Colorado Medicaid
Community Mental Health Services Program

FY 07–08 PIP VALIDATION REPORT

Improving Use and Documentation of Clinical Guidelines

for

Foothills Behavioral Health, LLC

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 **Foothills Behavioral Health, LLC**

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 test the effectiveness of **Foothills Behavioral Health, LLC's** (**FBH's**) new procedures for guideline development and training in increasing the documentation of clinical guideline use following the introduction of two new clinical guidelines. The goal was to assess whether the guideline development procedures improved provider documentation and provider perception of the guidelines. These goals were used to develop the indicators for the study.

Study Topic

FBH continued with **Improving Use and Documentation of Clinical Guidelines** as its nonclinical PIP topic for the fiscal year (FY) 07–08 validation cycle. The topic addressed CMS' requirements related to quality of care outcomes, specifically the use of evidence-based practices in guiding the provider's treatment decision-making. The study addressed bipolar disorder and depression, which account for 25 percent of all **FBH's** diagnoses of Medicaid consumers and are considered high-risk conditions.

FBH's study question was: "Do specially designed procedures for guideline development, dissemination, and training:

1. Improve **FBH** Network MHC provider documentation, during the first 6-months of treatment, of key recommendations included in newly developed **FBH** Depression and Bipolar Disorder clinical guidelines?

2. Improve **FBH** Network MHC provider perception of clinical guidelines, how useful, user-friendly, and accessible?"

Study Methodology

FBH had two study indicators:

- Study Indicator 1: "Number of audited medical records that meet criteria for each of the three categories of 'met,' 'partially met,' or 'not met' guideline documentation status."
- Study Indicator 2: "Number of Network MHC provider survey respondents that indicate agreement (rating '1' or '2'), disagreement (rating '3' or '4'), or don't know rating (rating '5') on the Clinical Practice Guideline Survey on items #6, 7, 8,12."

The study population for Study Indicator 1 included: "all **FBH** Members, admitted to the two Network MHCs, between 1/1/05 and 3/31/05 (pre-audit or baseline benchmark) and, for the first and second measure, between 1/1 and 3/31 for subsequent years, who were enrolled in the Medicaid program at the time of admission, with a primary diagnosis of Bipolar disorder (DSM-IV-RE codes 296.0x, 296.40, 296.4x, 296.6x, 296.7, 296.4x, 296.89, 301.13, 296.80) or Depression (DSM-IV-RE codes 296.2x, 296.3x, 300.4, 311), for all ages (no age restriction), and that have a minimum length of service of 6 months after their admission date."

The study population for Study Indicator 2 included: "all Network MHC providers, credentialed to provide services for Medicaid Members."

Study Results

FBH completed data analysis for baseline and for the first and second remeasurements of both study indicators. There was documented improvement from baseline to the first and second remeasurements for both study indicators; however, Study Indicator 2 demonstrated nonstatistically significant declines in performance from the first remeasurement to the second remeasurement period for item 6 (disagree), item 7 (disagree), item 8 (disagree), and item 12 (agree and disagree).

For Study Indicator 2, item 12, providers who "agree" declined from the first remeasurement to the second remeasurement (51.5 percent to 48.3 percent) and providers who "disagree" increased from 38.5 percent to 49.1 percent, while those providers who "don't know" declined from 10 percent to 2.6 percent. Table 1-1 illustrates results for both study indicators.

Table 1-1—Study Indicator Results

Study Indicators	Baseline Results January 1, 2005–October 31, 2005	Remeasurement 1 Results January 1, 2006–October 31, 2006	Remeasurement 2 Results January 1, 2007–October 31, 2007
Study Indicator 1: "Number of audited medical records that meet criteria for each of the three categories of 'met,' 'partially met,' or 'not met' guideline documentation status."	Met: 15% Partially Met: 28.3% Not Met: 56.7%	Met: 50% Partially Met: 30% Not Met: 20%	Met: 48.3% Partially Met: 31.7% Not Met: 20%
Study Indicator 2: "Number of Network MHC provider survey respondents that indicate agreement (rating '1' or '2'), disagreement (rating '3' or '4'), or don't know rating (rating '5') on the Clinical Practice Guideline Survey on items # 6, 7, 8, 12."	**Item 6** Agree 51% Disagree 11.1% Don't Know 37.9% **Item 7** Agree 47.4% Disagree 15.2% Don't Know 37.4% **Item 8** Agree 39.9% Disagree 27.8% Don't Know 32.3% **Item 12** Agree 26.3% Disagree 56.5% Don't Know 16.2%	**Item 6** Agree 75.9% Disagree 4.7% Don't Know 19.4% **Item 7** Agree 71% Disagree 7.1% Don't Know 21.9% **Item 8** Agree 72.2% Disagree 14.8% Don't Know 13% **Item 12** Agree 51.5% Disagree 38.5% Don't Know 10%	**Item 6** Agree 82.8% Disagree 8.6% Don't Know 8.6% **Item 7** Agree 79.3% Disagree 12.1% Don't Know 8.6% **Item 8** Agree 76.3% Disagree 19% Don't Know 4.7% **Item 12** Agree 48.3% Disagree 49.1% Don't Know 2.6%

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, all activities with a total of 53 elements were validated. Of this number:
  - 48 evaluation elements were *Met*.
  - 2 evaluation elements were *Partially Met.*
  - 0 evaluation elements were *Not Met*.
  - 3 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 **FBH's** PIP showed an overall score of 96 percent, a critical element score of 100 percent, and *Met* validation status.

Conclusions

For the FY 07–08 validation cycle, all 10 activities were reviewed for this study. The study addressed quality of care outcomes; specifically, the use of evidence-based practices in guiding the provider's treatment decision-making. **FBH** provided data for both study indicators from baseline to the second remeasurement period. There was sustained improvement for both indicators over comparable time periods, and the declines that occurred were not statistically significant.

Requirements

There were no requirements identified during this review.

Recommendations

Future submissions of the PIP should clearly specify which interventions have been revised, and when they were revised. **FBH** should also include a discussion of how the interventions were standardized and how the interventions will be monitored for success on an ongoing basis.

Statistical testing was performed from baseline to the first remeasurement, from the first remeasurement to the second remeasurement, and from the baseline to the second remeasurement. For future PIP submissions with similar data, an alternate approach would be to test the status of the "Met," "Partially Met," and "Not Met" categories individually with a 2x2 Chi-square test between each measurement period.

For Study Indicator 2, item 12, providers who "agree" declined from the first remeasurement to the second remeasurement (from 51.5 percent to 48.3 percent) and providers who "disagree" increased from 38.5 percent to 49.1 percent, while those providers who "don't know" declined from 10 percent to 2.6 percent. The 2x3 Chi-square test from the first remeasurement to the second remeasurement indicated a statically significant change for all categories as a whole. The change appeared to be the result of providers moving from the "agree" and the "don't know" categories to the "disagree" category. Further assessment and/or intervention changes may be desired.

Comparison of Years 1 Through 3

In FY 05–06, **FBH** completed Activities I through VII, receiving scores of 94 percent for evaluation elements *Met*, 100 percent for critical elements *Met*, and a *Met* validation status. During this period, baseline results were reported. HSAG identified opportunities for improvement in Activity VI—Accurate/Complete Data Collection for **FBH** to address as the study progresses.

For the FY 06–07 validation cycle, **FBH** progressed through Activity IX, receiving scores of 100 percent for evaluation elements *Met*, 100 percent for critical elements *Met*, and a *Met* validation status. During this period, baseline and the first remeasurement results were reported. **FBH** addressed the opportunities for improvement identified by HSAG in Activity VI.

For the FY 07–08 validation cycle, **FBH** progressed through Activity X, receiving scores of 96 percent for evaluation elements *Met*, 100 percent for critical elements *Met*, and a *Met* validation status. During this period, baseline and two remeasurement periods were reported. **FBH** achieved sustained improvement for both study indicators over comparable time periods because the declines in performance in the second remeasurement were not statistically significant.

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 All 10 Activities. Table 2-1 and Table 2-2 show **FBH's** scores based on HSAG's PIP evaluation of ***Improving Use and Documentation of Clinical Guidelines***. Each activity has been reviewed and scored according to HSAG's validation methodology.

Table 2-1—FY 07-08 Performance Improvement Project Scores *for* Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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	6	0	0	1	3	3	0	0	0
IV. Use a Representative and Generalizable Study Population	3	3	0	0	0	2	2	0	0	0
V. Valid Sampling Techniques	6	6	0	0	0	1	1	0	0	0
VI. Accurate/Complete Data Collection	11	9	0	0	2	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	2	2	0	0	No Critical Elements				
X. Sustained Improvement Achieved	1	1	0	0	0	No Critical Elements				
Totals for All Activities	**53**	**48**	**2**	**0**	**3**	**11**	**11**	**0**	**0**	**0**

Table 2-2—FY 07-08 Performance Improvement Project Overall Score *for* Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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

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

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 purpose of the study was to test the effectiveness of the BHO's new procedures for guideline development and training in increasing the documentation of clinical guideline use following the introduction of two new clinical guidelines. The goal was to assess whether the guideline development procedures improved provider documentation and provider perception of the guidelines.

FBH developed two guidelines for implementation. These guidelines were designed to be user-friendly and easily understood. **FBH** then implemented special training for the Network Mental Health Center (NMHC) providers and assessed the effects of this strategy with these two NMHC providers. The external provider network (IPN) was also provided with a copy of the two new guidelines, with a plan to extend the training to the IPN in the future.

Activity I. Appropriate Study Topic

Study Topic

FBH continued with ***Improving Use and Documentation of Clinical Guidelines*** as its nonclinical PIP topic for the fiscal year (FY) 07–08 validation cycle.

Finding(s)

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

Strength(s)

The study topic reflected high-risk, high-volume conditions, and addressed a broad spectrum of care and services over time. All eligible consumers who met the study criteria were included. 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)

FBH's study question was: "Do specially designed procedures for guideline development, dissemination, and training:

1. Improve **FBH** Network MHC provider documentation, during the first 6-months of treatment, of key recommendations included in newly developed **FBH** Depression and Bipolar Disorder clinical guidelines?

2. Improve **FBH** Network MHC provider perception of clinical guidelines, how useful, user-friendly, and accessible?"

Finding(s)

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

Strength(s)

The study question was answerable and was stated in clear, simple terms, maintaining the focus of the study. It was formatted 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)

FBH had two study indicators:

- Study Indicator 1: Number of audited medical records that meet criteria for each of the three categories of "met," "partially met," or "not met" guideline documentation status."
- Study Indicator 2: "Number of Network MHC provider survey respondents that indicate agreement (rating '1' or '2'), disagreement (rating '3' or '4') or don't know rating (rating '5') on the Clinical Practice Guideline Survey on items #6, 7, 8, 12."

Finding(s)

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

Strength(s)

The study indicators were well-defined, objective, and measurable. They measured changes in quality of care related to using documentation of clinical guidelines.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

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

Activity IV. Use a Representative and Generalizable Study Population

Study Population

The study population for Study Indicator 1 included: "all **FBH** Members, admitted to the two Network MHCs, between 1/1/05 and 3/31/05 (pre-audit or baseline benchmark) and, for the first and second measure, between 1/1 and 3/31 for subsequent years, who were enrolled in the Medicaid program at the time of admission, with a primary diagnosis of Bipolar disorder (DSM-IV-RE codes 296.0x, 296.40, 296.4x, 296.6x, 296.7, 296.4x, 296.89, 301.13, 296.80) or Depression (DSM-IV-RE codes 296.2x, 296.3x, 300.4, 311), for all ages (no age restriction), and that have a minimum length of service of 6 months after their admission date."

The study population for Study Indicator 2 included: "all Network MHC providers, credentialed to provide services for Medicaid Members."

Finding(s)

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

Strength(s)

The method for identifying the eligible populations was accurately and completely defined, included the required length of consumer enrollment, 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)

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

Activity V. Valid Sampling Techniques

Sampling Technique(s)

A sample size of 60 was chosen. Random sampling was used, stratifying by MHC, with 30 cases from each NMHC for comparison.

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 identified as 60. The confidence level was reported as 95 percent, with an acceptable margin of error reported as +/-1.25 percent. The sampling technique that was used ensured a representative sample 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

FBH used medical/treatment record abstraction to capture all necessary data elements defined in the PIP study. Survey data were distributed at clinical staff team/program meetings by the MHC trainer. All medical record audit data and survey data were entered into the Statistical Package for the Social Sciences (SPSS) by the quality improvement (QI) data analyst.

Finding(s)

Nine of the 11 evaluation elements were *Met* for this activity, including one critical element. Two elements were *Not Applicable* because administrative data collection was not used for this PIP.

Strength(s)

The data elements that were collected were clearly identified, and a systematic process with a timeline for baseline and remeasurement data collection was provided in the PIP documentation. The PIP included documentation on the relevant education, experience, and training for all manual data collection personnel, and the manual data collection tool ensured consistent and accurate data collection.

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

FBH implemented special training for the NMHC providers and assessed the effects of this strategy and others with two NMHC providers. The IPN was provided with a copy of the two new guidelines, and the plan was to extend the training to the IPN in the future.

Finding(s)

All evaluation elements for this activity were *Met.*

Strength(s)

Improvement strategies were based on a causal/barrier analysis identified through quality improvement processes. System changes noted in the PIP were likely to induce permanent changes.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

Future submissions of the PIP should clearly specify which interventions have been revised, and when they were revised. **FBH** should also include a discussion of how the interventions were standardized and how the interventions will be monitored for success on an ongoing basis.

Activity VIII. Sufficient Data Analysis and Interpretation

Data Analysis and Interpretation

FBH completed data analysis for baseline and the first and second remeasurements for the two study indicators. Table 3-1 illustrates results for both study indicators.

Table 3-1—Study Indicator Results

Study Indicators	Baseline Results January 1, 2005–October 31, 2005	Remeasurement 1 Results January 1, 2006–October 31, 2006	Remeasurement 2 Results January 1, 2007–October 31, 2007
Study Indicator 1: "Number of audited medical records that meet criteria for each of the three categories of 'met,' 'partially met,' or 'not met' guideline documentation status."	Met: 15% Partially Met: 28.3% Not Met: 56.7%	Met: 50% Partially Met: 30% Not Met: 20%	Met: 48.3% Partially Met: 31.7% Not Met: 20%
Study Indicator 2: "Number of Network MHC provider survey respondents that indicate agreement (rating '1' or '2'), disagreement (rating '3' or '4'), or don't know rating (rating '5') on the Clinical Practice Guideline Survey on items # 6, 7, 8, 12."	**Item 6** Agree 51% Disagree 11.1% Don't Know 37.9% **Item 7** Agree 47.4% Disagree 15.2% Don't Know 37.4% **Item 8** Agree 39.9% Disagree 27.8% Don't Know 32.3% **Item 12** Agree 26.3% Disagree 56.5% Don't Know 16.2%	**Item 6** Agree 75.9% Disagree 4.7% Don't Know 19.4% **Item 7** Agree 71% Disagree 7.1% Don't Know 21.9% **Item 8** Agree 72.2% Disagree 14.8% Don't Know 13% **Item 12** Agree 51.5% Disagree 38.5% Don't Know 10%	**Item 6** Agree 82.8% Disagree 8.6% Don't Know 8.6% **Item 7** Agree 79.3% Disagree 12.1% Don't Know 8.6% **Item 8** Agree 76.3% Disagree 19% Don't Know 4.7% **Item 12** Agree 48.3% Disagree 49.1% Don't Know 2.6%

Finding(s)

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

Strength(s)

The data findings were presented in an accurate, clear, and easily understood format. The PIP identified factors that threatened the internal and external validity of the findings. Chi-square testing was used to determine statistical significance between remeasurement periods.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

Statistical testing was performed from baseline to the first remeasurement, from the first remeasurement to the second remeasurement, and from the baseline to the second remeasurement. For future PIP submissions with similar data, an alternate approach would be to test the status of the "Met," "Partially Met," and "Not Met" categories individually with a 2x2 Chi-square test between each measurement period.

Activity IX. Real Improvement Achieved

Real Improvement Achieved

There was documented improvement from baseline to the first and second remeasurement for both study indicators. Study Indicator 2 demonstrated nonstatistically significant declines from the first remeasurement to the second remeasurement period for item 6 (disagree), item 7 (disagree), item 8 (disagree), and item 12 (agree and disagree).

Finding(s)

Two evaluation elements for this activity were *Met* and two evaluation elements were *Partially Met.*

Strength(s)

The improvement noted in the PIP appeared to be the result of planned interventions.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

For Study Indicator 2, item 12, providers who "agree" declined from the first remeasurement to the second remeasurement (from 51.5 percent to 48.3 percent) and providers who "disagree" increased from 38.5 percent to 49.1 percent, while those providers who "don't know" declined from 10 percent to 2.6 percent. The 2x3 Chi-square test from the first remeasurement to the second remeasurement indicated a statically significant change for all categories as a whole. The change appeared to be the result of providers moving from the "agree" and the "don't know" categories to the "disagree" category. Further assessment and/or intervention changes may be desired.

Activity X. Sustained Improvement Achieved

Sustained Improvement Achieved

There was sustained improvement for both study indicators over comparable time periods, and the declines noted were not statistically significant.

Finding(s)

The evaluation element for this activity was *Met.*

Strength(s)

The PIP demonstrated sustained improvement over comparable time periods.

Requirement(s) (for Critical Elements)

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

Recommendation(s) (for Noncritical Elements)

HSAG acknowledges that this is the third year for this PIP. **FBH** may consider monitoring data internally to determine whether improvement can be achieved across all measures.

Section 4: Colorado FY 07-08 PIP Validation Tool: Improving Use and Documentation of Clinical Guidelines for Foothills Behavioral Health, LLC

DEMOGRAPHIC INFORMATION			
Health Plan Name:	Foothills Behavioral Health, LLC		
Study Leader Name:	Barbara Smith	Title:	Dir. Quality Assurance and Performance Improvement
Phone Number:	(303) 432-5952	E-mail Address:	bsmith@fbhcolorado.org
Name of Project/Study:	Improving Use and Documentation of Clinical Guidelines		
Type of Study:	Nonclinical		
Date of Study:	1/1/2007 to 10/31/2007		
Type of Delivery System:	BHO	Number of Medicaid Consumers in BHO: Number of Medicaid Consumers in Study:	3,167 122
Year 3 Validation:	Resubmission		
Results:	Remeasurement 2		

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 high-risk and high-volume conditions.
	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 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 in the study.
	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.
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 simple terms and was formatted 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 well-defined, objective, and measurable.
	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 based on practice guidelines.
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 quality of care related to using documentation of clinical guidelines.
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.
	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 on which each study indicator was developed was provided in the PIP documentation.

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

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

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

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 populations was completely and accurately defined.
	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 populations included the required length of enrollment.
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 populations 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 provided and considered in the sampling equation.
	2. Identify the sample size.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The sample size was identified as 60.
	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 +/- .125.
C*	5. Ensure a representative sample of the eligible population.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The sample size was representative 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 generally accepted principles of research design and 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 clearly 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 identified in the PIP documentation as medical record data and survey 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.
	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 both the baseline and remeasurement data collection was provided.
	5. Qualified staff and personnel to abstract manual data.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The PIP included documentation on the relevant education, experience, and training of all manual data collection personnel to ensure that only qualified staff members were involved with medical record abstraction.
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 was included.
	7. A manual data collection tool that supports interrater reliability.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	A manual data collection tool that supported the interrater reliablity process was included.
	8. Clear and concise written instructions for completing the manual data collection tool.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	Clear and concise written instructions were included with the manual data collection tool.

* "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.**		
	9. An overview of the study in written instructions.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	An overview of the study was included in the instructions for the data audit tool and in the clinical guideline survey.
	10. Administrative data collection algorithms/flow charts that show activities in the production of indicators.	☐ Met ☐ Partially Met ☐ Not Met ☑ NA	Administrative data collection was not used in this PIP.
	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	Administrative data collection was not used in this PIP.

Results for Activity VI				
# of Elements				
Critical Elements**	**Met**	**Partially Met**	**Not Met**	**Not Applicable**
1	9	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			
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 were related to causes/barriers identified through quality improvement processes.
	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	Original interventions were revised based on data analysis. Point of clarification: Future submissions of the PIP should clearly specify which interventions have been revised and when they were revised. Re-review March 2008: After review of the resubmitted PIP documentation, the point of clarification will remain. The resubmitted PIP did not specify which interventions had been revised or when they were revised. Future submissions of the PIP should provide this information.

** 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	The PIP reported that interventions within the first remeasurement period were standardized. Point of clarification: Future submissions of the PIP should include a discussion of how these standardized interventions will continue to be monitored for success. Re-review March 2008: After review of the resubmitted PIP documentation, the point of clarification will remain. The resubmitted PIP did not include a discussion of how interventions were standardized or how they will be monitored on an ongoing basis. Future submissions of the PIP should include this information.

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 PIP.
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 statistical techniques that were used supported 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 findings for each measurement period was provided.
	5.	Is presented in a way that provides accurate, clear, and easily understood information.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The data was presented in a clear, accurate, and easily understood format.
	6.	Identifies initial measurement and remeasurement of study indicators.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The initial measurement and remeasurement were identified for 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		
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 between the initial measurement and remeasurements were discussed. However, for this year's submission, it was unclear to HSAG which category was tested in the table in Activity IX. For example, was "Met", "Not Met", or "Partially Met" tested for Study Indicator 1? Also, the HSAG statistician could not replicate the Chi-square or p values provided by FHN. Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element was changed from Partially Met to Met. Statistical testing was performed from baseline to the first remeasurement, from the first remeasurement to the second remeasurement, and from baseline to the second remeasurement. For future PIP submissions with similar data, an alternate approach to using a 2x3 Chi-square would be to test the status of the "Met," "Partially Met," and "Not Met" categories individually with a 2x2 Chi-square between each measurement period. Individual category changes between measurement periods can be explained in more detail using the 2x2 Chi-square versus the 2x3 Chi-square which compares all category changes together. For example, using a 2x2 Chi-square to compare "Met" status between

* "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			
			baseline (9/60=15%) and the first remeasurement (30/60=50%) would yield a Chi-square value of 16.75 and a p-value of 0.000043.
	8. Identifies factors that affect the ability to compare initial measurement with remeasurement.	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The PIP identified factors that affected the ability to compare measurement periods.
	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	For Study Indicator 1, the PIP reported that there were no changes to the methodology. For Study Indicator 2, the PIP reported that there were changes in the way that the survey was administered (timing and larger number of staff members surveyed).

** 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.**		
	2. There is documented improvement in processes or outcomes of care.	☐ Met ☑ Partially Met ☐ Not Met ☐ NA	There was documented improvement from baseline to the first and second remeasurements for both study indicators; however, Study Indicator 2 demonstrated nonstatistically significant declines in performance from the first remeasurement to the second remeasurement period for item 6 (disagree), item 7 (disagree), item 8 (disagree), and item 12 (agree and disagree). Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element will remain Partially Met. For Study indicator 2, Item 12, providers who "agree" declined from the first remeasurement to the second remeasurement (51.5% to 48.3%) and providers who "disagree" increased from 38.5% to 49.1%, while those providers who "don't know" declined from 10% to 2.6%. The 2x3 Chi-square from the first remeasurement to the second remeasurement indicated a statistically significant change for all categories as a whole. The change appeared to be the result of providers moving from the "agree" and the "don't know" categories to the "disagree" category.
	3. The improvement appears to be the result of planned intervention(s).	☑ Met ☐ Partially Met ☐ Not Met ☐ NA	The improvement noted for both study indicators appeared to be the result of the planned interventions.

** 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.**		
4. There is statistical evidence that observed improvement is true improvement.	☐ Met ☑ Partially Met ☐ Not Met ☐ NA	Study Indicator 2 demonstrated non-statistically significant declines from the first remeasurement to the second remeasurement period for item 6 (disagree), item 7 (disagree), item 8 (disagree), and item 12 (agree and disagree). Re-review March 2008: After review of the resubmitted PIP documentation, the score for this evaluation element will remain Partially Met. For Study indicator 2, Item 12, providers who "agree" declined from the first remeasurement to the second remeasurement (51.5% to 48.3%) and providers who "disagree" increased from 38.5% to 49.1%, while those providers who "don't know" declined from 10% to 2.6%. The 2x3 Chi-square from the first remeasurement to the second remeasurement indicated a statistically significant change for all categories as a whole. The change appeared to be the result of providers moving from the "agree" and the "don't know" categories to the "disagree" category.

Results for Activity IX				
# of Elements				
Critical Elements**	**Met**	**Partially Met**	**Not Met**	**Not Applicable**
0	2	2	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	There was sustained improvement for both study indicators over comparable time periods and the declines noted were not statistically significant.

Results for Activity X				
# of Elements				
Critical Elements**	**Met**	**Partially Met**	**Not Met**	**Not Applicable**
0	1	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: Improving Use and Documentation of Clinical Guidelines for Foothills Behavioral Health, LLC											
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	6	0	0	1	3	3	0	0	0
IV.	Use a representative and generalizable study population	3	3	0	0	0	2	2	0	0	0
V.	Valid Sampling Techniques	6	6	0	0	0	1	1	0	0	0
VI.	Accurate/Complete Data Collection	11	9	0	0	2	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	2	2	0	0	0	No Critical Elements			
X.	Sustained Improvement Achieved	1	1	0	0	0	0	No Critical Elements			
Totals for All Activities		**53**	**48**	**2**	**0**	**3**	**11**	**11**	**0**	**0**	**0**

Table 4-2—FY 07-08 PIP Validation Report Overall Scores: Improving Use and Documentation of Clinical Guidelines for Foothills Behavioral Health, LLC	
Percentage Score of Evaluation Elements Met*	**96%**
Percentage Score of Critical Elements Met**	**100%**
Validation Status***	***Met***

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

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

*** Met equals confidence/high confidence that the PIP was valid.
Partially Met equals low confidence that the PIP was valid.
Not Met equals reported PIP results that were not credible.

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 X were assessed for this PIP Validation Report. Based on the validation of this PIP, HSAG's assessment determined high confidence in the results.

Introduction

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

- Appendix A: **Foothills Behavioral Health, LLC's** PIP Study: ***Improving Use and Documentation of Clinical Guidelines***
- Appendix B: CMS Rationale by Activity
- Appendix C: Definitions and Explanations by Activity

Appendix A: PIP Summary Form:
Improving Use and Documentation of Clinical Guidelines
for Foothills Behavioral Health, LLC

DEMOGRAPHIC INFORMATION			
BHO Name and ID:	Foothills Behavioral Health		
Study Leader Name:	Barbara Smith, PhD, RN	Title:	Director of Quality Assurance and Performance Improvement
Telephone Number:	303.432.5952	E-mail Address:	bsmith@fbhcolorado.org
Name of Project/Study:	Improving Use and Documentation of Clinical Guidelines		
Type of Study:	☐ Clinical	☒ Nonclinical	
Date of Study Period:	From January 1, 2007 through October 31, 2007 **All changes in this report are in red** **Changes 3/31/08 in blue**		

3,167 (open cases in January, 2006) — Number of Medicaid Consumers served by BHO 122 (# eligible consumers as of January 2006) — Number of Medicaid Consumers in Project/Study	**Section to be completed by HSAG** _____ Year 1 Validation _____ Initial Submission _____ Resubmission __X__ Year 2 Validation _____ Initial Submission __X__ Resubmission _____ Year 3 Validation _____ Initial Submission _____ Resubmission
	Section to be completed by HSAG _____ Baseline Assessment _____ Remeasurement 1 __X__ Remeasurement 2 _____ Remeasurement 3

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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: The rationale for having, and using, clinical guidelines is to ensure consumers receive, on a consistent basis, the most current mental health treatment with the best possible outcomes. Guidelines support this goal by providing clinical staff the necessary evidence-based information, in an abbreviated format, to "guide" their treatment decision-making. Because of the potential value to consumers, FBH, which began operations in January, 2005, established an on-going Guideline Subcommittee, charged with the primary responsibility of developing, updating, and disseminating guidelines throughout the FBH provider network. The Guideline Subcommittee's work is reviewed and approved by two FBH committees: Utilization Management and Quality Assessment Performance Improvement.

The Guideline Subcommittee developed, as the first FBH clinical guideline product, two guidelines for the major mental illnesses of depression and bipolar illness (please see all Guideline Attachment 2__.doc). These two diagnoses were chosen as the first set of guidelines for development because they account for approximately 25% of all diagnoses of Medicaid consumers presently in treatment and these two disorders have considerable evidence to guide treatment decision-making and improve outcomes (please see Guideline Attachment 2_references.doc).

Developing clinical guidelines though does not really ensure their use (Morrison, 2004; Sachs & Gaughan, 1999)(please see Guideline Attachment 2_references.doc). Reasons for inconsistent use of clinical guidelines and/or evidence-based practices include inadequate training, guidelines are not "user-friendly," lack of administrative support for guideline use, and lack of staff involvement in guideline development (Morrison, 2004; Sachs & Gaughan, 1999) (Please see Guideline Attachment 2_references.doc). Results of a FBH provider survey, administered to providers at the two Network MHCs, indicated that, although most providers believe that guidelines are useful and important to their practice, less than half found guidelines user-friendly and only 40% indicated that they were readily accessible (Guideline Attachment 1_MHC Provider Guideline Survey.doc). In addition, only about a fourth of respondents indicated that someone explained how guidelines should be used. Finally, although almost two-thirds (61.5%) of survey respondents agreed that it was important to document use of clinical guidelines, a medical record audit, conducted in November, 2005, of 60 medical records, indicated that more than half (56.7%) of the records did not document key guideline recommendations, such as regularly assessing suicide and substance use, that are included in all well-accepted clinical guidelines for Depression and Bipolar disorder (American Psychiatric Association, 2002) (Please see Guideline Attachment 2_references.doc).

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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).
The Guideline Subcommittee, recognizing the above issues/concerns about guidelines and overall problems with documentation of guideline use, focused its attention on developing specific methods and procedures that would improve clinical guideline accessibility, how user-friendly guidelines are, and improve clinical guideline dissemination and use. The study plan was to use the introduction of these two clinical guidelines, within the two Network MHCs, to test the effectiveness of the new procedures for guideline development and training. Although the two clinical guidelines were introduced to the external provider network (IPN) the project team decided to test the new processes within the MHCs first and then develop a similar process for IPN.

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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:

Do specially designed procedures for guideline development, dissemination, and training

1. Improve FBH Network MHC provider documentation, during the first 6-months of treatment, of key recommendations included in newly developed FBH Depression and Bipolar Disorder clinical guidelines?
2. Improve FBH Network MHC provider perception of clinical guidelines, how useful, user-friendly, and accessible?

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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:	The proportion of audited medical records that achieve "met" status (defined as 90% or more of the applicable items on the Audit Checklist tool are rated as "met"), "partially met" status (defined as 70 % to < 90% of the applicable items on the Audit Checklist tool are rated as "met"), or "not met" status (defined as less than 70% of the applicable items on the Audit Checklist are rated as "met"). Please see F. Activity 6b for explanation of criteria for documentation status groups and "Attachment_4 guideline checklist development.doc" for information on the checklist tools.
Numerator:	Number of audited medical records that meet criteria for each of the three categories of "met," "partially met," or "not met" guideline documentation status.
Denominator:	Total number of applicable audited medical records
First Measurement Period Dates:	First measurement period, post guideline implementation and training: January through October, 2006.
Baseline Benchmark:	56.7% of 60 audited medical records grouped as "Not Met" documentation status compared to 28.3% grouped as "partially met," and 15% grouped as "met."
Source of Benchmark:	Medical Record Audit conducted pre clinical guideline implementation and training (November, 2005)
Baseline Goal:	Significantly increase the percent of medical records that meet the documentation status of "Met" and "Partially Met" post guideline implementation and training, compared to the percent in the baseline benchmark

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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:	The percent of Network MHC provider respondents to the Clinical Practice Guideline Survey that agree (response of 1 or 2), disagree (response of 3 or 4), or don't know (response of 5) that clinical guidelines are easily understood (Item #6), user-friendly (item #7), readily accessible (item #8), and understand how to use clinical guidelines (item #12). Please see "Guideline Attachment 1_MHC provider guideline survey.doc" for a description of the survey.
Numerator:	The number of Network MHC provider survey respondents that indicate agreement (rating "1" or "2"), disagreement (rating "3" or "4") or don't know (rating "5") on the Clinical Practice Guideline Survey on items #6,7,8,12.
Denominator:	Total number of Network MHC provider survey respondents to the Clinical Practice Guideline Survey items #6,7,8,12.
First Measurement Period Dates:	May, 2006
Benchmark:	Pre guideline implementation survey results on the FBH Clinical Practice Guideline Survey: #6 with 51% agreement, #7 with 47.7% agreement, #8 with 39.9% agreement; and #12 with 27.6% agreement.
Source of Benchmark:	Survey conducted pre guideline implementation and training with MHC Network Providers
Baseline Goal:	Significantly increase the percent of Network MHC respondents that agree (indicate a "1" or "2") on items #6,7,8,12 of the Clinical Practice Guideline Survey, pre compared to post guideline implementation and training.
Study Indicator #3:	
Numerator:	
Denominator:	
First Measurement Period Dates:	
Benchmark:	
Source of Benchmark:	
Baseline Goal:	

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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.

Identified Study Population: The study population, for indicator #1, includes all FBH Members, admitted to the two Network MHCs, between 1/1/05 and 3/31/05 (pre-audit or baseline benchmark) and, for the first and second measure, between 1/1 and 3/31 for subsequent years, who were enrolled in the Medicaid program at the time of admission, with a primary diagnosis of Bipolar disorder (DSM-IV-TR codes 296.0x, 296.40, 296.4x, 296.6x, 296.5x,296.7,296.4x,296.89,301.13,296.80) or Depression (DSM-IV-TR codes 296.2x,296.3x,300.4,311), all ages (no age restriction), and that have a minimum length of service of 6 months after their admission date. Study population size was n=127 **(baseline)**.

Rationale for above study population parameters:

1. A 6-month treatment period was required to allow adequate time for documenting all checklist guideline items. For example, item #6 on the Bipolar Audit checklist requires a 6-month period to assess and item #7, on both the Bipolar and Depression checklist requires a 6-month Treatment Plan update.
2. New admissions were chosen to create a consistent period of time for documentation in each of the medical record audits, which would take place in November of the respective years. In addition, limiting the study population to an initial 6-8 month treatment period reduces the complexity of the checklist item definitions, which leads to improved checklist reliability, e.g. what to do when there is a change in provider. The goal was to create, through the parameters for the defined population, a consistent study period for documentation of guideline use that can be compared between the measurement periods

The study population, for indicator #2, includes all Network MHC providers, credentialed to provide services for Medicaid Members. At the time of the baseline survey n=488.

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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*)
Study Indicator #1 –	Based on the baseline sample results the sample error = .064. At the 95% CL the CI is +/- .125.	N=60	N=127	A sample of 60 was chosen stratified by MHC so that there were 30 from each. Of the two Network MHCs: 1. Sample size of 60 should meet the chi-square assumption that expected cell frequency be >0; 2. Although the strength of the relationship is unknown, a sample of n=60 allows for a power of .80 with a moderate effect size between .30 and .40.	Random sampling was used, stratifying by MHC – 30 cases from each MHC. With 30 from each MHC comparisons in improvement, between the MHCs, can be made.
Study Indicator #2: All MHC providers who met criteria were eligible.					

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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

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

[☒] Medical/treatment record abstraction

Record Type

[☒] Outpatient
[☐] Inpatient
[☐] Other _____

Other Requirements

[☒] Data collection tool attached (Attachment 3_final chart audit tool guideline.xls)
[☒] Data collection instructions attached (Attachment 5_instructions for clinical guideline audit) **see Attachment 5_revised_instructions for clinical guideline audit.doc (2/26/07)**
[☒] Summary of data collection training attached (Attachment 5_instructions for clinical guideline audit)
[☒] IRR process and results attached (Attachment 4_guideline checklist development.doc)

[☐] Other data

Description of Data Collection Staff **(updated in bold, 2/26/07)**

[☐] Administrative data

Data Source

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

Other Requirements

[☐] Data completeness assessment attached
[☐] Coding verification process attached

[☒] Survey Data (see Appendix A "Guideline Attachment 1_MHC provider guideline survey.doc" for copy of survey)

Fielding Method

[☐] Personal interview
[☐]
[☐] Phone with CATI script
[☐] Phone with IVR
[☐] Internet
[☒] Other distributed at clinical staff team/program meetings by MHC trainer – responses placed by provider in a sealed envelope and returned by the trainer to the FBH QI Dept_____

Other Requirements

[☒] Number of waves one_____

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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.

Included 3 licensed and one unlicensed master's degree senior clinicians, two from each of the Network MHCs, and the FBH Medical Director and Assistant Medical Director. **Required qualifications of auditors: at least one year experience in conducting medical record audits, at least two years experience as clinician documenting services, must be detail oriented, and must complete training for audit. Items on the checklist regarding psychiatric services must be audited by a licensed Prescriber or Psychiatrist.**

Required qualifications of staff distributing surveys: Brief training in procedures for distribution of surveys, ability to follow instructions, and are not involved in the training.

[☒] Response rate _40.6%_____
[☒] Incentives used none_____

F. Activity VIb: Determine the data collection cycle.

[☒] Once a year
[☐] Twice a year
[☐] Once a season
[☐] Once a quarter
[☐] Once a month
[☐] Once a week
[☐] Once a day
[☐] Continuous
[☐] Other (list and describe):

The first Remeasurement will occur in November, 2006; the second remeasurement will occur in November, 2007

Determine the data analysis cycle.

[☒] Once a year
[☐] Once a season
[☐] Once a quarter
[☐] Once a month
[☐] Continuous
[☐] Other (list and describe):

December/January 2006/2007 and December/January 2007/2008

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

Medical Record Audit Analysis

1. All medical record audit data entered into SPSS by the QI Data Analyst.

2. Results for each medical record audit categorized as to the adequacy of guideline documentation, using three groups, "Met," defined as 90% of all applicable guideline items scored as "met" on the checklist tool, "Partially Met," defined as 70% to <90% of applicable items scored as "met" on the checklist tool, and "Not Met," defined as less than 70% of guideline checklist items scored as "met" on the checklist tool. Three groups of documentation status are defined to provide more qualitative information on amount of documentation of guidelines and to show progress in improving documentation.

3. A chi-square test will be used to analyze differences in frequency of documentation status groups between medical records audited before and after implementation and training of the two FBH clinical guidelines. A signficance level of p=.05 will be used. The analyzes will be conducted on all data from the two Checklists (n=60).

Clinical Staff Guideline Survey Analysis:

1. All survey data entered into SPSS by the QI Data Analyst

2. Responses for each of the four survey items will be categorized as "agree," (item rating of "1" or "2"), "disagree" (item rating of "3" or "4"), or don't know (item rating of "5").

3. A chi-square test will be used to analyze differences in frequency of rating categories of "agree," "disagree," or "don't know between survey responses before and after implementation and training of two FBH guidelines. A signficiance level of p=.05 will be used.

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

G. Activity VII. 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.

Describe interventions.

The study plan is to implement special training for the Network MHC providers and assess effects of this strategy and others with just the two Network MHC providers. The external providers (IPN) were provided a copy of the two new guidelines with a description of their use in the Provider Manual. The plan is to use what we learn in the study with the Network MHCs to develop a similar training with the IPN provider network that will be web-based.

Baseline to Remeasurement 1

1. Developed two guidelines for implementation that were designed to be user friendly and easily understood. Guidelines were no more than 2 pages, key areas bolded, recommendations were as specific as possible. A one page "Ten Tips for Recovery" was developed to give to the consumers with specific helpful hints for each of the diagnoses (see all "Guideline Attachment 2_ which includes all components of the two guidelines).

2. Developed specific guideline training procedures for implementing the two new clinical guidelines as well as future guidelines

3. Training on the guidelines was conducted at the program/team level at both Network MHC, reviewing the purpose of the guidelines, specific key elements of each guideline, the importance of documentation specific to the guideline recommendations, tips on how/where to document use of guidelines, and how to use the guidelines in general. In addition, providers were given information on where to find these guidelines and specific plans for future guidelines.

4. Guidelines were posted on JCMH's Intranet, MHCBBC's shared drive, and each provider received a loose-leaf notebook with the two guidelines and Ten Tips in there. The provider could use the notebook for future guidelines.

5. A second clinical guideline training will be provided, in May, 2006, introducing two additional guidelines. During this training the first two guidelines, for Depression and Bipolar disorder, will be reviewed again, reminding providers regarding issues of documenting guideline use, where to find guidelines, and how to use guidelines.

Remeasurement 1 to Remeasurement 2

All strategies, identified under re-measurement 1, were established as standard procedures during this re-measurement period. In addition provider trainings, annual and at orientation, on clinical guidelines, were standardized during this study period. (see all “Guideline Attachment 6” documents)

Remeasurement 2 to Remeasurement 3

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

Results from the baseline Medical Record Audit and the baseline Clinical Practice Guideline Survey were entered, by the QI Department Data Analyst, into SPSS. Results for each medical record audit were coded into one of three documentation status categories: "Met" (defined as 90% or more of the applicable Guideline checklist items were rated as met), "partially met" (defined as 70% to <90% of the applicable Guideline checklist items were rated as met), and "not met" (defined as less than 70% of the applicable Guideline checklist items were rated as met). Frequencies were run on the proportion of audited medical records which achieved each of the three documentation statuses for baseline measurement results for this indicator.

Frequencies were run on the responses on the Clinical Practice Guideline Survey for the purpose of initial problem description for the project and to establish a baseline measure. Responses on items #6, 7, 8,and 12, which will be used for measuring study indicator #2, were coded into 3 categories: "agree" (if response was "1" or "2"), disagree (if response was "3" or "4") and don't know (if response was "5"). Frequencies were run on the proportion of surveys by the 3 categories to establish the baseline measure for the second indicator.

Remeasurement 1 (updated 2/26/07)

Results from the medical record audit, for the 1st re-measurement, were entered into SPSS by the QI Department Data Analyst. Each audit was categorized as to the adequacy of guideline documentation, using the three groups, "Met," defined as 90% of all applicable guideline items scored as "met" on the checklist, "Partially Met," defined as 70% to <90% of applicable items scored as "met" on the checklist, and "Not Met," defined as less than 70% of guideline checklist items scored as "met" on the checklist. A chi-square test was used to analyze differences in frequencies between medical record audit baseline results and results from the 1st remeasurement. Significance was determined at =<.05 level.

All survey data was entered into SPSS by the QI Data Analyst. Responses for each of the four study survey items were categorized as "agree," if the rating was "1" or "2," "disagree," when item rating was "3" or "4," or "don't know," when the item rating was "5." Chi-square tests were performed for each item, analyzing differences in frequencies between baseline results and the results from re-measurement 1.

Remeasurement 2

Results from the medical record audit, for re-measurement 2, were entered into SPSS by the QI Department Data Analyst. Each audit was categorized as to the adequacy of guideline documentation, using the three groups, "Met," defined as 90% of all applicable guideline items scored as "met" on the checklist, "Partially Met," defined as 70% to <90% of applicable items scored as "Met" on the checklist, and "Not Met," defined as <70% of guideline checklist items scored as "MET" on the checklist. A chi-square test was used to analyze differences in frequencies between medical record audit re-measurement 1and re-measurement 2 as well as baseline results (two-way contingency table analysis). If overall test was significant follow-up tests were conducted to evaluate differences in means among the 3 levels. Significance was determined at =<.05 level.

All survey data results were entered into SPSS by the QI Data Analyst. Responses for each of the four study survey items were categorized as "agree," if the rating was "1" or "2," "disagree," when the item rating was "3" or "4," or "don't know," when the item rating was "5." Chi-square tests were performed for each item, analyzing differences in frequencies between re-measurement 1 and re-measurement 2 as well as baseline results (two-way contingency table analysis). If overall test was significant follow-up tests were conducted to evaluate differences in means among the 3 levels for each item. Significance was determined at =<.05 level.

Remeasurement 3

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.

Baseline Measurement

Results of the data analyses, from the medical record audit for the baseline measure of study indicator #1, indicated that 43.3% of the medical records for consumers with Bipolar disorder and 70% of medical records for consumers with depressive disorders were grouped as "not met" for documentation status. That is, of the 60 medical records audited for the baseline (30 bipolar and 30 deprssion) more than half (56.7%) did not have documentation for 30% or more of the Clinical Guideline Checklist items. There were three items in both checklists that had the most significant problem with adequate documentation: 1). Documenting that a co-occuring medical condtion was assessed and if there was a one that there was consultation or referral to a medical provider; 2). Documenting that education, verbally or through materials, was provided on the mental illness; 3). Both the clinician and the prescriber, documented, if applicable, an assessment of key symptoms of suicide/homicide, substance abuse, and/or psychotic symptoms. Other common documentation problems included: absence from the medical record that specific evidence-based treatment recommended, e.g. CBT for depressive disorders or psycho-ed, CBT, interpersonal social rhythm for Bipolar disorder, that family involvement/enducation recommended for consumers with Bipolar disorder, and that there was evidence that DSM IV criteria met for Depressive Disorder diagnoses. Results of initial baseline analyses reported to the project team.

Results of the data analyses, from the provider Guideline Survey for the baseline measure of study indicator #2, indicated that the majority of MHC providers did believe clinical guidelines were inportant to their work and were useful but most did not find clinical guidelines accessible, user-friendly, and respondents reported little assistance was provided in how to use guidelines. Most of these findings were expected, as there has not been much of an emphais on clinical guidelines in either Network MHC. Because there has not been much emphasis it was a bit surprising that most staff found guidelines important and useful. This was a positive finding that will suppport clinical guideline introduction and documentation of use.

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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.

Remeasurement 1 (updated 2/26/07)

Results of the data analysis, from the medical record audit, for the first remeasurement of study indicator #1, indicated that only 12 (20%) of the 60 records audited, had "not met" status. At the same time 30 (50%) of the 60 medical records reviewed achieved "met" status and 18 (30%) "partially met" status. The differenece between the baseline and the remeasurement 1, regarding the frequency of audits with "not met," "met" or "partially met status," was significant, $x^2 = 21.86$, *df* = 2, p = .000. That is, from baseline to remeasurement 1, there was a significant increase in proportion of audited medical records that achieved "met" status and a concurrent signficant decrease in those that achieved "not met" status. More specifically, improvements were found, from baseline to remeasurement 1, in documentation of lethality risk and substance use, use of evidence-based treatment, and education regarding illness and how to build coping skills. Additional documentation, still in need of improvement include assessment of co-occuring medical condition and assessment of substance abuse as primary or secondary condition. Some of this documentation improvement can be attributed to the PIP strategies, including increased attention and training on clinical guidelines as well as improvement in the design and assessibiity of the guidelines. Other changes, not attributable to the PIP, affecting internal validity of these findings, include such issues as history, such as, at MHCBBC, the implementation of new forms and an electronic medical record that prompt and possibly improved guideline related documentation, and, at both MHCs, an increase in peer review audits, focusing staff attention on the quality of their documentation; selection, such as differneces in staff factors, such as longevity or age, that may affect documention; and instrumentation related to improved auditor skills in located appropriate documentation in the medical record. Along with issues related to internal validity there are also concerns regarding generalizability of these findings. For example, because this study focused on two specific guidelines and these indicator results are specific to these guidelines these findings may not be generalizable to documentation of other guidelines. In addition, since the study focused on documentation of new consumers entering treatment, improved documentation may not be generalizable to documentation of longer term consumers. There were no changes in study procedures or sample size, related to this indicator, to affect the ability to compare the baseline results with the remeasurement results

Results of the data analysis, of the provider Guideline Survey, for remeasurement 1 of study indicator #2, indicated a significant increase, from baseline to remeasurement 1, in number of staff that agreed and a concurrent decrease in number of staff that didn't know, on the responses for item #6, $x^2 = 24.29$, *df* = 2,p = .000; a significant increase in number of staff respondents that agreed and a concurrent decrease in number of those who disagreed or don't know, for survey item #7, $x^2=21.05$, *df* = 2, p=.000; a significant increase in number of staff respondents that agreed and a concurrent decrease in those who disagreed or didn't know, for survey item #8, $x^2=38.91$, *df*=2, p=.000; and a significant increase in number of staff respondents that agreed and a concurrent decrease in those who disagreed or didn't know, for survey item #12, $x^2=22.65$, *df*=2, p=.000. Results suggest a significant improvement in staff perception that clinical guidelines are written in a way that's easily understood, that guidelines are user-friendly, that guidelines are readily accessible, and that someone has explained to staff how guidelines should be used.

These results indicated that PIP strategies implemented had positive effect on staff attitude regarding guidelines, which may have also affected staff attention to documenting guideline use. Other changes affecting staff perception of guidelines, affecting internal validity of these findings, include guideline committee procedures that increasingly involve clinical staff in the development of guidelines (history) and staff differences in the two groups taking the survey (selection), such as differences in longevity, age, or other factors that may affect response on the survey. Issues related to generalizability of these survey findings include the fact that the survey was not administered to FBH providers outside of the MHCs, which , if replicated with this provider group, would provide a more complete picture of staff guideline perception. Issues related to comparing the baseline rsults to the remeasurement, for this indicator, include timing of the survey, in that the survey for the remeasurement was distributed right before a training on guidelines, whereas the survey for the baseline was distributed on a separate day as the guideline training. In addition, there was a smaller sample of staff completing the survey for the remeasurement, compared to baseline.

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.

Remeasurement 2

Results of the data analysis, from the medical record audit, for the second re-measurement of study indicator #1, indicated that only 12 (20%) of the 60 records audited, were categorized as "not met" status. At the same time 29 (48.3%) of the 60 medical records reviewed achieved "met" status and 19 (31.7%) "partially met" status. The difference between audit results for re-neasurement 1 and re-measurement 2 was non significant. That is, the overall two-way contingency table analysis between these two measurement periods indicated no change in frequencies between the three audit levels. On the other hand, the difference between the baseline and re-measurement 2, regarding the frequency of audits with "not met, " "met," or "partially met status," was significant, x^2=21.01, *df*=2, p=.000. That is, from baseline to re-measurement 2, there was a significant increase in proportion of audited medical records that achieved "met" status and a concurrent significant decrease in those that achieved "not met" status. These results were similar to what was found in the analysis between re-measurement 1 and baseline. Follow-up tests, between the paired audit levels, for baseline to re-measurement 2 indicated significant improvement for met vs. partially met, met vs. not met, and partially met vs. not met, although the largest change was between met vs. not met, x^2=21.01 p=.000. More specifically, there were significant improvements, from baseline to re-measurement 2, in documentation of assessment of co-occurring medical conditions, substance abuse, and a larger improvement in documentation, from re-measurement 1, of lethality risk. In addition, improvement was found in documentation of evidence-based practice and family education regarding their illness. Documentation of consumer education about their illness and how to build coping skills was improved from baseline but continues to be documented in the medical record only about a third of the time. As indicated above, under the "Re-measurement 1" section, there appears to be reasonable evidence that the PIP strategies, in particular increased attention and training on clinical guidelines and the importance of documenting their use, positively affected the audit results, supporting the internal validity of the findings. Other system changes, such as implementation of new forms and, at MHCBBC, the electronic medical record, which prompt documentation of key guideline elements, staff factors, such as new staff hired during this two year study, that may be better attuned to the importance of guidelines and their documentation, and, as indicated above, improved auditor skills in locating appropriate documentation, may be a threat to the internal validity of this study. In addition, there are issues regarding the generalizability or external validity of findings, including the assessment of documentation of newly admitted consumers, limiting generalizability to longer term consumers, and the focus on documentation of two disorders, limiting generalizability to other disorder guidelines, such as schizophrenia, which may be more difficult to affect. There were no changes in study procedures or sample size, related to this indicator, to affect the ability to compare the baseline or re-measurement 1 results with re-measurement 2 results.

Results from the provider Guideline Survey or study indicator #2, item #6, indicated a significant increase in frequency, from re-measurement 1 to re-measurement 2, x^2=11.41 p=.003, specifically an increase in number of staff that agreed and a concurrent decrease in number of staff that don't know and a significant increase in staff that disagree and a concurrent decrease in staff that don't know. For item #7 there was a significant increase in frequency between re-measurement 1 to re-measurement 2, x^2=15.43, p=.000, specifically an increase in staff that agree and a concurrent decrease in staff that don't know and an increase in staff that disagree and a decrease in staff that don't know. For item #8 there was a significant increase in frequency between re-measurement 1 to re-measurement 2, x^2=9.35, p=.009, specifically an increase in staff that agree vs. don't know and staff that disagree vs. don't know. Last for item #12 there was a significant increase in frequency between re-measurement 1 and re-measurement 2, x^2=12.22, p=.002, specifically an increase in staff that agree vs. don't know and staff that disagree vs. don't know.

Although there continued to be significant improvement from baseline to re-measurement 2 for all four survey items, with number of staff who agreed with the four times there was not a decrease in staff that disagreed. Rather, there was a decrease in number of staff who "didn't know." That is, a significantly larger percent of staff reported, in survey responses for re-measurement 2, compared to baseline, that clinical guidelines were written in a way that is easily understood, that clinical guidelines were user-friendly, that guidelines were readily accessible, and that they were told how they should use the clinical guidelines but these increases were

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.

reflective of fewer staff indicating "don't know" rather than fewer staff indicated they "disagreed." These results provide reasonable evidence that PIP strategies had a significant positive effect on staff attitudes regarding clinical guidelines, from baseline through re-measurement 2, at least in regard to an increase in agreement on guidelines usefulness, etc, and an increase in staff knowledge regarding clinical guidelines. At the same time there are extraneous variables that may have affected internal validity of these results, including staff respondent differences in the two groups (selection issues), e.g. differences in longevity, age, or other staff characteristics, that may affect response on the survey, as well as increased involvement of clinical staff in guideline development. Threats to external validity continue to be the fact that this survey was only administered to MHC staff, limiting generalizability to other FBH providers. In addition, as mentioned under "remeasurement 1" there were some differences in survey administration timing and a larger number of staff surveyed for re-measurement 2, possibly affecting the comparability of results from baseline to re-measurement 2.

Remeasurement 3

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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

#1 Quantifiable Measure: The proportion of audited medical records that achieve "met" status (90% or more of applicable items on the Audit Checklist tool were rated as "met"), "partially met" status (70% to <90% of applicable items on the Audit Checklist tool are rated as "met"), or "Not Met" status (<70% of the applicable items on the Audit Checklist tool were rated as "met").

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
January-October, 2005	***Baseline: Study Indicator #1***	**9 (met status) 17 (partially met) 34 (not met status)**	60	**15% (met) 28.3% (partially met) 56.7% (not met)**	none	**x^2=21.86 *df*=2, p=.000 (re-measurement 1)**
January-October, 2006 (updated in bold 2/26/07)	**Remeasurement 1:**	**30 (met status) 18 (partially met status) 12 (not met status)**	**60**	**50% (met) 30% (partially met) 20% (not met)**	**none**	
January-October, 2007	**Remeasurement 2:**	29 (met status) 19 (partially met status) 12 (not met status)	60	48.3% (met) 31.7% (partially met) 20% (not met)	none	**re-measurement 1 to 2: overall 2x3 x^2=.044 *df*=2, p=.978 baseline to re-measurement 2: overall 2x3 x^2=21.16 *df*=2, p=.000 follow-up tests: met vs. partially met: x^2=4.49 *df*=1, p=.034 met vs. not met: : x^2=21.01 *df*=1, p=.000 partially met vs. not met: x^2=6.12 *df*=1, p=.013**
	Remeasurement 3:					
	Remeasurement 4:					
	Remeasurement 5:					

Appendix A: PIP Summary Form:
Improving Use and Documentation of Clinical Guidelines
for Foothills Behavioral Health, LLC

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

#2 Quantifiable Measure: The percent of Network MHC provider respondents on the Clinical Practice Guideline Survey that agree (response of 1 or 2), **disagree (response of 3 or 4), or don't know (response of 5)** that clinical guidelines are easily understood (Item #6), user-friendly (item #7), readily accessible (item #8), and understand how to use clinical guidelines (item #12)

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
Oct through Nov, 2005	***Baseline: Study Indicator #2***	**Item 6:** **101 (agree)** **22 (disagree)** **75 (don't know)** **Item 7:** **94 (agree)** **30 (disagree)** **74 (don't know)** **Item 8:** **79 (agree)** **55 (disagree)** **64 (don't know)** **Item 12:** **54 (agree)** **112 (disagree)** **32 (don't know)**	**Item 6:** 198 **Item 7:** 198 **Item 8:** 198 **Item 12:** 198	**Item 6:** **51% (agree)** **11.1% (disagree)** **37.9% (don't know)** **Item 7:** **47.4% (agree)** **15.2% (disagree)** **37.4% (don't know)** **Item 8:** **39.9% (agree)** **27.8% (disagree)** **32.3% (don't know)** **Item 12:** **26.3% (agree)** **56.5% (disagree)** **16.2% (don't know)**	none	

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

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

Time Period Measurement Covers	Baseline Project Indicator Measurement	Numerator	Denominator	Rate or Results	Industry Benchmark	Statistical Test and Significance*
May - June, 2006 (updated in bold 2/26/07)	**Remeasurement 1:**	**Item 6:** **129(agree)** **8 (disagree)** **33 (don't know)** **Item 7:** **120 (agree)** **12 (disagree)** **37 (don't know)** **Item 8:** **122(agree)** **25 (disagree)** **22 (don't know)** **Item 12:** **87 (agree)** **65 (disagree)** **17 (don't know)**	**Item 6: 170** **Item 7: 169** **Item 8: 169** **Item 12: 169**	**Item 6:** **75.9% (agree)** **4.7% (disagree)** **19.4% (don't know)** **Item 7:** **71% (agree)** **7.1% (disagree)** **21.9% (don't know)** **Item 8:** **72.2% (agree)** **14.8% (disagree)** **13% (don't know)** **Item 12:** **51.5% (agree)** **38.5% (disagree)** **10% (don't know)**	none	**Item 6: x^2=24.29, df=2, p=.000** **Item 7: x^2=21.05, df=2, p=.000** **Item 8: x^2=38.91, df=2, p=.000** **Item 12: x^2=22.65, df=2, p=.000** **Item 6 re-measurement 1 to 2: x^2=11.41, *df*=2, p=.003** **Follow-up tests:** **Agree vs. disagree x^2=1.46, *df*=1, p=.227** **Agree vs. don't know x^2=9.03, *df*=1, p=.003** **Disagree vs. don't know x^2=8.32, *df*=1, p=.004** **Item 6 baseline to re-measurement 2: x^2=57.87, *df*=2, p=.000**

Appendix A: PIP Summary Form: Improving Use and Documentation of Clinical Guidelines *for* Foothills Behavioral Health, LLC

I. Activity IX: Report improvement. Describe any meaningful change in performance observed and demonstrated during baseline measurement.						
May-June, 2007	**Remeasurement 2:**	Item 6: 192 (agree) 20 (disagree) 20 (don't know) Item 7: 184 (agree) 28 (disagree) 20 (don't know) Item 8: 177 (agree) 44 (disagree) 11 (don't know) Item 12: 112 (agree) 114 (disagree) 6 (don't know)	Item 6: 232 Item 7: 232 Item 8: 232 Item 12: 232	Item 6: 82.8% (agree) 8.6% (disagree) 8.6% (don't know) Item 7: 79.3% (agree) 12.1% (disagree) 8.6% (don't know) Item 8: 76.3% (agree) 19% (disagree) 4.7% (don't know) Item 12: 48.3% (agree) 49.1% (disagree) 2.6% (don't know)	None	**Item 7 re-measurement 1 to 2: x^2=15.43, *df*=2, p=.000** **Follow-up tests:** **Agree vs. disagree x^2=1.34, *df*=1, p=.247** **Agree vs. don't know x^2=12.64, *df*=1, p=.000** **Disagree vs. don't know x^2=11.46, *df*=1, p=.001** **Item 7 baseline to re-measurement 2 : x^2=57.90, *df*=2, p=.000** **Item 8 re-measurement 1 to 2: x^2=9.35, *df*=2, p=.009** **Follow-up tests:** **Agree vs. disagree x^2=.488, *df*=1, p=.485** **Agree vs. don't know x^2=8.10, *df*=1, p=.004**
	Remeasurement 3:					
	Remeasurement 4:					

I. Activity IX: Report improvement. Describe any meaningful change in performance observed and demonstrated during baseline measurement.						
	Remeasurement 5:					**Disagree vs. don't know x^2=8.32, *df*=1, p=.004** **Item 8 baseline to re-measurement 2: x^2=73.97, *df*=2, p=.000** **Item 12 re-measurement 1 to 2: x^2=12.22, *df*=2, p=.002** **Follow-up tests:** **Agree vs. disagree x^2=2.15, *df*=1, p=.143** **Agree vs. don't know x^2=7.55, *df*=1, p=.006** **Disagree vs don't know x^2=11.95, *df*=1, p=.001** **Item 12: x^2=35.61, *df*=2, p=.000**

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

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

There was a sustained improvement in indicator one from baseline to re-measurement one and baseline to re-measurement two (repeated measurement), with both re-measurements showing a statistically significant increase, from baseline, at the .05 level, in the percent of audited charts that achieved "met" status and a concurrent significant reduction in percent of audited charts that achieved "not met" status. In addition, although there was a slight decrease in percent of audited charts achieving "met status" from re-measurement one to re-measurement two, this difference was non significant. For indicator #2, there was a sustained improvement, from baseline to re-measurement one and baseline to re-measurement two (repeated measurment), with both re-measurements showing a statistically significant increase, from baseline, at the .05 level, in the percent of survey responses "agreeing" with item #6, #7, #8, and #12. At the same time, there wasn't a sustained concurrent significant decrease in percent of survey responses "disagreeing" with each item, with re-measurement 2 results indicating the same proportion of staff disagreeing with each of the four items. In addition, although there was an increase, for all four items, in percent "agree," from re-measurement one to re-measurement two, as well as a decrease in percent staff that "don't know," the percent of staff "disagreeing" with all four items, at re-measurement 2, actually increased from re-measurement 1. Change in indicator #2 survey response patterns, in particular for Item #6 and #12, from re-measurement one to re-measurement two, seemed to be a movement from "don't know" to a "disagree" versus "agree status." This change may have been due to changes in the staff sample, with staff becoming more aware, over the study period, of guidelines and more critical of how understandable the guidelines are or how to really use them in practice.

There was no difference in the sampling error between re-measurement one and re-measurement two for indicator one. For indicator two, item #6, there was a slight decrease in the sampling error from re-measurement one to re-measurement two, with a 95% confidence that the percent agree was +-.049 (re-measurement two) and +-.064 (re-measurement one). This difference in sampling error, between re-measurement one and two, was similar and smaller for items 7, 8, 12, with only about a 1% difference. with 95% confidence that the percent agree was between +- .052 and +-.074. The decrease in sampling error was likely due to an increase in sample size for this indicator in re-measurement 2.

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 Foothills Behavioral Health, LLC

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.