Outcomes Management

ALERT[SM] ANNUAL REPORT JANUARY 2001

History and Overview

ALERT (**AL**gorithms for **E**ffective **R**eporting and **T**reatment) is an outcomes management system, first developed by PacifiCare Behavioral Health (PBH) in 1998, now successfully implemented within employer-based and public sector delivery systems. ALERT is unique because it:

- analyzes change based on member self-report and assessment in combination with practitioner-reported data.
- monitors data *during* a treatment episode rather than merely measuring outcomes before and after treatment.
- makes real-time information available for clinicians to help them develop better-informed treatment planning decisions.
- compares actual change from session-to-session against *expected change*.

At the core of the ALERT program are two outcome assessment tools developed exclusively for PBHI by Michael Lambert, Ph.D. and Gary Burlingame, Ph.D. of Brigham Young University. They conducted an item analysis of the original 45 questions of the OQ-45 and 65 questions of the YOQ, and created two new 30-item instruments, the Life Status Questionnaire (LSQ) and Youth Life Status Questionnaire (YLSQ). Drs. Lambert and Burlingame also continue to consult with PBH on report development and data analyses to continually refine the program.

The primary consultant working with PBH since the inception of ALERT is Jeb Brown, Ph.D., Director of the Center for Clinical Informatics. He has taken the lead in developing the clinical algorithms and reports that are at the core of the system, and he conducted the original data analysis for this annual report. Dr. Brown works closely with Drs. Lambert and Burlingame, and they have consulted

in the development of the procedures and statistical methods utilized in this report.

The ALERT program was developed in the second half of 1998 and first implemented in February 1999 with three PBH Preferred Group Practices (PGPs) in California. By April 1999, the program was implemented with all PGPs in California and New Mexico. In July 1999, the YLSQ was implemented with all PGPs, and in addition, the program was implemented with approximately 100 of the highest volume individual practitioners. In March 2000 the program was implemented with all practitioners in California. Finally, the program was implemented in Q4 2000 with members in all PBH markets being managed in the Extended Care Management (ECM) and Assertive Care Management (ACM) models.

The ALERT program is also operational in Oregon with the five-county ABHA Medicaid population. The LSQ and YLSQ have been in place since November 1999 with the five ABHA county clinics and with the network of contracted practitioners used by three of the counties. In general, the ABHA population of 23,000 members enter treatment with more severe behavioral health problems, and a separate analysis of their outcomes will be completed and distributed.

ALERT Outcome Measures

The LSQ and YLSQ are investigational instruments based on a revised set of 30 items from the OQ-45 and Youth Outcome Questionnaire (YOQ). The LSQ and YLSQ were developed based on the assumption that the items from parent instruments would exhibit similar psychometric properties as part of a 30-item measure. Items were selected for inclusion in the new measures based on three criteria:

1) High face validity for symptoms and concerns common to patients in mental health treatment

2) Ability to differentiate treatment from non-treatment samples

3) Sensitivity to change in treatment samples while remaining relatively stable in non-treatment samples.

Drs. Lambert and Burlingame have accumulated a large data repository of the OQ-45 and YOQ items through collaboration with several large managed care companies. The size and scope of this repository permitted complex modeling of how the new instruments could be expected to perform in real world clinical settings. A sample, consisting of over 3200 adults and 600 children/adolescents, was utilized to model the properties of the new tests. The sample came from a

population of commercially insured outpatients in managed care settings. This sample serves as a benchmark against which to compare PBH's outcomes.

A critical element of the ALERT system is the ability to predict the expected change for individual patients using the outcome measures. These trajectory of change formulas for the ALERT system were derived from the data repository of OQ-45 and YOQ questionnaires, again with the assumption that these calculations would remain constant when the items were presented in a 30-item scale. The formulas provide a target outcome for each case against which to compare actual results. This provides the basis for case mix adjustment and benchmarking LSQ results against the norms from the data repository.

The assumption that the 30-item instruments would perform as predicted proved accurate. An expected change score was established at intake for each patient in the ALERT system, using the formulas derived from the data repository. At the end of treatment, it is possible to compare the difference between the predicted change and the actual change for individual patients or large groups of patients. In the case of the LSQ, the average difference between the predicted change and actual change in the sample for this report is .33 points (.018 effect size). This represents less than 5% of the total change.

The YLSQ outcomes have tended to show more improvement than the formulas predict, reducing the overall accuracy. However, the YLSQ sample (n=479) is less than 20% the size of the LSQ sample, so some of the variance may be due to small sample size.

The YLSQ permits completion by either an adult or an adolescent. The protocols completed by adults had slightly higher scores at intake than adolescent-completed tests (43.7 versus 39.5) and also showed slightly more change after controlling for intake severity. However, neither difference approached statistical significance ($p>.2$).

In addition to these full-length measures, three items from the LSQ are used as part of the telephonic intake and triage procedure. These three items are used to estimate improvement between the first phone contact and subsequent treatment sessions. The results using this three-item measure are discussed in a subsequent section.

Treatment Outcomes

Treatment outcomes are reported as "effect size," a type of standardized change score calculated by dividing change during treatment (raw score difference in scores from start to end of treatment) by the standard deviation (variability of change scores) of the outcome measure. An effect size of zero means no change on average resulted from psychotherapy. The average effect size for

psychotherapy studies is estimated to be .82, while minimal treatment controls effect size averages .42.[1] The overall combined effect size for the PBH treatment sample is .36. While on the surface this appears to be a moderate effect size at best, this conclusion is misleading. The probability of improvement with treatment is proportional to the severity of distress at intake. Patients with more symptoms are much more likely to show improvement than those with minimal distress.

The PBH sample contains a large percentage of cases with minimal levels of distress as measured by the outcomes questionnaires. Since psychotherapy studies often select for individuals that meet certain diagnostic criteria, the subjects are more likely to have high levels of symptoms and thus show more improvement. If the PBH sample is reanalyzed after excluding cases with intake scores below 44, the effect size is .8, comparable to that found in psychotherapy studies.

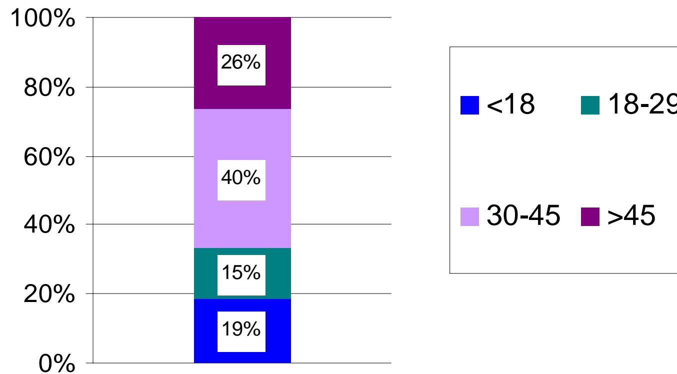
The ALERT system avoids the problem of the dependency of effect size on the severity of the sample by using the Outcome Index statistic. The Outcome Index is case mix adjusted in that it represents the deviation from a target outcome, and thus is independent of the magnitude of the effect size per se.

Treatment Sample

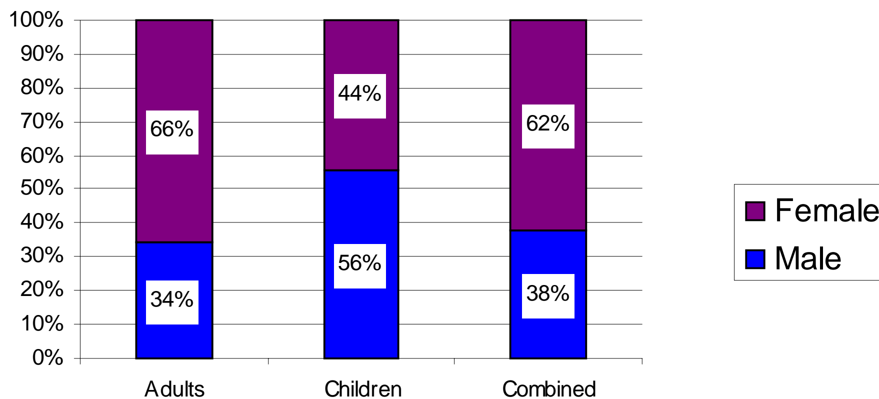
By the end of 2000 PBH had analyzed over 38,000 patient self-report questionnaires. The ALERT system involves multiple measures during treatment, and so this total count represents participation in ALERT by approximately 14,000 commercially insured members. The sample of cases analyzed in this report began outpatient treatment between May 1 of 1999 and April 31 of 2000. Most have completed treatment, some are continuing in treatment, and all have had at least three months of treatment. The primary sample used for this report consists of 7390 cases, of which 3015 (41%) have at least two measurement points in treatment.

The following bar graph presents a breakdown of the sample by age.

[1] Lambert, M.J. Weber, F.D. & Sykes, J.D. (1993, April) *Psychotherapy versus placebo.* Poster presented at the annual meeting of the Western Psychological Association, Phoenix.



The total sample consists of a majority of females (62%). However, in the under 18 age group, males outnumbered females (55%). The following bar graphs display the break down of sex and age (adults and children).



The Results

This report summarizes results for the ALERT outcomes management system and examines the validity of the assumptions underlying the design of the system. The Outcome Index results presented are for the PBH commercially insured population, and breaks out the results by age group and severity levels. Note that one of the severity ranges is designated “Normal.”

PacifriCare Behavioral Health Care Aggregated Outcomes Report: Commerical population

Cases included in this report began between May 1, 1999 and April 31, 2000.

Age Group		**> 1 data point**		**Change**		***Change Index***
Severity at intake	**Total**	Number	Sessions/	(effect size)		*(actual-expected)*
Adults	**Cases**	cases	Case	actual	expected	
Normal range	1554	546	6.39	-0.09	-0.13	0.04
Mildly distressed	1520	592	6.47	0.24	0.20	0.05
Moderately distressed	1541	706	6.74	0.47	0.44	0.03
Severely distressed	1508	692	6.45	0.75	0.79	-0.04
Combined Adult	**6123**	**2536**	**6.52**	**0.37**	**0.36**	**0.02**
Children & Adolescents						
Normal range	423	157	6.17	-0.05	-0.07	0.03
Mildly distressed	385	158	6.63	0.24	0.15	0.09
Moderately distressed	275	95	7.04	0.45	0.38	0.06
Severely distressed	184	69	6.65	0.85	0.62	0.24
Combined Child/Adolescent	**1267**	**479**	**6.56**	**0.27**	**0.19**	**0.09**

Aggregate Results for All Age Groups

		Change		***Change Index***
Total number of cases:	**7390**			
Number of cases with > one data point:	**3015**			
% of cases with > one data point:	**41%**	*actual*	*expected*	*(actual-expected)*
Sessions Per Case:	***6.53***	**0.36**	**0.33**	**0.03**

The target outcome (average expected effect size using the case mix formulas) is .33. Since the actual effect size is .36, the Outcome Index is .03 (actual change – expected change). This result is above the target at the 95% confidence level. This indicates that PBH results are equivalent if not better than the benchmark

The results for children exceed the benchmark to a greater degree than the results for adults (Change Index of .09 as opposed to .02). The children's outcomes are above the benchmark at the 95% confidence level, and this pulls up the overall results for all members. The adult results do not exceed the benchmark at the 90% confidence level, and so they can be interpreted as being essentially consistent with expected change.

The LSQ and YLSQ displayed the expected psychometric properties. The case mix adjustment formulas used to establish the benchmark for outcomes have been found to be valid for the PBH commercially insured treatment population. Encouragingly, all indications are that these norms are also appropriate for PBH's public sector population.

Overall results for PBH's treatment population are at least on par, and possibly superior to, the national benchmark established by the data repository for the outcome instruments. None of the PGPs have outcomes below the target (at a 95% confidence level), while four of the groups have results that are above the target (95% confidence). Members treated by one of these groups average 33% more improvement than the target.

Measuring Change from Phone Intake to Treatment Follow-Up

The **ALERT** system assesses patient change from the point of initial phone contact to a follow-up assessment six months after treatment. There are of course many challenges to the systematic collection of so many repeated measures in a naturalistic setting across a long period of time. Missing data at the various assessment points compound to limit the number of cases with complete data across all measurement points. Despite these limitations, these data are unique in permitting a glimpse at patient change across an entire episode of care.

The initial measure of severity is captured during the telephone intake by the use of three items from the LSQ. The member is asked to respond to the items using a scale from 0 (or "Never") to 4 (or "Almost Always"). The three items were selected due to their ability to differentiate a treatment from a non-treatment sample. A fourth item is used as a self-report screening item for substance abuse problems.

- *I am satisfied with my life*
- *I feel hopeless about the future*
- *I feel sad. (Note: wording changed from "blue" to "sad" after feedback from customer service associates.)*

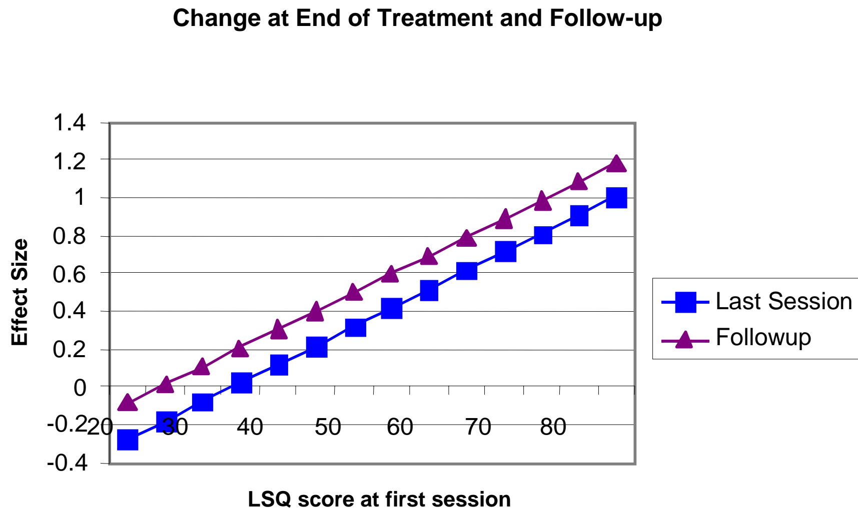
In order to permit ready comparison of the three-item scale to the full-scale LSQ scores, a regression analysis was conducted to provide the formula for converting the 3-item scale score to its 30-item equivalent. A sample of 442 cases was identified with data at triage and the first session. The mean score at triage for this sample (using the 30-item equivalent score) was 54.5, and the standard deviation was 18. The mean score for the first session was 53.2 for this sample. The mean score at the first session for all members in the PBH data repository is 52.5, and this non-significant difference supports the premise that this triage sample is comparable in severity to the entire PBH patient population. The total average change from the point of phone triage to the first session is a nominal 1.3 points.

The follow-up measure was collected by systematically mailing out LSQs to all adult members in the ALERT system after no additional data had been received for six months. These follow-up mailings did not begin until Q1 2000. Of the first 2700 questionnaires mailed out, approximately 10% were returned as undeliverable due to an address change. A total of 393 were completed and returned, for a return rate of 16%, which is comparable to return rates for large HMOs using a similar method.

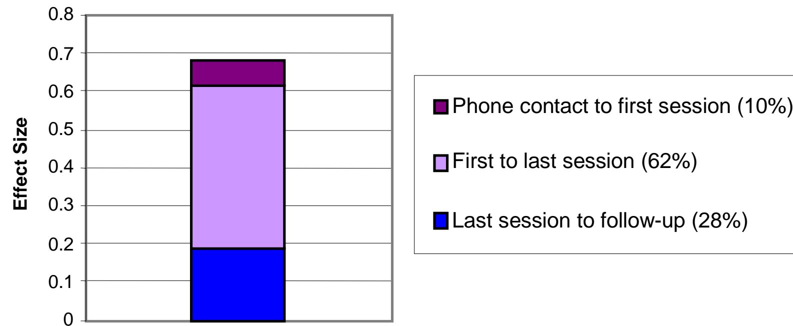
This follow-up sample provides an estimate of the rate of change during the post-treatment period. A total of 286 follow-up LSQ protocols were matched with complete LSQ scores at the first session. Of these 286 cases, 139 also contained a second data point during the treatment episode.

In virtually all respects this sample appears comparable to the PBH data repository. The mean LSQ score at intake was 53.6, with a standard deviation of 18. Forty eight percent (48%) contained two data points, compared to the overall PBH average of 42% with at least two data points. Those cases with two data points displayed virtually identical patterns of change to the PBH data repository. The patterns are so close that the regression formulas to predict change, when rounded to two decimal places, are identical. The status at follow-up for patients with a single measurement at the first session was comparable to those with multiple measurement points in treatment.

The following graph shows the regression lines for change at the end of treatment and at six-month follow-up. Across the range of severity, members tended to show continued improvement after treatment. The amount of change is 3.4 points on average, or an effect size of .19.



Taken together, these samples provide a unique opportunity to look at patterns of change across the entire treatment episode, from initial phone contact, to the end of treatment, to the post-treatment consolidation of treatment gains. The following graph displays the percentage of overall change that occurs during each phase of the treatment process.



Cases at Risk for Poor Outcomes

One of the key features of the ALERT system is the ability to track at-risk cases based upon trajectory of change projections. Once the LSQ/YLSQ has been administered at least twice, the system calculates the most likely outcome using trajectory of change projections. These are based on formulas derived from cases in the data repository. The system is designed to target approximately 10% of cases at highest risk for a poor outcome. In order to test the validity of this logic, a sample of 2,343 cases (adults and children combined) with assessments at the first and third sessions was used to identify at-risk cases and track the eventual outcome of these cases.

This analysis used calculations identical to those employed every day by the ALERT system in generating the daily High Risk Report. The result was that 204 (9%) of the cases were targeted at the *third* session as high risk for a poor outcome. This high risk group has scores averaging well into the severe range at the third session, and they have worsened an average of 16 points (-.85 effect size) since the first session.

The most probable outcome for these cases is simple—members are likely to end treatment based on the fact that they are feeling worse after three visits. Only 31% of these cases have a subsequent assessment at the fifth or later session, compared to 42% of members not identified as being at-risk. Those at-risk cases that fortunately do continue in treatment to at least the fifth session show significant improvement beyond the third session, averaging 13 points of improvement (.7 effect size) between the third session and last assessment point.

This analysis indicates that the logic for targeting at-risk cases is valid. The percentage of cases targeted in this analysis is similar to the percentage targeted by the algorithms built into the daily High Risk Report (9% versus 10%). These cases appear to leave treatment prematurely at a higher rate than the comparison group. Those members who continue in treatment show significant improvement

from that point forward. (Nonetheless, it should be noted the gains do not fully offset the deterioration seen in the first 3 sessions.) These findings suggest that quality improvement efforts should focus on identifying methods to keep these at-risk patients engaged in treatment for a sufficient duration to experience benefits.

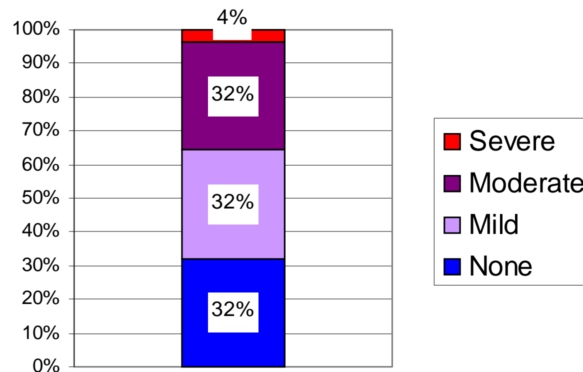
Identification of Patients at Risk for Suicide and/or Substance Abuse

Another function of the ALERT system is to target cases with recurrent suicidal ideation and/or high potential for substance abuse in order to ensure that services of adequate intensity are provided. The system incorporates information on these risk factors from the Provider Assessment Report (PAR) and from the LSQ/YLSQ.

In order to check the congruence between the provider assessment and the patient self-report, an analysis was conducted on the entire sample. The report design matched PARs with LSQs administered within 7 days of one another. This produced 1,391 cases with concurrent PAR and LSQ data.

Of these cases, 56 (4%) reported suicidal ideation “Frequently” or “Almost Always” on the LSQ item “I have thoughts of ending my life.” The average LSQ score for patients reporting this level of suicidal ideation is 77 points, which is at the high end of the severe range. In fact, a score of 77 or higher occurs in approximately 5% of cases. These same cases also have an average response of 3 (“Frequently”) on the item, “I feel hopeless about the future.” Over 70% of these members with recurrent suicidal thoughts have scores of 3 or higher on the hopelessness item.

These are clearly very distressed individuals with high levels of depression and anxiety, frequent suicidal ideation, and a high level of hopelessness. Surprisingly, almost two thirds of these cases were assessed by the practitioner as presenting either with no suicidal thoughts or only a mild level of ideation. The following graph shows the breakdown of provider assessment of suicidal ideation for these cases.



It is not possible to offer a definitive explanation for the discrepancy in assessments, but numerous conversations with practitioners suggest there may be two very different sources. First of all, there are instances when practitioners seem genuinely surprised by the discrepancy when contacted by phone by Care Managers, and it would seem that these practitioners did not detect the presence of suicidal thinking. However, the other source of discrepancy is a thorough assessment by the practitioner who rates the patient's suicidal thinking as passive and "mild" in terms of risk. In other words, the practitioner may be well aware of the patient's frequent suicidal thoughts, but the rating of "mild" ideation on the PAR reflects a more comprehensive analysis of the risk factors associated with those thoughts.

A similar analysis was performed to check the congruence between the provider assessment and the patient self-report regarding substance abuse. This analysis compares responses on the LSQ chemical dependency (CD) items and the clinician assessment of substance abuse problems on the PAR. The LSQ has three CD-related items. These are scored on a five-point scale, from 0 for "Never" to 4 for "Almost Always." The CD scale score is derived by adding the three items.

1) I use alcohol or a drug to get going in the morning.

2) People criticize my drinking (or drug use).

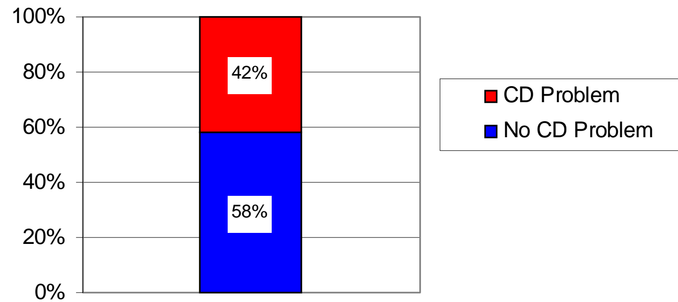
3) I have trouble at work/school or other daily activities because of drinking or drug use.

Arguably, any response other than a "Never" on all three items could be indicative of a problem. Twenty percent (20%) of the sample responded with at least one "Rarely." However, it is difficult to have confidence that a single "rarely" is strong evidence of substance abuse problem.

For purposes of this analysis, a more stringent requirement of a score of four or higher was used to indicate presence of a substance abuse problem. Less than five percent (5%) of members scored this on this scale. It appears that virtually all patients self-reporting this degree of difficulty have a substance abuse problem.

Similar to the case with the suicide item, a significant percentage of these self-admitted substance abusers are not identified by the provider as having a problem. The following graph displays the results. Fifty eight percent (58%) of the cases targeted by the LSQ were not identified by providers as having a problem.

PAR Assessment of CD Problem for Patients scoring in the upper 5% on LSQ CD scale

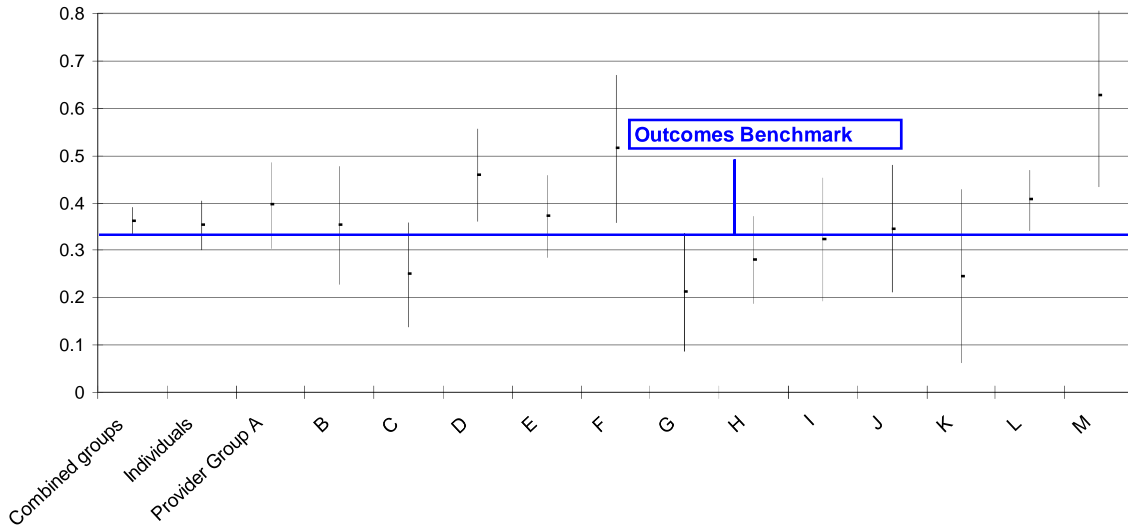


Preferred Group Practice Results

The ALERT system provides monthly reports comparing results for PGPs, and a graph is generated that depicts the results for all groups with at least 50 cases with two data points. The following graph presents the results for 13 groups that meet this criterion of at least 50 cases. The chart also plots the combined results for all groups and for all solo practitioners.

The Outcomes Benchmark is the overall expected effect size for the entire sample (.33 effect size). The individual group results are plotted by adding their individual Change Index to this benchmark value. While the actual effect size for each group will vary slightly from the results plotted here, this method provides a fair means of comparison while conveying information about the general magnitude of the effect size.

The results are plotted with a 90% confidence interval, meaning that there is a 5% chance that the results are above the interval and 5% below the interval. Based on this criterion, four of the 13 groups exceed the benchmark at the 95% confidence level. None of the groups are significantly below the benchmark.



These four groups are responsible for elevating the overall results. If these groups are removed from the calculation of aggregated effect size, the Outcome Index score for the remaining groups and individual providers is .00, meaning that the PBH outcomes are virtually identical to those in the data repository.

A patient seen in one of these four groups averages over 33% more improvement per episode of care than patients treated elsewhere in the network. The reasons for the superior results are not apparent. These results present an opportunity to determine how care is delivered in these groups and then promote “best practices” within other groups.

A Practitioners's Guide to the ALERT System

The Required Data Set

Patient Self-Report Tools

WHAT: PBH utilizes two patient self-report instruments for monitoring clinical change: the Life Status Questionnaire **(LSQ)** for adults (18 and over) and the Youth Life Status Questionnaire **(YLSQ)** for children and adolescents. The YLSQ can be completed by the parent (generally for children under 12) or by the adolescent. Higher scores reflect higher levels of distress, and so treatment generally lowers scores on these tests.

WHEN: Administer for all outpatient episodes prior to the first, third, and fifth sessions, and then in increments of five, at sessions 10, 15, 20, and so on. (Note: the questionnaires can be administered prior to every session if the clinician desires.)

WHY: The Y/LSQ are assessment tools sensitive to changing levels of symptomatic distress for patients in behavioral health treatment. Frequent administration in the first five sessions is intended to detect patients with a clinical profile suggesting risk for suicide or chemical dependency. In addition, the Y/LSQ can detect patients at high risk for prematurely terminating treatment due to escalating levels of distress (generally 10% of a treatment population). In the case of longer episodes of care for more complex conditions, administration every 5th session is necessary to monitor clinical risk. PBH seeks to ensure throughout each episode of care that patients receive the appropriate intensity and types of clinical services.

HOW: It is best to use a black pen or a soft pencil when completing the form, and the circle should be darkened completely rather than checked. It is helpful to educate patients on the value of completing the Y/LSQ. An important element of that education would include a quick review of the patient's responses in the session, always including a review of the critical items related to risk. What about the resistant patient? We would never suggest that you force someone to complete a form when they are firmly opposed, but we usually find that there are two reasons people resist completing them. They often have mistaken ideas about how the form is used, and secondly, they refuse when they perceive that the clinician does not value the process. If you convey that these brief assessment tools can help you track progress together, most people are willing to complete the forms.

WHO: When a patient has multiple providers for medication monitoring and psychotherapy, each provider may utilize the Y/LSQ. However, we request that the Y/LSQ always be administered by the psychotherapy provider.

Provider Assessment Report (PAR) & Telephonic Review

The **PAR** is a one-page assessment form completed by the clinician following the initial treatment sessions authorized at intake, designed to provide PBH with data on clinical severity and risk from the practitioner's perspective. When a treatment episode requires more visits than authorized upon PAR review, psychotherapists should initiate telephonic review with a PBH Outpatient Care Manager. Prescribing practitioners have two choices: 1) submit a new PAR for ongoing medication management, or 2) call PBH for a brief telephonic review.

Y/LSQ and PAR data are preferably **faxed to (800) 992-2809 with no fax cover sheet**, or if necessary, mailed to the PBH regional office.

Please Note: *The most critical information on these forms for administrative purposes is the* ***Reference ID – Suffix. <u>Please be sure this number is correct.</u>*** *Your patient is given this number at the time of the initial call to PBH, and it is identified prominently on the authorization form sent to your office.*