Colorado Medicaid
Community Mental Health Services Program

FY 07–08 PIP VALIDATION REPORT

Screening For Bipolar Disorder

for

Behavioral HealthCare, Inc.

May 2008

This report was produced by Health Services Advisory Group, Inc. for the Colorado Department of Health Care Policy & Financing.

HSAG HEALTH SERVICES ADVISORY GROUP

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

HSAG HEALTH SERVICES ADVISORY GROUP

CONTENTS

for **Behavioral HealthCare, Inc.**

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 1-4
Summary of Validation Findings 1-4
Conclusions 1-5
Requirements 1-5
Recommendations 1-5
Comparison of Years 1 Through 3 1-5
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 3-5
Activity VII. Appropriate Improvement Strategies 3-5
Activity VIII. Sufficient Data Analysis and Interpretation 3-6
Activity IX. Real Improvement Achieved 3-8
Activity X: Sustained Improvement Achieved 3-9
4. FY 07–08 PIP Validation Tool 4-1

Appendices

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

Overview

The Balanced Budget Act of 1997 (BBA), 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 compliance with requirements set forth in the Code of Federal Regulations (CFR), at 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 improve the screening of adults and children/adolescents for bipolar disorder, and to improve the recommendation and completion of medication evaluations.

Study Topic

Behavioral HealthCare, Inc. (BHI) continued its nonclinical PIP, ***Screening For Bipolar Disorder***, for the fiscal year (FY) 07–08 submission. The topic addressed the CMS' requirement related to access to, and quality of, care. The topic specifically examined whether consumers diagnosed with a mental health disorder were being screened for bipolar disorder, and if medication evaluations were being recommended and conducted for consumers with a positive screening result.

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

Study Methodology

BHI changed its study methodology this year by adding two study indicators. These indicators measured whether or not intake assessments with a positive screening for bipolar disorder resulted in medication evaluations being recommended and conducted. **BHI** used a hybrid method for data collection, which consisted of both administrative data and medical record review.

BHI collected data on four study indicators, which were defined as follows:

- The number of adults screened for bipolar disorder who were screened using the mood disorder questionnaire (MDQ).
- The number of children or adolescents screened for bipolar disorder who were screened using the MDQ or young mania rating scale/parents (YMRS-P).
- The number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation. (Separate data were collected for the adult and child/adolescent populations.)
- The number of positive screens that resulted in a medication evaluation being conducted. (Separate data were collected for the adult and child/adolescent populations.)

Study Results

For the first two study indicators, although two centers had statistically significant declines from the first remeasurement to the second remeasurement, their rates were above the baseline rates. For the second set of study indicators, there were no baseline data reported; however, all but Adult Center B and Youth Center B demonstrated improvement from the first remeasurement to the second remeasurement. One center, Adult Center C, remained at the same percentage. There was statistical evidence that demonstrated improvement was true improvement for some, but not all, indicators. Table 1-1 illustrates the results for each study indicator.

Table 1-1—Study Indicator Results				
Study Indicators	**Center**	**Baseline Results July 1, 2004–December 31, 2004**	**Remeasurement 1 Results August 1, 2006–December 31, 2006**	**Remeasurement 2 Results August 1, 2007–December 31, 2007**
Study Indicator 1: Number of adults screened for bipolar disorder who were screened using the MDQ.	**A**	0%	74%	79%
	B	3%	56%	79.2%
	C	0%	36%	7%
Study Indicator 2: Number of children or adolescents screened for bipolar disorder who were screened using the MDQ or YMRS-P.	**A**	3%	62%	87%
	B	0%	37%	47%
	C	0%	36%	9%
Study Indicator 3: **ADULT** Number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation.	**A-Adult**	**N/A**	89%	93%
	B-Adult	**N/A**	62%	42%
	C-Adult	**N/A**	33%	33%
Study Indicator 3: **YOUTH** Number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation.	**A-Youth**	**N/A**	50%	68%
	B-Youth	**N/A**	56%	38%
	C-Youth	**N/A**	50%	80%
Study Indicator 4: **ADULT** Number of positive screens that resulted in a medication evaluation being conducted.	**A-Adult**	**N/A**	67%	79%
	B-Adult	**N/A**	31%	42%
	C-Adult	**N/A**	0%	67%

Table 1-1—Study Indicator Results				
Study Indicators	**Center**	**Baseline Results July 1, 2004–December 31, 2004**	**Remeasurement 1 Results August 1, 2006–December 31, 2006**	**Remeasurement 2 Results August 1, 2007–December 31, 2007**
Study Indicator 4: **YOUTH** Number of positive screens that resulted in a medication evaluation being conducted.	**A-Youth**	**N/A**	50%	77%
	B-Youth	**N/A**	44%	54%
	C-Youth	**N/A**	50%	100%

Scoring

HSAG validates a total of 10 activities for each PIP. PIP validation takes place annually and 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, 9 activities with a total of 52 elements were validated. Of this number:
  - 49 evaluation elements were *Met*.
  - 1 evaluation element was *Partially Met.*
  - 0 evaluation elements were *Not Met*.
  - 2 evaluation elements were *Not Applicable (NA).*
- 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 *NA.*

The final validation finding for **BHI's** PIP showed an overall score of 98 percent, a critical element score of 100 percent, and *Met* validation status.

Conclusions

The study successfully addressed access to and quality of care related to screening for bipolar disorder and medication evaluations in the behavioral health population.

For this validation cycle, **BHI** collected baseline and two remeasurement periods of data. However, due to changes in the methodology (the addition of two study indicators), sustained improvement could not be determined until the study has reported baseline and at least two annual remeasurement periods of data for these new indicators. True improvement was seen for some, but not all, study indicators.

Requirements

There were no requirements identified during this review.

Recommendations

As the study moves forward, **BHI** should update the effective date for all new admissions in Activity IV to reflect the current year.

The percentage for Study Indicator 3B–Adult for 8/1/2007 through 12/31/02007 (second remeasurement) was reported as 77 percent, while the actual calculated percentage was 79 percent. **BHI** should make this correction.

Comparison of Years 1 Through 3

BHI completed Activities I through VII for the FY 05–06 validation cycle because, at the time of the submission, **BHI** had only completed a baseline measurement. **BHI** received a score of 69 percent for evaluation elements *Met* and 67 percent for critical elements *Met*, with a *Not Met* validation status. HSAG identified many areas for improvement during this validation cycle.

For the FY 06–07 validation cycle, **BHI** progressed through Activity IX. **BHI** collected baseline and first remeasurement data, which showed a statistically significant improvement in the rates of adults, children, and adolescents screened for bipolar disorder from the first year to the second year. **BHI** addressed all the areas of improvement identified by HSAG during the FY 05–06 review, resulting in 96 percent for evaluation elements *Met*, 100 percent for critical elements *Met*, and a *Met* validation status.

For the FY 07–08 validation cycle, **BHI** remained at Activity IX. There was a change to the study methodology by adding two new study indicators. **BHI** progressed to reporting an additional remeasurement year of data for Study Indicators 1 and 2; however, the two additional indicators reported only two measurement periods of data, with no baseline. HSAG determined that sustained improvement could not be determined until there was another measurement period of data reported for these new indicators. The first two study indicators that measured screening for bipolar disorder continued to show overall improvement; however, the declines for Adult Center C and Youth Center C were statistically significant from the first remeasurement to the second remeasurement.

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:

Met	(1) All critical elements were *Met* **and** (2) 80 percent to 100 percent of all critical and noncritical elements were *Met*. No action required.
Partially Met	(1) All critical elements were *Met* and 60 percent to 79 percent of all critical and noncritical elements were *Met* **or** (2) One critical element or more was *Partially Met*. Requires revision and resubmission of the PIP.
Not Met	(1) All critical elements were *Met* and less than 60 percent of all critical and noncritical elements were *Met* **or** (2) One critical element or more was *Not Met*. Requires revision and resubmission of the PIP.
NA	Not Applicable elements (including critical elements if they were not assessed) were removed from all scoring.

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 07–08 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 *NA*	Total Possible Critical Elements	Total Critical Elements *Met*	Total Critical Elements *Partially Met*	Total Critical Elements *Not Met*	Total Critical Elements *NA*
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	5	0	0	2	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	11	0	0	0	1	1	0	0	0
VII. Appropriate Improvement Strategies	4	4	0	0	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	3	1	0	0	No Critical Elements				
X. Sustained Improvement Achieved	1	Not Assessed				No Critical Elements				
Totals for All Activities	**53**	**49**	**1**	**0**	**2**	**11**	**11**	**0**	**0**	**0**

Table 2-2—FY 07–08 Performance Improvement Project Overall Score *for* Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

Percentage Score of Evaluation Elements *Met**	**98%**
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.

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 the CMS rationale for each activity.

The PIP evaluated access to and quality of care and services. **BHI** used four study indicators to collect data and asses outcomes for this study. The study indicators measured adults and children/adolescents who were screened for bipolar disorder, and the intake assessments with positive screenings that recommended and conducted medication evaluations.

Activity I. Appropriate Study Topic

Study Topic

For the FY 07–08 validation cycle, **BHI** continued with ***Screening For Bipolar Disorder*** as its clinical PIP topic. The topic addressed CMS' requirements related to access to, and quality of, care.

Finding(s)

All evaluation elements for this activity were *Met,* including one critical element.

Strength(s)

The study topic reflected high-risk and high-volume conditions, and addressed a broad spectrum of care and services over time. All eligible consumers who met the study criteria were included, and consumers with special health care needs were not excluded. The study topic had the potential to affect consumer health and functional status.

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)

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

Finding(s)

All evaluation elements for this activity were *Met,* including one critical element.

Strength(s)

The study question was answerable; was stated in clear, simple terms; and was in the correct format to meet CMS Protocols.

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)

BHI collected data for four study indicators during this validation cycle:

- The number of adults screened for bipolar disorder who were screened using the mood disorder questionnaire (MDQ).
- The number of children or adolescents screened for bipolar disorder who were screened using the MDQ or young mania rating scale/parents (YMRS-P).
- The number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation. (Separate data were collected for the adult and child/adolescent populations.)

- The number of positive screens that resulted in a medication evaluation being conducted. (Separate data were collected for the adult and child/adolescent populations.)

Finding(s)

Five of the seven evaluation elements were *Met* for this activity, including three critical elements. Two elements were *Not Applicable*, because the study indicators were not nationally recognized measures and were not based on current, evidence-based practice guidelines, pertinent peer review literature, or consensus expert panels.

Strength(s)

The study indicators are well-defined, allowed for the study question to be answered, and measured changes in valid process alternatives.

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

BHI defined the study population as:

- "**BHI** eligible consumers (CCAR MHASA=BH)
- All new admissions (CCAR Action Type=1) during study period (Effective Date: 7/1/04-12/31/04)
- 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."

Finding(s)

All evaluation elements for this activity were *Met*, including two critical elements.

Strength(s)

The method for identifying the eligible population was completely and accurately defined, and included the requirement for length of consumer enrollment in the BHO, and 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)

BHI should update the effective date for all new admissions in Activity IV to reflect the current year.

Activity V. Valid Sampling Techniques

Sampling Technique(s)

For baseline and the first remeasurement, chart review data collection was conducted and was broken into two separate data collection periods: August 2006–October 2006 (review period 1) and November 2006–December 2006 (review period 2). This was done to ensure that a reliable sample size was produced from each mental health center (MHC).

For the second remeasurement, chart review data collection was conducted in January and February 2008 on all 628 charts from new admits to the MHCs. An oversample of 33 charts was reviewed to ensure adequate sample size.

Finding(s)

All evaluation elements for this activity were *Met,* including one critical element.

Strength(s)

The true or estimated frequency of occurrence was considered in the sampling equation. The sample size was provided, and the confidence level was reported as 95 percent with an acceptable margin of error reported as .05 percent. The sampling technique that was used ensured a representative sample of the eligible population and was 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

BHI used a hybrid method for data collection, which consisted of both administrative data and medical record review.

Finding(s)

All evaluation elements were *Met* for this activity, including one critical element.

Strength(s)

The data elements collected were clearly identified, and a defined and systematic process for the collection of baseline and remeasurement data was used. The PIP included documentation on the relevant education, experience, and training for all manual data collection personnel, and the algorithm demonstrating the administrative data collection process was provided. The manual data collection tool ensured consistent and accurate data collection and supported the interrater reliability process.

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 VII. Appropriate Improvement Strategies

Improvement Strategies

New interventions were implemented in May 2007 and again in September 2007. A two-phase training program was provided using a PowerPoint presentation on the importance of bipolar screening and how to use the MDQ and YMRS-P screening tools. The second phase emphasized the need for positive screens to be followed up with medication evaluations to confirm diagnostic changes. Phase II training was incorporated into the daily operations of the MHCs, such as integrating them into the intake assessment packets and into the consumers' electronic medical records.

Finding(s)

All evaluation elements for this activity were *Met.*

Strength(s)

A fishbone diagram was provided that illustrated the quality improvement process; improvement strategies were based on the causal/barrier analysis noted in the fishbone diagram. System changes noted in the PIP were likely to induce permanent change.

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 VIII. Sufficient Data Analysis and Interpretation

Data Analysis and Interpretation

BHI completed data analysis for baseline and the first remeasurement for two study indicators. For the second remeasurement, two additional study indicators were added to measure whether medication evaluations were recommended and conducted on positive screenings. These two study indicators were further broken down into adult and child/adolescent data. Table 3-1 illustrates the results for each study indicator.

Table 3-1—Study Indicator Results

Study Indicators	Center	Baseline Results July 1, 2004–December 31, 2004	Remeasurement 1 Results August 1, 2006–December 31, 2006	Remeasurement 2 Results August 1, 2007–December 31, 2007
Study Indicator 1: Number of adults screened for bipolar disorder who were screened using the MDQ.	**A**	0%	74%	79%
	B	3%	56%	79.2%
	C	0%	36%	7%
Study Indicator 2: Number of children or adolescents screened for bipolar disorder who were screened using the MDQ or YMRS-P.	**A**	3%	62%	87%
	B	0%	37%	47%

Table 3-1—Study Indicator Results				
Study Indicators	**Center**	**Baseline Results July 1, 2004–December 31, 2004**	**Remeasurement 1 Results August 1, 2006–December 31, 2006**	**Remeasurement 2 Results August 1, 2007–December 31, 2007**
	C	0%	36%	9%
Study Indicator 3: **ADULT** Number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation.	**A-Adult**	**N/A**	89%	93%
	B-Adult	**N/A**	62%	42%
	C-Adult	**N/A**	33%	33%
Study Indicator 3: **YOUTH** Number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation.	**A-Youth**	**N/A**	50%	68%
	B-Youth	**N/A**	56%	38%
	C-Youth	**N/A**	50%	80%
Study Indicator 4: **ADULT** Number of positive screens that resulted in a medication evaluation being conducted.	**A-Adult**	**N/A**	67%	79%
	B-Adult	**N/A**	31%	42%
	C-Adult	**N/A**	0%	67%
Study Indicator 4: **YOUTH** Number of positive screens that resulted in a medication evaluation being conducted.	**A-Youth**	**N/A**	50%	77%
	B-Youth	**N/A**	44%	54%
	C-Youth	**N/A**	50%	100%

Finding(s)

All evaluation elements for this activity were *Met*, including two critical elements.

Strength(s)

The data analysis was conducted according to the analysis plan in the study. The data analysis allowed for the generalization of the results to the study population. The PIP identified factors that threatened the internal and external validity of the findings; factors that could affect the ability to compare measurement periods were also discussed.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

The percentage for Study Indicator 3B–Adult for 8/1/2007 through 12/31/02007 (second remeasurement) was reported as 77 percent, while the actual calculated percentage was 79 percent. **BHI** should make this correction.

Activity IX. Real Improvement Achieved

Real Improvement Achieved

For the first two study indicators, although two centers showed statistically significant declines from the first remeasurement to the second remeasurement, their rates were above the baseline rates. For the second set of study indicators, there were no baseline data reported; however, all but Adult Center B and Youth Center B demonstrated improvement from the first remeasurement to the second remeasurement. One center, Adult Center C, remained at the same percentage. There was statistical evidence that demonstrated improvement was true improvement for some, but not all, indicators.

Finding(s)

Three of four evaluation elements for this activity were *Met*. One evaluation element was *Partially Met*.

Strength(s)

The remeasurement methodology was not the same as the baseline methodology; however, **BHI** provided the rationale for the additional study indicators. There was documented improvement in outcomes of care.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

There was statistical evidence that demonstrated improvement was true improvement for some, but not all, study indicators.

Activity X: Sustained Improvement Achieved

Activity X was not assessed for this year's validation cycle. **BHI** progressed to reporting an additional remeasurement year of data for Study Indicators 1 and 2; however, the two additional indicators reported only two measurement periods of data, with no baseline. HSAG determined that sustained improvement could not be determined until there was another measurement period of data reported for these new indicators. The first two study indicators, which measured screening for bipolar disorder, continued to show overall improvement; however, the declines for Adult Center C and Youth Center C were statistically significant from the first remeasurement to the second remeasurement.

Section 4: Colorado FY 07-08 PIP Validation Tool: Screening for Bipolar Disorder for Behavioral HealthCare, Inc.

DEMOGRAPHIC INFORMATION			
Health Plan Name:	Behavioral HealthCare, Inc.		
Study Leader Name:	Melissa Kulasekere	Title:	Program Evaluator
Phone Number:	(303) 627-2015	E-mail Address:	melissa_kulasekere@bhiinc.org
Name of Project/Study:	Screening for Bipolar Disorder		
Type of Study:	Clinical		
Date of Study:	7/1/2004 to 12/31/2007		
Type of Delivery System:	BHO	Number of Medicaid Consumers in BHO:	9,869
		Number of Medicaid Consumers in Study:	1,700
Year 3 Validation:	Resubmission		
Results:	Remeasurement 3		

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). NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study topic reflected a high-risk and high-volume condition.
	2. Is selected following collection and analysis of data. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study topic was selected following the collection and analysis of national and plan-specific data.
	3. Addresses a broad spectrum of care and services (or was selected by the State). The score for this element will be Met or Not Met.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study topic addressed a broad spectrum of care and services over time.
	4. Includes all eligible populations that meet the study criteria. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	All eligible populations that met the study criteria were included.
	5. Does not exclude consumers with special health care needs. The score for this element will be Met or Not Met.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	Consumers with special health care needs were not excluded from the study.
C*	6. Has the potential to affect consumer health, functional status, or satisfaction. The score for this element will be Met or Not Met.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	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.

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. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study question was stated in clear and simple terms, and was in the correct format to meet CMS Protocols.
C*	2. Is answerable. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	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. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study indicators were not completely well-defined with this PIP submission. Study Indicator 3 should actually be Study Indicator 1 at this time, and the numbering of the remaining study indicators should have been updated for this year's submission. It was unclear to HSAG why Study Indicator 1 was still referenced in Study Indicators 3 and 4 in this submission. Also, Study Indicator 3, stated that it was to be measuring adults screened with the mood disorder questionnaire (MDQ) tool; however, in the numerator, young mania rating scale/parents (YMRS-P) was also referenced. For the new study indicators (5 and 6 in this submission), the description/rationale for the study indicators was the same as the numerators. For Study Indicator 5, the numerator should be "the number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation" and for Study Indicator 6, the denominator should have also stated "that resulted in a medication evaluation being recommended." The time period for the new study indicators was four months, and the time period for the first two study indicators

* "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			
			was five months. Study Indicators 1 and 2 can now be removed from the PIP documentation. It was noted that corrections were made with the resubmission of last year's PIP resulting in the Met score; however, these corrections were not carried over for this year's submission. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element has been changed from Partially Met to Met. The corrections to the study indicators have been made and the study indicators are now well-defined.
	2. Are based on current, evidence-based practice guidelines, pertinent peer review literature, or consensus expert panels.	☐ Met ☐ Partially Met ☐ Not Met ☑ NA	The study indicators were not based on current, evidence-based practice guidelines, pertinent peer review literature, or consensus expert panels.
C*	3. Allow for the study question to be answered. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study indicators allowed for the study question to be answered.
	4. Measure changes (outcomes) in health or functional status, consumer satisfaction, or valid process alternatives. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The study indicators measured changes in valid process alternatives.
C*	5. Have available data that can be collected on each indicator. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	There were data available to be collected on each study indicator.

* "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.**		
	6. Are nationally recognized measures such as HEDIS specifications, when appropriate. The scoring for this element will be Met or NA.	☐ Met ☐ Partially Met ☐ Not Met ☑ NA	The study indicators were not nationally recognized measures.
	7. Includes the basis on which the indicator(s) was adopted, if internally developed.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The basis for Study Indicators 3 and 4 was provided; however, the basis for the new study indicators (5 and 6) in this year's submission was not provided. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element has been changed from Partially Met to Met. The basis for all study indicators was provided.

Results for Activity III				
# of Elements				
Critical Elements**	**Met**	**Partially Met**	**Not Met**	**Not Applicable**
3	5	0	0	2

* "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			
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. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The method for identifying the eligible population was completely and accurately defined. Point of clarification: For all new admissions (CCAR Action Type = 1) during study period (effective date: 7/1/04-12/31/04). It appears that the study period date should be updated to reflect this year's PIP submission. Re-review March 2008: After review of the resubmitted PIP documentation, the point of clarification will remain. The point of clarification had not been addressed with the resubmission documents.
	2. Includes requirements for the length of a consumer's enrollment in the BHO.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The method for identifying the eligible population included the requirements for length of consumer enrollment in the BHO.
C*	3. Captures all consumers to whom the study question applies. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The method for identifying the eligible 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.

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.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The true or estimated frequency of occurrence was considered in the sampling equation.
	2. Identify the sample size.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The sample sizes were identified in the PIP documentation.
	3. Specify the confidence level.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The confidence level was reported as 95 percent.
	4. Specify the acceptable margin of error.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The acceptable margin of error was reported as .05 percent.
C*	5. Ensure a representative sample of the eligible population.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	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.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The sampling techniques were in accordance with the 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. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The data elements collected were identified in the PIP documentation.
	2. Clearly identified sources of data. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The sources for data collection were specified as medical record review and administrative data.
	3. A clearly defined and systematic process for collecting data that includes how baseline and remeasurement data will be collected. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	A defined and systematic process for collecting baseline and remeasurement data was discussed in the PIP documentation.
	4. A timeline for the collection of baseline and remeasurement data. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	A timeline for the collection of baseline and remeasurement data was provided.
	5. Qualified staff and personnel to abstract manual data.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The qualifications and training for data collection personnel were provided.
C*	6. A manual data collection tool that ensures consistent and accurate collection of data according to indicator specifications.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	A manual data collection tool that ensured consistent and accurate data collection was included.
	7. A manual data collection tool that supports interrater reliability.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The manual data collection tool supported the interrater reliability process.
	8. Clear and concise written instructions for completing the manual data collection tool.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	Written instructions for completing the manual data collection tool were provided.
	9. An overview of the study in written instructions.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	An overview of the study was included with the training.
	10. Administrative data collection algorithms/flow charts that show activities in the production of indicators.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	An algorithm demonstrating the administrative data collection process was provided with the PIP submission.

* "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.**		
	11. An estimated degree of administrative data completeness. Met = 80 - 100% Partially Met = 50 - 79% Not Met = <50% or not provided	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The estimated degree of administrative data collection was reported as 100 percent for this year's submission. The supporting documentation on how this percentage was derived was also included.

Results for Activity VI				
# of Elements				
Critical Elements**	**Met**	**Partially Met**	**Not Met**	**Not Applicable**
1	11	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			
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. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The interventions undertaken were not related to causes/barriers identified through data analysis and quality improvement processes. Behavioral HealthCare, Inc. (BHI) provided a fishbone diagram upon resubmission last year showing the quality improvement process used to identify barriers. The fishbone diagram was not provided with this year's submission. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element has been changed from Not Met to Met. The PIP included the fishbone diagram illustrating the quality improvement process used to identify the causes/barriers from which improvement strategies were developed.
	2. System changes that are likely to induce permanent change.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The system changes noted in the PIP were likely to induce permanent change.
	3. Revised if the original interventions were not successful.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	New interventions were implemented in May 2007 and again in September 2007.

** 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.**		
4. Standardized and monitored if interventions were successful.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	It appeared that many of the same interventions were ongoing; however, there was no discussion regarding these interventions being standardized and monitored for their continued success. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element has been changed from Partially Met to Met. The resubmitted documentation discussed the standardization of the curriculum training and assessment packets and how this intervention was monitored for ongoing success.

Results for Activity VII				
# 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			
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. NA is not applicable to this element for scoring.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The data analysis was conducted according to the analysis plan in the study.
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 NA.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The data analysis allowed for the generalization of the results to the study population.
	3. Identifies factors that threaten internal or external validity of findings.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The PIP identified factors that threatened the internal and external validity of the findings.
	4. Includes an interpretation of findings.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	An interpretation of the data analysis findings was provided.

* "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.**		
	5. Is presented in a way that provides accurate, clear, and easily understood information.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The data was provided in an accurate, clear, and easily understood format. Point of clarification: The percentage for Study Indicator 3B-Adult for 8/1/2007 through 12/31/2007 (remeasurement 2) was reported as 77 percent, while the actual calculated percentage was 79 percent. BHI should make this correction going forward with the PIP. There were no data presented in attachment step 9 for Study Indicators 5 and 6. The data for these two indicators were in another attachment. It would be more easily understood if the data for all study indicators for each measurement period were presented in the provided data table in Activity IX or in the step 9 spreadsheet. Re-review March 2008: After review of the resubmitted PIP documentation, the point of clarification addressing the miscalculation will remain. The percentage for Study Indicator 3B-Adult for 8/1/2007 through 12/31/2007 (remeasurement 2) was reported as 77 percent, while the actual calculated percentage was 79 percent. BHI should make this correction going forward with the PIP.
	6. Identifies initial measurement and remeasurement of study indicators.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The initial measurement and remeasurement for each study indicator were identified in the PIP documentation.

* "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.**		
7. Identifies statistical differences between initial measurement and remeasurement.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	Statistical testing was performed and statistical differences identified for each study indicator; however, HSAG was unable to replicate some of the Chi-square values and p values. HSAG calculated the following values: For Study Indicator 3A, baseline to remeasurement 1: Chi-square of 25.37 and p value of 0.00000. For Study Indicator 3A, remeasurement 1 to remeasurement 2: Chi-square of 0.333 and p value of 0.563. For Study Indicator 3B, remeasurement 1 to remeasurement 2: Chi-square of 10.48 and p value of 0.0012. For Study Indicator 3C, remeasurement 1 to remeasurement 2: Chi-square 18.435 and p value of 0.00001. For Study Indicator 4A, baseline to remeasurement 1: Chi-square of 24.863 and p value of 0.000000. For Study Indicator 4A, remeasurement 1 to remeasurement 2: Chi-square of 7.8753 and p value of 0.0050. For Study Indicator 4B, baseline to remeasurement 1: p value of 0.00005. For Study Indicator 4B, remeasurement 1 to remeasurement 2: Chi-square of 2.161 and p value 0.1415. For Study Indicator 4C, remeasurement 1 to remeasurement 2: Chi-square of 13.2467 and p value of 0.00027. BHI should provide the information on how it calculated its values and correct as necessary. Also, HSAG noted that the data for Study

* "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			
			Indicators 5 and 6 that were in narrative format in attachment Step 8B did not provide numerators, denominators, or p values. Future submissions of the PIP should provide this information. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element has been changed from Partially Met to Met. The p values provided in the resubmission could be replicated and the numerators, denominators, and p values were provided for all study indicators.
	8. Identifies factors that affect the ability to compare initial measurement with remeasurement.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	Factors that could affect the ability to compare measurement periods were discussed in the PIP documentation.
	9. Includes interpretation of the extent to which the study was successful.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	An interpretation of the extent to which the study was successful was provided.

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.

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.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The remeasurement methodology was not the same as the baseline methodology; however, BHI provided the rationale for the additional two study indicators.
	2. There is documented improvement in processes or outcomes of care.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	There was documented improvement in outcomes of care.
	3. The improvement appears to be the result of planned intervention(s).	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The improvement noted appeared to be the result of the planned interventions.
	4. There is statistical evidence that observed improvement is true improvement.	☐ Met ☑ Partially Met ☐ Not Met ☐ NA	There was statistical evidence that demonstrated improvement was true improvement for some, but not all, indicators. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element will remain Partially Met. There was true improvement demonstrated for some but not all study indicators.

Results for Activity IX				
# of Elements				
Critical Elements**	**Met**	**Partially Met**	**Not Met**	**Not Applicable**
0	3	1	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.	☐ Met ☐ Partially Met ☐ Not Met ☐ NA	Not assessed. For the first two study indicators, although two centers had statistically significant declines from the first remeasurement to the second remeasurement, their rates were above the baseline rates. For the second set of study indicators, there were no baseline data reported; however, all but Adult Center B and Youth Center B demonstrated improvement from the first remeasurement to the second remeasurement. One center, Adult Center C, remained at the same percentage. The HSAG PIP Review Team acknowledges that this PIP is in its third year; however, HSAG determined that sustained improvement could not be assessed for the PIP because Study Indicators 5 and 6 had only two measurement periods with no baseline reported. Sustained improvement cannot be determined until indicators have reported a baseline and at least two annual remeasurement periods of data.

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.

Table 4-1—FY 07-08 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 NA**	**Total Possible Critical Elements**	**Total Critical Elements Met**	**Total Critical Elements Partially Met**	**Total Critical Elements Not Met**	**Total Critical Elements NA**
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	5	0	0	2	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	11	0	0	0	1	1	0	0	0
VII.	Appropriate Improvement Strategies	4	4	0	0	0	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	3	1	0	0	0	No Critical Elements			
X.	Sustained Improvement Achieved	1		Not Assessed			0	No Critical Elements			
Totals for All Activities		**53**	**49**	**1**	**0**	**2**	**11**	**11**	**0**	**0**	**0**

Table 4-2—FY 07-08 PIP Validation Report Overall Scores: Screening for Bipolar Disorder for Behavioral HealthCare, Inc.	
Percentage Score of Evaluation Elements Met*	**98%**
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.

EVALUATION OF THE OVERALL VALIDITY AND RELIABILITY OF PIP 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.

***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

*	**	***
X		
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, HSAG's assessment determined high confidence in the results.

Appendices

for Behavioral HealthCare, Inc.

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 ***BHI*** submitted to HSAG for review, Appendix B is the 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 HealthCare, Inc.

DEMOGRAPHIC INFORMATION			
BHO Name or ID:	BHI		
Study Leader Name:	Melissa Kulasekere	Title:	Program Evaluator
Telephone Number:	303-627-2015	E-Mail Address:	melissa_kulasekere@bhiinc.org
Name of Project/Study:	Screening for Bipolar disorder		
Type of Study:	☐ Clinical ☐ Nonclinical		

9,869 (FY04)	Number of Medicaid Consumers	**Section to be completed by HSAG** _____ Year 1 Validation _____ Initial Submission _____ Resubmission _____ Year 2 Validation _____ Initial Submission _____ Resubmission __X__ Year 3 Validation _____ Initial Submission __X__ Resubmission
1,700	Number of Medicaid Consumers in Study	
		Section to be completed by HSAG _____ Baseline Assessment _____ Remeasurement 1 _____ Remeasurement 2 __X__ Remeasurement 3

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

A. Activity I: Choose the study topic. PIP topics should target improvement in relevant areas of services and reflect the population in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of the disease. Topics may be derived from utilization data (ICD-9 or CPT coding data related to diagnoses and procedures; NDC codes for medications; state HCPC codes for medications, medical supplies, and medical equipment; adverse events; admissions; readmissions; etc.); grievances and appeals data; survey data; provider access or appointment availability data; consumer characteristics data such as race/ethnicity/language; other fee-for-service data; local or national data related to Medicaid risk populations; etc. The goal of the project should be to improve processes and outcomes of health care or services in order to have a potentially significant impact on consumer health, functional status, or satisfaction. The topic may be specified by the State Medicaid agency or CMS and be based on input from consumers. Over time, topics must cover a broad spectrum of key aspects of consumer care and services, including clinical and nonclinical areas, and should include all enrolled populations (i.e., certain subsets of consumers should not be consistently excluded from studies).

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 outcomes include poor response to treatment, as well as 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 plans. The addition of an antidepressant without first stabilizing mood can lead to the increased energy of mania without the remission of suicidal ideation. This leaves 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% were not correctly diagnosed for 10 or more years. (_1_) Forty-four percent were not correctly diagnosed between one and ten 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% 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.

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

A. Activity I: Choose the study topic. PIP topics should target improvement in relevant areas of services and reflect the population in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of the disease. Topics may be derived from utilization data (ICD-9 or CPT coding data related to diagnoses and procedures; NDC codes for medications; state HCPC codes for medications, medical supplies, and medical equipment; adverse events; admissions; readmissions; etc.); grievances and appeals data; survey data; provider access or appointment availability data; consumer characteristics data such as race/ethnicity/language; other fee-for-service data; local or national data related to Medicaid risk populations; etc. The goal of the project should be to improve processes and outcomes of health care or services in order to have a potentially significant impact on consumer health, functional status, or satisfaction. The topic may be specified by the State Medicaid agency or CMS and be based on input from consumers. Over time, topics must cover a broad spectrum of key aspects of consumer care and services, including clinical and nonclinical areas, and should include all enrolled populations (i.e., certain subsets of consumers should not be consistently excluded from studies).

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

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

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. The results showed that 7 (17%) records indicated no assessment of history of mania, 5 (17%) records had documentation of history of mania, mixed symptoms of depression and mania, and 2 (5%) records had documentation that the 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 both 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_)

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

A. Activity I: Choose the study topic. PIP topics should target improvement in relevant areas of services and reflect the population in terms of demographic characteristics, prevalence of disease, and the potential consequences (risks) of the disease. Topics may be derived from utilization data (ICD-9 or CPT coding data related to diagnoses and procedures; NDC codes for medications; state HCPC codes for medications, medical supplies, and medical equipment; adverse events; admissions; readmissions; etc.); grievances and appeals data; survey data; provider access or appointment availability data; consumer characteristics data such as race/ethnicity/language; other fee-for-service data; local or national data related to Medicaid risk populations; etc. The goal of the project should be to improve processes and outcomes of health care or services in order to have a potentially significant impact on consumer health, functional status, or satisfaction. The topic may be specified by the State Medicaid agency or CMS and be based on input from consumers. Over time, topics must cover a broad spectrum of key aspects of consumer care and services, including clinical and nonclinical areas, and should include all enrolled populations (i.e., certain subsets of consumers should not be consistently excluded from studies).

In 2004, 4112 (51.2%) BHI consumers were diagnosed with one of the above disorders. Because this represents over half of all BHI consumers, the risk for missing or delaying diagnoses are significant. Valid screening tools exist and are recommended by the APA in their treatment guideline, therefore, a focused effort towards improving screening for Bipolar disorder is indicated.

References:

1. National Depressive and Manic-Depressive Association. (2001). Living with bipolar disorder: How far have we really come? Constituency survey. Chicago, IL. Retrieved from www.dbsa.org.
2. Goodwin, F. K., & Jamison, K. R. (1990). Manic-depressive illness. New York: Oxford University Press.
3. Bowden, C. L. (1997). Update on bipolar disorder: Epidemiology, etiology, diagnosis, and prognosis. Medscape Mental Health, 2(6). Retrieved from http://wwwmedscape.com.
4. American Psychiatric Association. (2002). Practice guideline for the treatment of patients with bipolar disorder (2nd ed.). Washington, D.C.
5. American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, D.C.

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

B. Activity II: Define the study question(s). 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 HealthCare, Inc.

C. Activity III: Select the 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 an influenza vaccination in the last twelve months), or a status (e.g., a consumer's blood pressure is/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.

Study Indicator 1	**Describe rationale for selection of study indicator:** ~~Adult Screening for Bipolar Disorder *DELETED-see explanation on BHI response form~~ Indicators were not solely developed based on practice guidelines, Initially BHI needed to document screening of mania, 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. For purpose of resubmission, study indicator 1 is being deleted for this PIP since it not being used to remeasure anything.
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~~
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 HealthCare, Inc.

C. Activity III: Select the 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 an influenza vaccination in the last twelve months), or a status (e.g., a consumer's blood pressure is/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.

Study Indicator 2	**Describe rationale for selection of study indicator:** ~~Youth Screening for bipolar Disorder *DELETED-see explanation on BHI response form~~ Indicators were not solely developed based on practice guidelines, Initially BHI needed to document screening of mania, 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. For purpose of resubmission, study indicator 2 is being deleted for this PIP since it not being used to remeasure anything.
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 HealthCare, Inc.

C. Activity III: Select the 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 an influenza vaccination in the last twelve months), or a status (e.g., a consumer's blood pressure is/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.

Study Indicator #3: ***Please note that this should have been Study Indicator #1 for this PIP.***	**Describe rationale for selection of study indicator:** Adult Screening with MDQ Tool
Numerator	Number of adult individuals screened for Bipolar disorder ~~(indicator #1)~~ who were screened using the Mood disorder questionnaire (MDQ).
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
Benchmark	none
Source of Benchmark	N/A
Baseline Goal	N/A—none expected

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

C. Activity III: Select the 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 an influenza vaccination in the last twelve months), or a status (e.g., a consumer's blood pressure is/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.

Study Indicator #4: ***Please note that this should have been Study Indicator #2 for this PIP.***	Youth Screening with MDQ or YMRS-P Tool
Numerator:	Number of child or adolescent individuals screened for Bipolar disorder ~~(indicator #1)~~ who were screened using the Mood disorder questionnaire or Young Mania Rating scale/Parents.
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 HealthCare, Inc.

C. Activity III: Select the 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 an influenza vaccination in the last twelve months), or a status (e.g., a consumer's blood pressure is/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.

Study Indicator #5: ***Please note that this should have been Stud.y Indicator #3 for this PIP.***	Number of intake assessments with positive screens that recommended a medication evaluation. This indicator was developed based on BHI's Psychosocial Bipolar Tx Guideline that if a Bipolar diagnosis is indicated a Med Eval should be referred.
Numerator:	The number of intake assessments (MDQ or YMRS-P) with positive screens that recommended a medication evaluation.
Denominator:	Total number of intake assessments completed for the statistically valid sample of youth and adult 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:	08/01/2006 – 12/31/2006
Baseline Benchmark:	None
Source of Benchmark:	N/A
Baseline Goal:	N/A—none expected
Study Indicator #6: ***Please note that that should have been Study Indicator #4 for this PIP.***	Number of positive screens that resulted in a medication evaluation being conducted. This indicator was developed based on BHI's Psychosocial Bipolar Tx Guideline that if a Bipolar diagnosis is indicated a Med Eval should be referred. In addition, BHI's Phase II training on screening for Bipolar strongly indicates that a positive screen should have a documented Med Eval. The information from the Phase II training was incorporated into the centers' Core Curriculum Training for all intake workers.
Numerator:	Number of positive MDQ or YMRS-P that resulted in a medication evaluation being conducted.
Denominator:	Total number of positive MDQ or YMRS-P in the statistically valid sample of youth and adult 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:	08/01/2006 – 12/31/2006
Baseline Benchmark:	None
Source of Benchmark:	N/A
Baseline Goal:	N/A-None Expected

Use this area for the provision of additional information:

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

D. Activity IV: Use a representative and generalizable study population. The selected topic should represent the entire Medicaid enrolled population, with system wide measurement and improvement efforts to which the study indicators apply. Once the population is identified, a decision must be made whether to review data for the entire population or a sample of that population. The length of a consumer's enrollment needs to be defined in order to meet the study population criteria.

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 HealthCare, Inc.

E. Activity V: Use sound 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 the first time a topic is studied.

Measure	Sample Error and Confidence Level	Sample Size	Population	Method for Determining Size (*describe*)	Sampling Method (*describe*)
#1 *DELETED	~~.05, 95% CI~~				
#2*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)~~
#3 See E_Step5.doc		~~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~~

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

E. Activity V: Use sound 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 the first time a topic is studied.

Measure	Sample Error and Confidence Level	Sample Size	Population	Method for Determining Size (*describe*)	Sampling Method (*describe*)
					~~BHO level, generating larger sample size requirements)~~
#4 See E_Step5.doc			~~100 % population~~		
#5 See E_Step5.doc					
#6 See E_Step5.doc					

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

F. Activity VIa: Use valid and reliable data collection procedures. Data collection must ensure that the data collected on study 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

[X] Hybrid (medical/treatment records and administrative)

[X] Medical/Treatment Record Abstraction

Record Type

[X] Outpatient
[ ] Inpatient
[X] Other _____

Other Requirements

[X] Data collection tool attached
[X] Data collection instructions attached
[X] Summary of data collection training attached
[X] IRR process and results attached

[ ] Other data

Description of data collection staff (include training, experience and qualifications):

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

Re-measurement 2: QI Director, QI Program Evaluator

[X] Administrative Data

Data Source

[ ] Programmed pull from claims/encounters
[ ] Complaint/appeal
[ ] Pharmacy data
[ ] Telephone service data /call center data
[ ] Appointment/access data
[X] Delegated entity/vendor data _____
[X] Other _____

Other Requirements

[ ] Data completeness assessment attached
[ ] 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 HealthCare, Inc.

F. Activity VIb: Determine the data collection cycle.	Determine the data analysis cycle.
[X] Once a year for re-measurement 2 [X] Twice a year for baseline and re-measurement 1. [ ] Once a season [ ] Once a quarter [ ] Once a month [ ] Once a week [ ] Once a day [ ] Continuous [ ] Other (list and describe):	[X] Once a year for re-measurement 2 [ ] Once a season [ ] Once a quarter [ ] Once a month [ ] Continuous [X] Other (list and describe): Twice a year for baseline and re-measurement 1

F. Activity VIc. Data analysis plan and other pertinent methodological features. Complete only if needed.

Estimated percentage degree of administrative data completeness: _____ percent.

Supporting documentation:

BASELINE
Data enter the medical/treatment record abstraction results into SPSS database
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
RE-MEASUREMENT 1
A. Analyze the number of consumers screened between baseline and re-measurement 1.
B. Compute overall number of consumers screened.
C. Monthly cumulative of clinicians trained and charts screened by MHC.
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.

F. Activity VIc. Data analysis plan and other pertinent methodological features. Complete only if needed.

G. Percent Tested Positive Who Got Med Evals.

RE-MEASUREMENT 2

A. Compute overall number of consumers screened.
B. Analyze the number of consumers screened between re-measurement 1 and re-measurement 2.
C. Compute number of Positive Screens.
D. Compute number of Positive Screens by Dx.
E. Compute number of intake assessments with positive screens that recommended a medication evaluation.
F. Analyze the number of intake assessments with positive screens that recommended a medication evaluation between re-measurement 1 and re-measurement 2.
G. Compute number of positive screens that resulted in a medication evaluation being conducted.
H. Analyze the number of positive screens that resulted in a medication evaluation being conducted between re-measurement 1 and re-measurement 2.

AD HOC

I. Compute number of screens for consumers with exclusionary diagnoses.
J. Compute number of positive screens for consumers with exclusionary diagnoses.
K. Compute the number of intake assessments with positive screens for exclusionary diagnoses that recommended a medication evaluation.
L. Compute number of positive screens for exclusionary diagnoses that resulted in a medication evaluation being conducted.

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

G. Activity VIIa: Include improvement strategies (interventions for improvement as a result of analysis). List chronologically the interventions that have had the most impact on improving the measure. Describe only the interventions and provide quantitative details whenever possible (e.g., "Hired four customer service representatives" as opposed to "Hired customer service representatives"). Do not include intervention planning activities.

Date Implemented (MMYY)	Check if Ongoing	Interventions	Barriers That Interventions Address

G. Activity VIIb: Implement intervention and 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. Describe interventions designed to change behavior at an institutional, practitioner, or consumer level.

Baseline to Re-measurement 1

Presented screening PIP to Provider Advisory Committee and received commitment to conduct screening on new consumers. However, not all CMHC'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

G. Activity VIIb: Implement intervention and 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. Describe interventions designed to change behavior at an institutional, practitioner, or consumer level.

(Moved from the section below to allow for description of interventions for Re-Measurement 1 to Re-measurement 2)

BHI would like to continue this PIP for another year to get another set of re-measurement 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.

Re-measurement 1 to Re-measurement 2

The training for re-measurement 1 was a powerpoint presentation on the importance of Bipolar screening and how to use the MDQ and YMRS-P screening tools. This was considered Phase I. A second (Phase II) powerpoint presentation was developed and sent to the centers in Summer 2007. Phase II training was identical to Phase I training in several areas. The major difference was in emphasizing the need for positive screens to be followed-up with medication evaluations to confirm diagnostic changes. Unlike Phase I training, Phase II training was incorporated into the daily operations of the mental health centers. For REACH, only the Brighton site participated in both the trainings and screenings. From here forward whenever there is a reference to REACH, this only includes the Brighton site. All clinicians were trained in the content of the training material on an on-going basis through new employee orientation, team meetings, and one-on-one sessions. The Bipolar screenings were also integrated into the intake assessment packets at all three centers.

In May 2007, ADMHN included a text box to indicate a positive or negative Bipolar screening in the consumer's electronic record on the intake diagnosis tab.

Starting in September 2007, all new ADMHN Bipolar screens to be scanned into the electronic medical records.

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

Baseline Measurement

Indicator 1,2: frequency analysis run on number of records where all seven DSMIV elements were assessed (0) Frequence 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 receievd a Med Eval (see H_Step8.doc).

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

Re-measurement 2

A. Compute overall number of consumers screened (See H_Step8.doc).
B. Analyze the number of consumers screened between re-measurement 1 and re-measurement 2 (See I_Step9.xls).
C. Compute number of Positive Screens (See H_Step8.doc).
D. Compute number of Positive Screens by Dx (See H_Step8.doc).
E. Compute number of intake assessments with positive screens that recommended a medication evaluation (See H_Step8.doc).
F. Analyze the number of intake assessments with positive screens that recommended a medication evaluation between re-measurement 1 and re-measurement 2 (See I_Step9.xls).
G. Compute number of positive screens that resulted in a medication evaluation being conducted (See H_Step8.doc).
H. Analyze the number of positive screens that resulted in a medication evaluation being conducted between re-measurement 1 and re-measurement 2 (see I_Step9.xls).

AD HOC

I. Compute number of screens for consumers with exclusionary diagnoses (See H_Step8.doc).
J. Compute number of positive screens for consumers with exclusionary diagnoses (See H_Step8.doc).
K. Compute the number of intake assessments with positive screens for exclusionary diagnoses that recommended a medication evaluation (See H_Step8.doc).
L. Compute number of positive screens for exclusionary diagnoses that resulted in a medication evaluation being conducted (See H_Step8.doc).

H. Activity VIIIb. Interpretation of study results: Describe the results of the statistical analysis, interpret the findings, 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.

Interpretation of study results:

Address factors that threaten internal or external validity of the findings for each measurement period.

Baseline Measurement:

Remeasurement 1:

Remeasurement 2:

Remeasurement 3:

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

I. Activity IX: Report improvement. Describe any meaningful change in performance observed and demonstrated during baseline measurement.

Quantifiable Measure No. 1: #1 Quantifiable Measure: *DELETED

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance* Test statistic and p-value
7/1/04-12/31/04	***Baseline:***	1	91	1.1%		
	Remeasurement 1					
	Remeasurement 2					
	Remeasurement 3					
	Remeasurement 4					
	Remeasurement 5					

Quantifiable Measure No. 2: *DELETED

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance* Test statistic and p-value
7/1/04-12/31/04	***Baseline:***	1	93	1.1%		
	Remeasurement 1					
	Remeasurement 2					
	Remeasurement 3					
	Remeasurement 4					
	Remeasurement 5					

Appendix A: PIP Summary Form: Screening For Bipolar Disorder *for* Behavioral HealthCare, Inc.

I. Activity IX: Report improvement. Describe any meaningful change in performance observed and demonstrated during baseline measurement.

Quantifiable Measure No. 3: 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* Test statistic and p-value
~~7/1/04-12/31/04~~ 8/1/2006 – 12/31/2006 8/1/2007 – 12/31/2007	***Baseline:*** ***Recalculated 1 & 2***					

#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* Test statistic and p-value
~~7/1/04-12/31/04~~ 8/1/2006 – 12/31/2006 **8/1/2007 – 12/31/2007**	***Baseline:*** ***Recalculated 1 & 2***					

#5 Quantifiable Measure: Number of Intake Assessments with Positive Screens that Recommended a Medication Evaluation – See I_Step9.xls

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance* Test statistic and p-value
8/1/2006 – 12/31/2006 8/1/2007 – 12/31/2007	***Baseline:*** ***Recalculated***					

#6 Quantifiable Measure: Number of Positive Screens that Resulted in a Medication Evaluation being Conducted – See I_Step9.xls

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
8/1/2006 – 12/31/2006 8/1/2007 – 12/31/2007	***Baseline:*** ***Recalculated***					

* 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 HealthCare, Inc.

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

Sustained improvement:

~~Cannot show sustained improvement with just 1 remeasure. BHI would like to continue this PIP for another year to get another set of re-measurement 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.~~

All centers except one (Center C for adults) showed increased screenings. Screenings for Youth for Centers A and C increased significantly. Screenings for Adults at Center B increased significantly.

Clinician trainings on the purpose and use of bipolar screens which have been incorporated into daily programming at the centers have contributed to sustained increase in screenings since baseline.

PIPs provide a structured method of assessing and improving the processes, and thereby the 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 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 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 study 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 members, 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 members.
3. The number of data collection staff members used for a given project affects the reliability of the data. A smaller number of staff members promote 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. The BHO must identify and develop appropriate interventions for each PIP to ensure the likelihood of measurable change.

If repeated measurements 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 a 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 the protocol for this activity does not specify a level of statistical significance that a reported change in performance must meet, it does require that EQROs assess the extent to which any performance changes reported by a BHO can be found to be statistically significant. States may choose to establish their own numerical thresholds for the significance of reported improvements (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.

Concepts	Definitions and Explanations
Activity I. Appropriate Study Topic	
Broad spectrum of care	◆ 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	◆ May be defined as consumers who meet the study population parameters.
Selected by the State	◆ If the study topic was selected by the state Medicaid agency, this information is included as part of the description under Activity I: "Choose the Selected Study Topic" in the PIP Summary Form.
Activity II. Clearly Defined, Answerable Study Question	
Study question	◆ 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: 1. Does educational outreach about immunizations 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?

Concepts	Definitions and Explanations
Activity III. Clearly Defined Study Indicator(s)	
Study indicator	◆ 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	◆ 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	◆ Refers to consumers who are included in the study. ◆ Includes age, conditions, enrollment criteria, and measurement periods. ◆ Example: The eligible population includes all children 0–2 years of age 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	◆ This may not be known the first time a topic is studied. In this case, the BHO should assume the need for a maximum sample size to establish a statistically valid baseline for the study. HSAG will review whether the BHO defined the impact the topic has on the population or the number of eligible consumers in the population.
Sample size	◆ Indicates the size of the sample to be used.
Representative sample	◆ Refers to the sample reflecting the entire population.
Confidence level	◆ 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®** is a registered trademark of the National Committee for Quality Assurance (NCQA).

Concepts	Definitions and Explanations
Activity VI. Accurate/Complete Data Collection	
Data elements	◆ Identification of data elements includes unambiguous definitions of data that will be collected (e.g., the numerator/denominator, laboratory values).
Interrater reliability (IRR)	◆ 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 for the review of a minimum percentage of records? ◆ Examples: A policy that includes how IRR is tested, documentation of training, and instruments and tools used.
Algorithms	◆ 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 expects for the BHO 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	◆ 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 that the data has a high degree of 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.

Concepts	Definitions and Explanations
Activity VII. Appropriate Improvement Strategies	
Causes and barriers	◆ Interventions for improvement are identified through evaluation or barrier analysis. If there is no improvement, what problem-solving processes are 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	◆ If the interventions result in successful outcomes, the interventions should continue and the BHO should monitor them to ensure that the outcomes remain. ◆ Examples: If an intervention is the use of practice guidelines, then the BHO continues to use them. If mailers are a successful intervention, then the BHO continues the mailings and monitors the outcomes.
Activity VIII. Sufficient Data Analysis and Interpretation	
Analysis plan	◆ 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	◆ Study results can be applied to the general population with the premise that comparable results will occur.
Factors that threaten internal and external validity	◆ 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	◆ 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	◆ Clearly identify in the report which measurement period the indicator results reflect.
Statistical differences between initial measurement and remeasurement periods	◆ 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	◆ The HSAG review team looks for improvement over several measurement periods. ◆ Both interpretation and analysis should be based on continuous improvement philosophies, with the BHO documenting data results and the follow-up steps that will be taken for improvement.

Concepts	Definitions and Explanations
Activity IX. Real Improvement Achieved	
Remeasurement methodology is the same as baseline	◆ The HSAG review team looks to see that the study methodology remains the same for the entire study.
Documented improvement in processes or outcomes of care	◆ The study 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	
Sustained improvement	◆ 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.