found in one document. They were spread among the chart in three different forms, hence not allowing for an adequate screen. During intake, clinicians are checking off a number of boxes on several different forms and would never remember all of the assessment criteria for mania in this format. Clinician need one tool to screen for mania and the MDQ and YMRS were recommended by the APA (see powerpoint training). For purpose of resubmission, study indicators 1 & 2 are being deleted for this PIP since they are not being used to remeasure anything.*

Study Indicator #1:	Adult Screening for Bipolar Disorder *DELETED-see explanation on BHI response form
Numerator:	Number of adult consumer records that have documentation of screening for mania at intake (see definition in Step 2 above) for Bipolar disorder.
Denominator:	Statistically valid sample of adult individuals opened to center during review period with primary diagnosis codes per CCAR: 296.9, 311, 300.4, 301.13, 296.2-296.3x, 295.7, 295, 298.9, 297.1, 300.02, 301.83, 313.81, 314, 312.9, 293.83, 292.84.
First Measurement Period Dates:	7/1/04- 12/31/04
Baseline Benchmark:	N/A — data sets and study methodology for 2000 and 2004 differed so no tests of significance could be performed
Source of Benchmark:	2000 BHI Bipolar pilot study Internal research project.
Baseline Goal:	95%
Study Indicator #2:	Youth Screening for bipolar Disorder *DELETED-see explanation on BHI response form
Numerator:	Number of child or adolescent consumers records that have documentation of screening for mania at intake(see definition in Step 2 above) for Bipolar disorder.
Denominator:	Statistically valid sample of child or adolescent consumers opened to center during review period with primary Diagnosis codes per CCAR:296.9, 311, 300.4, 301.13, 296.2-296.3x, 295.7, 295, 298.9, 297.1, 300.02, 301.83, 313.81, 314, 312.9, 293.83, 292.84.
First Measurement Period Dates:	7/1/04-12/31/04
Benchmark:	N/A — data sets and study methodology for 2000 and 2004 differed so no tests of significance could be performed
Source of Benchmark:	2000 BHI Bipolar pilot study Internal research project.
Baseline Goal:	95%

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

C. Activity III: Selected Study Indicators. A study indicator is a quantitative or qualitative characteristic or variable that reflects a discrete event (e.g., rates of hospital readmissions within 30 or 90 days), or a status (e.g., percent of consumers reporting that they actively participate in treatment planning) that is to be measured. The selected indicators should be appropriate for the study topic and question as well as track performance or improvement over time. The indicators should be objective, clearly and unambiguously defined, and based on current clinical knowledge or health services research. **Indicators were not solely developed based on practice guidelines, Initially BHI needed to document screening of mania (see Step 2), but each of these items were not found in one document. They were spread among the chart in three different forms, hence not allowing for an adequate screen. During intake, clinicians are checking off a number of boxes on several different forms and would never remember all of the assessment criteria for mania in this format. Clinician need one tool to screen for mania and the MDQ and YMRS were recommended by the APA (see powerpoint training). For purpose of resubmission, study indicators 1 & 2 are being deleted for this PIP since they are not being used to remeasure anything.*

Study Indicator #3:	Adult Screening with MDQ Tool
Numerator:	Number of adult individuals screened for Bipolar disorder using the Mood Disorder Questionnaire (MDQ).
Denominator:	Study indicator # 1 numerator Statistically valid sample of adult individuals opened to center during review period with primary diagnosis codes per CCAR: 296.9, 311, 300.4, 301.13, 296.2-296.3x, 295.7, 295, 298.9, 297.1, 300.02, 301.83, 313.81, 314, 312.9, 293.83, 292.84.
First Measurement Period Dates:	7/1/04-12/31/04
Benchmark:	none
Source of Benchmark:	N/A
Baseline Goal:	N/A—none expected
Study Indicator #4:	Youth Screening with MDQ or YMRS-P Tool
Numerator:	Number of child or adolescent individuals screened for Bipolar disorder using the MDQ or Young Mania Rating Scales- Parents (YMRS-P).
Denominator:	Study indicator #2 numerator Statistically valid sample of child or adolescent consumers opened to center during review period with primary Diagnosis codes per CCAR:296.9, 311, 300.4, 301.13, 296.2-296.3x, 295.7, 295, 298.9, 297.1, 300.02, 301.83, 313.81, 314, 312.9, 293.83, 292.84.
First Measurement Period Dates:	7/1/04-12/31/04
Baseline Benchmark:	none
Source of Benchmark:	N/A
Baseline Goal:	N/A—none expected

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

D. Activity IV: Identified Study Population. The study population should be clearly defined to represent the entire population to which the PIP study question and indicators apply. The length of recipient enrollment should be considered and defined. All selection criteria should be listed here. Once the population is identified, a decision must be made whether to review data for the entire population or a sample of that population.

Identified Study Population:

BHI eligible consumers (CCAR MHASA=BH)

All new admissions (CCAR Action Type=1) during study period (Effective Date: 07/1/04- 12/31/04).

All ages,

all enrollment lengths- new or established

Admitted to Arapahoe/Douglas Mental Health Network (ADMHN), Community Reach Center (Reach), Aurora Mental Health Center (AUMHC) (CCAR Agencies= 11, 15, 48)

With a CCAR Primary Psychiatric Diagnosis: code of : 295, 295.7, 292.84, 293.83, 296.2x-296.3x 296.9, 297.1, 298.9, 311, 300.4, 300.02, 301.13, 301.83, 312.9, 313.81, 314.

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

E. Activity V: Sampling Methods. If sampling is to be used to select consumers of the study, proper sampling techniques are necessary to provide valid and reliable information on the quality of care provided. The true prevalence or incidence rate for the event in the population may not be known for the first time a topic is studied. In this case, an estimate should be used and the basis for that estimate indicated. **See E_A 5.doc**

Measure	Sample Error and Confidence Level	Sample Size	Population	Method for Determining Size (<i>describe</i>)	Sampling Method (<i>describe</i>)
*DELETED	.05, 95% CI	91 charts reviewed	149 adults (total of adult consumers opened to center with study dx in six month period)	Where .05 is the acceptable difference between the estimated mean and the population mean as determined by inferential standards. Average standard deviation was estimated to be 1. Formula= $((4 * pop) / (4 + (0.05 * pop)))$ Used NCQA formula for 2006 sample $n = N / (1 + (N * 0.0025))$	from list of MCAID numbers for all adults (18+ yrs) meeting study requirements during study period, randomizing by the first letter of the first name to arrive at sample of 87 (total was determined at the level of population at each of the three MHC's, rather than BHO level generating larger sample size requirements)
*DELETED	.05, 95% CI	93 charts reviewed	158 youth (total of adults consumers opened to center with study dx in six month period)	Where .05 is the acceptable difference between the estimated mean and the population mean as determined by inferential standards. Average standard deviation was estimated to be 1. Formula= $((4 * pop) / (4 + (0.05 * pop)))$	from list of MCAID numbers for all youth (0-17 yrs) meeting study requirements during study period, randomizing by the first letter of the first name to arrive at sample of 91 (total was determined at the level of the population at each of the three MHC's rather than the BHO level, generating larger sample size requirements)
See E_Step5.doc				100 % population	
See E_Step5.doc				100% population	

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

F. Activity Via: Data Collection Procedures. Data collection must ensure that the data collected on the PIP indicators are valid and reliable. Validity is an indication of the accuracy of the information obtained. Reliability is an indication of the repeatability or reproducibility of a measurement.

Data Sources

Hybrid (medical/treatment records and administrative)

Medical/treatment record abstraction

Record Type

Outpatient

Inpatient

Other _____

Other Requirements

Data collection tool attached – (see F_Step6.doc)

Data collection instructions attached

Summary of data collection training attached

IRR process and results attached

Other data

See F_Step6.doc for new data collection tool, instructions for usage, and IRR.

Description of Data Collection Staff

QI Director, QI Research Coordinator, other QI staff and contractor

Administrative data

Data Source

Programmed pull from claims/encounters

Complaint/appeal

Pharmacy data

Telephone service data /call center data

Appointment/access data

Delegated entity/vendor data _____

Other _____

Other Requirements

Data completeness assessment attached - NA

Coding verification process attached

Survey Data

Fielding Method

Personal interview

Mail

Phone with CATI script

Phone with IVR

Internet

Other _____

Other Requirements

Number of waves _____

Response rate _____

Incentives used _____

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

F. Activity VIb: Data Collection Cycle.	Data Analysis Cycle.
<input type="checkbox"/> Once a year <input checked="" type="checkbox"/> Twice a year <input type="checkbox"/> Once a season <input type="checkbox"/> Once a quarter <input type="checkbox"/> Once a month <input type="checkbox"/> Once a week <input type="checkbox"/> Once a day <input type="checkbox"/> Continuous <input type="checkbox"/> Other (list and describe):	<input type="checkbox"/> Once a year <input type="checkbox"/> Once a season <input type="checkbox"/> Once a quarter <input type="checkbox"/> Once a month <input type="checkbox"/> Continuous <input checked="" type="checkbox"/> Other (list and describe): twice a year

F. Activity VIc. Data Analysis Plan and Other Pertinent Methodological Features
<p>BASELINE Data enter the medical/treatment record abstraction results into SPSS data base Run frequency analysis on study indicators 1,2,3,4 variable above Run test to check for statistical differences between n adults and youth. MHC & age group</p> <p>REMEASUREMENT 1 A. Analyze the number of consumers screened between baseline and remeasurement . B. Compute overall number of consumers screened. C. Monthly cumulative of clinicians trained and charts screened by MHC.</p> <p>AD HOC D. Examine the number of consumers screened by clinicians who were trained vs non-trained. E. Number of Positive Screens. F. Number of Positive Screens by Dx. G. Percent Tested Positive Who Got Med Evals.</p>

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

Activity VII. Improvement Strategies. Real, sustained improvements in care result from a continuous cycle of measuring and analyzing performance, and developing and implementing system wide improvements in care. Describe interventions designed to change behavior at an institutional, practitioner, or beneficiary level.

Describe interventions.

Baseline to Remeasurement 1

Presented screening PIP to Provider Advisory Committee and received commitment to conduct screening on new consumers. However, not all MHC's screened all their consumers, (see F_Step6.doc).

Planned Interventions FY06:

Educating clinicians on screening guidelines from Bipolar practice guidelines-- clinicians trainings in April, May, July and August 2005 (see study question (see section B. Step two, Study Question).

Present pilot and baseline study findings at BHI operational and MHC clinical meetings and supervisions and BHI Intranet site Done: Presented findings and a rough draft of the educational PowerPoint presentation to Standards of Practice Committee. Done See attachment B1 and B2.

Conduct a desk-top PowerPoint presentation on screening for Bipolar disorder and screening tools in Summer of 2006. (see folder Final Training Materials).

June – rolled out training at Center A – training completed 9/30/2006 (see G_Step7.doc)

- BHI delivered screening tools (MDQ & YMRS) to all MHCs.

July - rolled out training at Center B – training completed 12/15/2006

August – rolled out training at Center C – training completed 11/27/2006

- Screening began at all MHCs

December – Chart review conducted on new intakes from 8/1/2006 – 10/31/2006

January – Chart review conducted on new intakes from 11/1/2006 – 12/31/2006

Remeasurement 1 to Remeasurement 2- BHI would like to continue this PIP for another year to get another set of remeasurement data on study indicators 3 & 4, and to add more Indicators to address the issue of a consumer with a positive screen going on to a Med Eval. Even though there is a significant difference (see I_step9.xls) in the number of consumers receiving screening, the current process does not allow for the ensurance of the screening tool to get into the right hands. It seems that clinicians who received a positive screen did not get those consumers into a med eval. Ancillary analyses of results revealed that 16 out of 39 (41%) of consumers with a positive screen were referred for psychiatric evaluation/confirmation of Dx. Also, several psychiatric evaluations of consumers with positive screens had no evidence that the screen had been received as part of the psychiatric work-up.

Remeasurement 2 to Remeasurement 3

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

H. Activity VIIIa. Data analysis: Describe the data analysis process in accordance with the analysis plan and any adhoc analysis done on the selected clinical or nonclinical study indicators. Include the statistical analysis techniques utilized and *p* values.

Baseline Measurement

Indicator 1,2: frequency analysis run on number of records where all seven DSMIV elements were assessed (0) Frequency analysis run on presence of completed MDQ or YMRS(1). Two items combined. No difference between adult and youth scores. Low number negated validity of comparing across diagnoses.

Ad Hoc analysis comparing adequacy of CCAR, MMSE and standard intake procedures to capture DSMIV elements (discussed in 8B)

Indicator 3,4: extracted from indicators one and two, frequencies of just MDQ and YMRS. No difference between adult and youth scores. Low number negated validity of comparing across diagnoses

Remeasurement 1 (analysis discussed in 8B)

A. Analyze number of consumers screened between baseline and remeasurement 1 (I_Step9.xls). 2006 data was aggregated by age group within each MHC. Baseline data was also re-aggregated from youth and adult groups to age groups within each MHC for purpose of comparing data and statistical significance.

B. Compute overall number of consumers trained. (see H_Step8.doc)

C. Monthly cumulative of clinicians trained and charts screened by MHC. (see G-Step7.doc)

AD HOC

D. Examine number of consumers screened by clinicians who were trained vs untrained. (see H_Step8.doc)

E. Number of Positive Screens. (see H_Step8.doc)

F. Examined the number of positive screens received by Dx. (see H_Step8.doc)

G. Examined the percentage of those with a positive screen who received a Med Eval. (see H_Step8.doc).

Remeasurement 2

Remeasurement 3

**Appendix A: PIP Summary Form:
Screening for Bipolar Disorder**
for **Behavioral Health Care, Inc.**

H. Activity VIIIb. Interpretation of study results: Describe the results of the statistical analysis, interpret the findings, and discuss the successfulness of the study and indicate follow-up activities. Also, identify any factors that could influence the measurement or validity of the findings. **(See H_Step8.doc)**

Baseline Measurement

Remeasurement 1

Remeasurement 2

Remeasurement 3

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

I. Activity IX. Study Results Summary and Improvement: List study results and describe any meaningful change in performance observed during the time period of analysis. **SEE ATTACHED (I_STEP 9 for Remeasurement on I3 & I4.)**

#1 Quantifiable Measure: *DELETED

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
7/1/04-12/31/04	Baseline:	1	91	1.1%		

#2 Quantifiable Measure: *DELETED

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
7/1/04-12/31/04	Baseline:	1	93	1.1%		

#3 Quantifiable Measure: Adult Screening With MDQ Tool – See I_Step9.xls

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
7/1/04-12/31/04 8/1/2006 – 12/13/2006	Baseline: Recalculated					

#4 Quantifiable Measure: Youth Screening With MDQ or YMRS Tool – See I_Step9.xls

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
7/1/04-12/31/04 8/1/2006 – 12/13/2006	Baseline: Recalculated					

* If used, specify the test, *p* value, and specific measurements (e.g., baseline to remeasurement #1, remeasurement #1 to remeasurement #2, etc., or baseline to final remeasurement) included in the calculations.

Appendix A: PIP Summary Form: Screening for Bipolar Disorder for Behavioral Health Care, Inc.

J. Activity X. Sustained improvement: Describe any demonstrated improvement through repeated measurements over comparable time periods. Discuss any random year-to-year variation, population changes, and sampling error that may have occurred during the remeasurement process.

Cannot show sustained improvement with just 1 remeasure. BHI would like to continue this PIP for another year to get another set of remeasurement data on study indicators 3 & 4, and to add more Indicators to address the issue of a consumer with a positive screen going on to a Med Eval.

PIPs provide a structured method of assessing and improving the processes, and thereby outcomes, of care for the population that a BHO serves. This structure facilitates the documentation and evaluation of improvements in care or service. PIPs are conducted by the BHOs to assess and improve the quality of clinical and nonclinical health care services received by consumers.

The PIP evaluation is based on CMS guidelines as outlined in the CMS publication, *Validating Performance Improvement Projects, A Protocol for Use in Conducting Medicaid External Quality Review Activities*, Final Protocol, Version 1.0, May 1, 2002 (CMS PIP Protocol).

This document highlights the rationale for each activity as established by CMS. The protocols for conducting PIPs can be used to assist the BHOs in complying with requirements.

CMS Rationale

Activity I. Appropriate Study Topic

All PIPs should target improvement in relevant areas of clinical care and nonclinical services. Topics selected for study by Medicaid managed care organizations must reflect the BHO's Medicaid enrollment in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of disease (CMS PIP Protocol, page 2).

Activity II. Clearly Defined, Answerable Study Question

It is important for the BHO to clearly state, in writing, the question(s) the study is designed to answer. Stating the question(s) helps maintain the focus of the PIP and sets the framework for data collection, analysis, and interpretation (CMS PIP Protocol, page 5).

Activity III. Clearly Defined Study Indicator(s)

A study indicator is a quantitative or qualitative characteristic (variable) reflecting a discrete event (e.g., an older adult has/has not received an influenza vaccination in the last 12 months) or a status (e.g., a consumer's blood pressure is/is not below a specified level) that is to be measured.

Each project should have one or more quality indicators for use in tracking performance and improvement over time. All indicators must be objective, clearly and unambiguously defined, and based on current clinical knowledge or health services research. In addition, all indicators must be capable of objectively measuring either consumer outcomes, such as health status, functional status, or consumer satisfaction, or valid proxies of these outcomes.

Indicators can be few and simple, many and complex, or any combination thereof, depending on the study question(s), the complexity of existing practice guidelines for a clinical condition, and the availability of data and resources to gather the data.

Indicator criteria are the set of rules by which the data collector or reviewer determines whether an indicator has been met. Pilot or field testing is helpful in the development of effective indicator criteria. Such testing allows the opportunity to add criteria that might not have been anticipated in the design phase. In addition, criteria are often refined over time based on results of previous studies. However, if criteria are changed significantly, the method for calculating an indicator will not be consistent and performance on indicators will not be comparable over time.

It is important, therefore, for indicator criteria to be developed as fully as possible during the design and field testing of data collection instruments (CMS PIP Protocol, page 5).

Activity IV. Use a Representative and Generalizable Study Population

Once a topic has been selected, measurement and improvement efforts must be systemwide (i.e., each project must represent the entire Medicaid enrolled population to which the PIP study indicators apply). Once that population is identified, the BHO must decide whether to review data for that entire population or use a sample of that population. Sampling is acceptable as long as the samples are representative of the identified population (CMS PIP Protocol, page 8). (See “Activity V. Valid Sampling Techniques.”)

Activity V. Valid Sampling Techniques

If the BHO uses a sample to select consumers for the study, proper sampling techniques are necessary to provide valid and reliable (and therefore generalizable) information on the quality of care provided. When conducting a study designed to estimate the rates at which certain events occur, the sample size has a large impact on the level of statistical confidence in the study estimates. Statistical confidence is a numerical statement of the probable degree of certainty or accuracy of an estimate. In some situations, it expresses the probability that a difference could be due to chance alone. In other applications, it expresses the probability of the accuracy of the estimate. For example, a study may report that a disease is estimated to be present in 35 percent of the population. This estimate might have a 95 percent level of confidence, plus or minus 5 percentage points, implying a 95 percent certainty that between 30 percent and 40 percent of the population has the disease.

The true prevalence or incidence rate for the event in the population may not be known the first time a topic is studied. In such situations, the most prudent course of action is to assume that a maximum sample size is needed to establish a statistically valid baseline for the project indicators (CMS PIP Protocol, page 9).

Activity VI. Accurate/Complete Data Collection

Procedures used by the BHO to collect data for its PIP must ensure that the data collected on the PIP indicators are valid and reliable. Validity is an indication of the accuracy of the information obtained. Reliability is an indication of the repeatability or reproducibility of a measurement. The BHO should employ a data collection plan that includes:

- ◆ Clear identification of the data to be collected.
- ◆ Identification of the data sources and how and when the baseline and repeat indicator data will be collected.
- ◆ Specification of who will collect the data.
- ◆ Identification of instruments used to collect the data.

When data are collected from automated data systems, development of specifications for automated retrieval of the data should be devised. When data are obtained from visual inspection of medical records or other primary source documents, several steps should be taken to ensure the data are consistently extracted and recorded:

1. The key to successful manual data collection is in the selection of the data collection staff. Appropriately qualified personnel, with conceptual and organizational skills, should be used to abstract the data. However, their specific skills should vary depending on the nature of the data collected and the degree of professional judgment required. For example, if data collection involves searching throughout the medical record to find and abstract information or judge whether clinical criteria were met, experienced clinical staff, such as registered nurses, should collect the data. However, if the abstraction involves verifying the presence of a diagnostic test report, trained medical assistants or medical records clerks may be used.
2. Clear guidelines for obtaining and recording data should be established, especially if multiple reviewers are used to perform this activity. The BHO should determine the necessary qualifications of the data collection staff before finalizing the data collection instrument. An abstractor would need fewer clinical skills if the data elements within the data source are more clearly defined. Defining a glossary of terms for each project should be part of the training of abstractors to ensure consistent interpretation among project staff.
3. The number of data collection staff used for a given project affects the reliability of the data. A smaller number of staff members promotes interrater reliability; however, it may also increase the amount of time it takes to complete this task. Intrarater reliability (i.e., reproducibility of judgments by the same abstractor at a different time) should also be considered (CMS PIP Protocol, page 12).

Activity VII. Appropriate Improvement Strategies

Real, sustained improvements in care result from a continuous cycle of measuring and analyzing performance and developing and implementing systemwide improvements in care. Actual improvements in care depend far more on thorough analysis and implementation of appropriate solutions than on any other steps in the process.

An improvement strategy is defined as an intervention designed to change behavior at an institutional, practitioner, or consumer level. The effectiveness of the intervention activity or activities can be determined by measuring the BHO's change in performance, according to predefined quality indicators. Interventions are key to an improvement project's ability to bring about improved health care outcomes. Appropriate interventions must be identified and/or developed for each PIP to ensure the likelihood of causing measurable change.

If repeat measures of quality improvement (QI) indicate that QI actions were not successful (i.e., the QI actions did not achieve significant improvement), the problem-solving process begins again with data analysis to identify possible causes, propose and implement solutions, and so forth. If QI actions were successful, the new processes should be standardized and monitored (CMS PIP Protocol, page 16).

Activity VIII. Sufficient Data Analysis and Interpretation

Review of the BHO data analysis begins with examining the BHO's calculated plan performance on the selected clinical or nonclinical indicators. The review examines the appropriateness of, and the BHO's adherence to, the statistical analysis techniques defined in the data analysis plan (CMS PIP Protocol, page 17).

Activity IX. Real Improvement Achieved

When an BHO reports a change in its performance, it is important to know whether the reported change represents real change, is an artifact of a short-term event unrelated to the intervention, or is due to random chance. The external quality review organization (EQRO) will need to assess the probability that reported improvement is actually true improvement. This probability can be assessed in several ways, but is most confidently assessed by calculating the degree to which an intervention is statistically significant. While this protocol does not specify a level of statistical significance that must be met, it does require that EQROs assess the extent to which any changes in performance reported by an BHO can be found to be statistically significant. States may choose to establish their own numerical thresholds for finding reported improvements to be significant (CMS PIP Protocol, page 18).

Activity X. Sustained Improvement Achieved

Real change results from changes in the fundamental processes of health care delivery. Such changes should result in sustained improvements. In contrast, a spurious, one-time improvement can result from unplanned accidental occurrences or random chance. If real change has occurred, the BHO should be able to document sustained improvement (CMS PIP Protocol, page 19).

Appendix C. Definitions and Explanations by Activity for Behavioral HealthCare, Inc.

This document was developed by HSAG as a resource to assist BHOs in understanding the broad concepts in each activity related to PIPs. The specific concept is delineated in the left column, and the explanations and examples are provided in the right column.

Definitions and Explanations	
Activity I. Appropriate Study Topic	
Broad Spectrum of Care	<ul style="list-style-type: none"> ◆ Clinical focus areas: includes prevention and care of acute and chronic conditions and high volume/high-risk services. High-risk procedures may also be targeted (e.g., care received from specialized centers). ◆ Nonclinical areas: continuity or coordination of care addressed in a manner in which care is provided from multiple providers and across multiple episodes of care (e.g., disease-specific or condition-specific care).
Eligible Population	<ul style="list-style-type: none"> ◆ May be defined as consumers who meet the study topic parameters.
Selected by the State	<ul style="list-style-type: none"> ◆ If the study topic was selected by the state Medicaid agency, this information is included as part of the description under Activity One: Choose the Selected Study Topic in the PIP tool.
Activity II. Clearly Defined, Answerable Study Question	
Study Question	<ul style="list-style-type: none"> ◆ The question(s) directs and maintains the focus of the PIP and sets the framework for data collection, analysis, and interpretation. The question(s) must be measurable and clearly defined. ◆ Examples: <ol style="list-style-type: none"> 1. Does outreach immunization education increase the rates of immunizations for children 0–2 years of age? 2. Does increasing flu immunizations for consumers with chronic asthma impact overall health status? 3. Will increased planning and attention to follow-up after inpatient discharge improve the rate of mental health follow-up services?

Definitions and Explanations	
Activity III. Clearly Defined Study Indicator(s)	
Study Indicator	<ul style="list-style-type: none"> ◆ A quantitative or qualitative characteristic reflecting a discrete event or status that is to be measured. Indicators are used to track performance and improvement over time. ◆ Example: The percentage of enrolled consumers who were 12–21 years of age who had at least one comprehensive well-care visit with a primary care practitioner or an obstetrician-gynecologist during the measurement year.
Sources Identified	<ul style="list-style-type: none"> ◆ Documentation/background information that supports the rationale for the study topic, study question, and indicators. ◆ Examples: HEDIS^{®1} measures, medical community practice guidelines, evidence-based practices, or provider agreements. ◆ Practice guideline examples: American Academy of Pediatrics and American Diabetes Association.
Activity IV. Use a Representative and Generalizable Study Population	
Eligible Population	<ul style="list-style-type: none"> ◆ Refers to consumers who are included in the study. ◆ Includes age, conditions, enrollment criteria, and measurement periods. ◆ Example: the eligible population includes all children ages 0–2 as of December 31 of the measurement period, with continuous enrollment and no more than one enrollment gap of 30 days or less.
Activity V. Valid Sampling Techniques	
True or Estimated Frequency of Occurrence	<ul style="list-style-type: none"> ◆ This may not be known the first time a topic is studied. In this case, assume that a maximum sample size is needed to establish a statistically valid baseline for the study. HSAG will review whether the BHOs defined the impact the topic has on the population or the number of eligible consumers in the population.
Sample Size	<ul style="list-style-type: none"> ◆ Indicates the size of the sample to be used.
Representative Sample	<ul style="list-style-type: none"> ◆ Refers to the sample resembling the entire population.
Confidence Level	<ul style="list-style-type: none"> ◆ Statistical confidence is a numerical statement of the probable degree of certainty or accuracy of an estimate (e.g., 95 percent level of confidence with a 5 percent margin of error).

¹ HEDIS[®] refers to the Health Plan Employer Data and Information Set and is a registered trademark of the National Committee for Quality Assurance (NCQA).

Definitions and Explanations	
Activity VI. Accurate/Complete Data Collection	
Data Elements	<ul style="list-style-type: none"> ◆ Identification of data elements includes unambiguous definitions of data that will be collected (e.g., the numerator/denominator, laboratory values).
Interrater Reliability (IRR)	<ul style="list-style-type: none"> ◆ The HSAG review team evaluates if there is a tool, policy, and/or process in place to verify the accuracy of the data abstracted. Is there an over-read (IRR) process of a minimum-percentage review? ◆ Examples: a policy that includes how IRR is tested, documentation of training, and instruments and tools used.
Algorithms	<ul style="list-style-type: none"> ◆ The development of any systematic process that consists of an ordered sequence of steps. Each step depends on the outcome of the previous step. ◆ The HSAG review team looks for the BHOs to describe the process used in data collection. What are the criteria (e.g., what Current Procedural Terminology and/or source codes were used)?
Data Completeness	<ul style="list-style-type: none"> ◆ For the purposes of PIP scoring, data completeness refers to the degree of complete administrative data (e.g., encounter data or claims data). BHOs that compensate their providers on a fee-for-service basis require a submission of claims for reimbursement. However, providers generally have several months before they must submit the claim for reimbursement, and processing claims by the health plan may take several additional months, creating a claims lag. Providers paid on a capitated or salaried basis do not need to submit a claim to be paid, but should provide encounter data for the visit. In this type of arrangement, some encounter data may not be submitted. ◆ PIPs that use administrative data need to ensure the data has a high degree of data completeness prior to its use. Evidence of data completeness levels may include claim processing lag reports, trending of provider submission rates, policies and procedures regarding timeliness requirements for claims and encounter data submission, encounter data submission studies, and comparison reports of claims/encounter data versus medical record review. Discussion in the PIP should focus on evidence at the time the data was collected for use in identifying the population, sampling and/or calculation of the study indicators. Statements such as, “Data completeness at the time of the data pull was estimated to be 97.8 percent based on claims lag reports (see attached Incurred But Not Reported report),” along with the attachment mentioned, usually (but not always) are sufficient evidence to demonstrate data completeness.

Definitions and Explanations	
Activity VII. Appropriate Improvement Strategies	
Causes and Barriers	<ul style="list-style-type: none"> ◆ Interventions for improvement are identified through evaluation or barrier analysis. If there was no improvement, what problem-solving processes were put in place to identify possible causes and proposed changes to implement solutions? ◆ It is expected that interventions associated with improvement of quality indicators will be system interventions.
Standardized	<ul style="list-style-type: none"> ◆ If the interventions have resulted in successful outcomes, the interventions should continue and the BHO should monitor to assure the outcomes remain. ◆ Examples: if an intervention is the use of practice guidelines, then the BHOs continue to use them; if mailers are a successful intervention, then the BHOs continue the mailings and monitor outcomes.
Activity VIII. Sufficient Data Analysis and Interpretation	
Analysis Plan	<ul style="list-style-type: none"> ◆ Each study should have a plan for how data analysis will occur. ◆ The HSAG review team will ensure that this plan was followed.
Generalization to the Study Population	<ul style="list-style-type: none"> ◆ Study results can be applied to the general population with the premise that comparable results will occur.
Factors that Threaten Internal and External Validity	<ul style="list-style-type: none"> ◆ Did the analysis identify any factors (internal or external) that would threaten the validity of study results? ◆ Example: there was a change in record extraction (e.g., a vendor was hired or there were changes in HEDIS methodology).
Presentation of the Data Analysis	<ul style="list-style-type: none"> ◆ Results should be presented in tables or graphs with measurement periods, results, and benchmarks clearly identified.
Identification of Initial Measurement and Remeasurement of Study Indicators	<ul style="list-style-type: none"> ◆ Clearly identify in the report which measurement period the indicator results reflect.
Statistical Differences Between Initial Measurement and Remeasurement Periods	<ul style="list-style-type: none"> ◆ The HSAG review team looks for evidence of a statistical test (e.g., a t-test, or chi square test).
Identification of the Extent to Which the Study Was Successful	<ul style="list-style-type: none"> ◆ The HSAG review team looks for improvement over several measurement periods. ◆ Both interpretation and analysis should be based on continuous improvement philosophies such that the BHO document data results and what follow-up steps will be taken for improvement.

Definitions and Explanations

Activity IX. Real Improvement Achieved

<p>Remeasurement Methodology Is the Same as Baseline</p>	<ul style="list-style-type: none"> ◆ The HSAG review team looks to see that the study methodology remained the same for the entire study.
<p>Documented Improvement in Processes or Outcomes of Care</p>	<ul style="list-style-type: none"> ◆ The study report should document how interventions were successful in impacting system processes or outcomes. ◆ Examples: there was a change in data collection or a rate increase or decrease demonstrated in graphs/tables.

Activity X. Sustained Improvement Achieved

<p>Sustained Improvement</p>	<ul style="list-style-type: none"> ◆ The HSAG review team looks to see if study improvements have been sustained over the course of the study. This needs to be demonstrated over a period of several (more than two) remeasurement periods.
-------------------------------------	---