

## Identifying Highly Effective Psychotherapists in a Managed Care Environment

G. S. (Jeb) Brown, PhD; Michael J. Lambert, PhD; Edward R. Jones, PhD;  
and Takuya Minami, PhD

**Objective:** To investigate the variability and stability of psychotherapists' effectiveness and the implications of this differential effectiveness for quality improvement in a managed care environment.

**Study Design:** Subset archival outcome data for patients receiving behavioral health treatment were divided into 2 time periods to cross-validate the treating therapists' effectiveness. After categorizing the therapists as "highly effective" and "others" during the baseline period, the stability of their individual effectiveness was cross-validated in the remaining time period.

**Methods:** Outcomes for 10 812 patients (76.0% adults, 24.0% children and adolescents) treated by 281 therapists were included. Patients initiated treatment between January 1999 and June 2004. Mean residual change scores obtained by multiple regression were used to adjust for differences in case mix among therapists. Raw change scores as well as mean residualized change scores were compared between the 71 psychotherapists identified as highly effective (25%) and those identified as other (remaining 75%).

**Results:** During the cross-validation period, mean differences in residualized change score between highly effective therapists and others were statistically significant (difference = 2.8;  $P < .001$ ), which corresponded to an average of 53.3% more change in raw change scores with the highly effective therapists. Results could not be explained by case mix differences in diagnosis, age, sex, intake scores, prior outpatient treatment history, length of treatment, or therapist training/experience.

**Conclusion:** Behavioral health outcomes for a large system of care could be significantly improved by measuring clinical outcomes and referring patients to therapists with superior outcomes.

(*Am J Manag Care.* 2005;11:513-520)

Managed behavioral health care organizations (MBHOs) are responsible to their customers and enrollees to ensure that psychotherapy services meet accepted standards of quality and effectiveness. To this end, MBHOs engage in a variety of credentialing and quality-improvement activities. With respect to credentials, typical MBHOs (1) verify that participating therapists have valid licenses to provide behavioral health services, (2) specify a minimum number of years of experience, and (3) require that there be no evidence of malpractice. To ensure quality, most MBHOs encourage the use of empirically supported treatments established in psychotherapy clinical trials.<sup>1,2</sup> Under these standards, MBHO care managers

review treatment plans submitted by qualified therapists and approve only those deemed appropriate. Thus, quality assurance in most MBHOs is to determine the following: the ability of the licensed therapist to propose a treatment plan to a care manager that involves an empirically supported treatment for a specified disorder.

However, advances in research methodology have enabled researchers to more critically examine the accumulating clinical evidence upon which these quality assurance standards are based. One advancement is that of meta-analysis, which is a statistical method for combining results from a large number of studies and thereby permitting the investigator to draw conclusions that could not be drawn from individual studies.<sup>3</sup> A second advancement is that of hierarchical linear modeling, which is an extension of multiple regression and allows for analysis of data that are hierarchical.<sup>4</sup>

Recent reviews of clinical trials based on these advancements have indicated that the commonly utilized quality assurance standards may no longer be adequate. The past 2 decades of meta-analytic studies have revealed that differences in the relative efficacy of various psychotherapies are minimal at most.<sup>5-9</sup> Therefore, treatment plans based on empirically supported treatments for specified disorders do not ensure that the proposed treatment will be more effective than other treatments.

One could certainly argue that although treatment plans based on empirically supported treatments do not ensure the most effective treatment, this approach at least ensures that the treatment being delivered is at an acceptable standard. Despite its logical appeal, this assertion holds true only if (1) therapists are sufficient-

From the Center for Clinical Informatics, Salt Lake City, Utah (GSB); Brigham Young University, Provo, Utah (MJL); PacificCare Behavioral Health, Santa Ana, Calif (ERJ); and the University of Utah, Salt Lake City, Utah (TM).

PacificCare Behavioral Health, Inc. (PBH) provided funding for this research as part of an ongoing program of outcomes management and quality-improvement research activities. Dr Jones is employed by PBH as vice president and chief clinical officer. Dr. Brown and Dr Lambert are independent researchers who have consulted for PBH over a number of years and received compensation for this work.

Address correspondence to: G. S. (Jeb) Brown, PhD, 1821 Meadowmoor Rd, Salt Lake City, UT 84117. E-mail: jebbrown@clinical-informatics.com.

ly trained to provide an empirically supported treatment for the specific disorder, (2) therapists do deliver the treatment that they proposed, and (3) all therapists deliver these treatments with equal competence. Supposedly, the therapists' credentials speak to the first issue, whereas the therapists' performance with regard to the second issue is almost impossible for an MBHO to determine.

However, data from clinical trials do speak to the third issue: the assumption of equal competence. Reanalysis of past clinical trial data, as well as a number of more recent studies, reveal that there is a considerable amount of variability in outcomes across individual therapists, even in well-controlled clinical trials.<sup>8,10-18</sup> This variability is far from trivial.<sup>18</sup> Furthermore, even the most well-controlled and well-executed clinical trial to date, the National Institute of Mental Health Treatment of Depression Collaborative Research Program, found that the variance due to therapist far exceeded that of the treatment.<sup>15</sup> Therefore, variability among therapists has become one of the most critical areas of psychotherapy research.

It is beyond the scope of this article to review the broad area of research on what may contribute to this variability in therapist outcomes. However, a recent comprehensive review of the literature in this area arrived at the conclusion that readily measurable therapist variables such as age, sex, race, years of training, and type of degree explain little of the variance in outcomes.<sup>19</sup> Although research on inferred therapist traits, such as interpersonal style and the role of the therapeutic relationship, showed more promises in explaining treatment outcomes, evidence is too scant to warrant any conclusion.<sup>19,20</sup> If quality includes a domain of clinical outcomes, then quality assurance initiatives that ignore this variability in outcomes at the clinician level are unlikely to improve quality as promised.

PacifiCare Behavioral Health, Inc (PBH) differs from other national MBHOs with regard to how patient self-report outcome questionnaires are used as a critical component of its comprehensive quality-improvement program. PBH encourages its panel of psychotherapy providers to administer the outcome questionnaires at regular intervals in treatment to as many patients as possible. The PBH outcome data provide a unique opportunity to investigate this variability among therapist outcomes and its practical importance for behavioral health care management.<sup>21</sup>

PBH developed its ALERT clinical information system to help care managers and clinicians monitor and manage clinical outcomes. The first 3 authors (GSB, MJL, ERJ) participated in the development of the system, which was first implemented by PBH in 1999. The

current study analyzed the outcome data contained in the ALERT system to investigate the stability of this variability in therapist outcomes. In addition, the practical implications of this variability are discussed in light of quality improvement in an MBHO environment.

---

## METHODS

### Sample Description

The ALERT database contains outcomes data for 69 503 unique patients (79 748 episodes of care) who initiated treatment during the period from January 1999 through June 2004. Of these, 46 052 were treated by 1 of 5834 psychotherapists in private practice. Patients treated at a multidisciplinary group practice are excluded from this count and from subsequent analyses because the available dataset does not permit us to identify the treating clinician. The 46 052 patients treated by an individually identifiable clinician will be referred to as the total sample.

A subset of 281 therapists was selected for inclusion in this study based on their having a sample of at least 15 cases with change scores between January 1999 and December 2002 and at least 5 cases in the subsequent cross-validation period between January 2002 and June 2004. These clinicians treated a total of 10 812 patients (study sample) during the study period. This number constitutes 23.5% of the total sample.

Overall, the study sample was highly comparable to the total sample with respect to diagnoses and test scores. **Table 1** provides the breakdown by diagnostic groups. It should be noted that diagnostic data, which was obtained from the Provider Assessment Report, was only available for 33% of this sample.

Likewise, the study sample was comparable to the total patient sample with regard to sex and age. As is typical of outpatient treatment samples, approximately two thirds were female (63.5% in study sample; 64.5% in total sample), and juveniles under the age of 18 years comprised a quarter of the sample. Test scores at intake and change scores during treatment also were comparable between the study sample and total sample. Space limitations do not permit the use of tables to present this data, but these tables are available upon request.

The therapists in the study sample were comparable to the total sample of therapists with regards to age and years of experience. The mean age of the therapists in the study sample was 55 years (SD = 7 years), compared with 54 years (SD = 8 years) for the total sample. The therapists in the study sample had a mean of 22 years (SD = 7 years) of postlicensure experience, compared

**Table 1.** Breakdown of Diagnoses by Samples and Age Groups

| Diagnostic Groups                        | Study Sample, % |             |          | Total Sample, % |             |          |
|--|-----------------|-------------|----------|-----------------|-------------|----------|
|  | Adults          | Adolescents | Children | Adults          | Adolescents | Children |
| Attention deficit/hyperactivity disorder | 0.4             | 6.6         | 15.3     | 0.5             | 6.5         | 17.1     |
| Adjustment disorders                     | 32.8            | 24.8        | 27.8     | 28.3            | 23.3        | 31.2     |
| Anxiety disorders                        | 11.8            | 7.9         | 13.5     | 12.1            | 7.9         | 11.5     |
| Behavior disorders                       | 0.5             | 8.4         | 10.0     | 0.5             | 7.2         | 10.2     |
| Bipolar disorder                         | 2.8             | 2.7         | 1.0      | 3.6             | 3.2         | 1.2      |
| Depression (including dysthymia)         | 44.5            | 41.0        | 17.6     | 47.1            | 43.3        | 15.1     |
| Posttraumatic stress disorder            | 2.2             | 2.6         | 4.8      | 2.3             | 2.1         | 3.5      |
| Personality disorders                    | 2.5             | 3.5         | 6.6      | 2.8             | 3.7         | 7.4      |
| Psychotic disorders                      | 0.2             | 0.6         | 0.0      | 0.4             | 0.3         | 0.1      |
| Substance abuse disorders                | 1.0             | 1.0         | 0.2      | 1.2             | 1.2         | 0.2      |
| Other diagnoses                          | 1.4             | 0.9         | 3.5      | 1.2             | 1.3         | 2.5      |

with 22 years (SD = 8 years) for the total sample. Female therapists comprised 70% of the study sample, compared with 63% of the total sample. With regard to licensure type, marriage and family therapists were disproportionately present in the study sample, comprising 48% of the study sample compared with 28% of the total sample of therapists. The percentage of psychologists, social workers, and other licensed mental health professions was lower in the study sample. The reason for this disproportionate representation is unclear, unless marriage and family therapists as a whole are more inclined than other professions to use outcome measures.

**Outcome Measures**

The ALERT system uses 2 outcome measures: the Life Status Questionnaire (LSQ) for adults and Youth Life Status Questionnaire (YLSQ) for children and adolescents.<sup>22-25</sup> The YLSQ can be completed either by a parent or a guardian for younger children, or by adolescents on their own. In the remainder of the article, the abbreviation Y/LSQ will be used when referring to both measures simultaneously.

The majority of items on these measures inquire about psychiatric symptoms (primarily symptoms of anxiety and depression), while a subset of items also inquire about interpersonal relationships and functioning in daily activities. The items ask patients to indicate how often the statement is true for them over the past week, responding on a 5-point Likert scale with anchors ranging from “never” to “almost always” (values scored as 0 to 4). Higher scores indicate greater

severity of symptoms, subjective distress, and/or impaired functioning.

The outcome measures demonstrate excellent psychometric properties, with a Cronbach’s alpha of .93 and higher. The measures also have consistently high correlations with other well-established self-report questionnaires widely used in psychotherapy research.<sup>22,23</sup> Both the LSQ and YLSQ have been administered to large samples of patients in treatment (clinical sample) and individuals not seeking clinical services (community sample). These samples provide normative information on the means and standard deviations of the clinical and community samples, which were used to calculate clinical cutoff scores using the method recommended by Jacobson and Truax.<sup>26</sup> Scores at or above the clinical cutoff are determined as more characteristic of individuals seeking behavioral health services.

**Data Collection and Clinical Feedback**

Therapists are asked to administer the questionnaires at sessions 1, 3, and 5 and every fifth session thereafter. Completed Y/LSQs are faxed to a central toll-free number, where optical mark recognition software is used to read the data from the completed form. These files are then uploaded to the ALERT system, which scores the questionnaires, calculates the rate of each patient’s change compared with normative expectations, and checks for values on critical items (eg, self-harm, substance abuse).<sup>27-30</sup> The system also evaluates data obtained from the clinician such as the patient’s diagnosis.

The ALERT system notifies therapists via regular mail regarding cases with high-risk indicators, drawing the therapist's attention to the test scores and critical items, and offers to authorize more sessions as needed.<sup>30</sup> Therapists also are notified of cases with good outcomes, as evidenced by test scores within the normal range. On a quarterly basis, therapists are provided summary data on all of their cases within the past 36 months.

Therapists are given no financial incentives to use the outcome questionnaires. However, in late 2002, the system was enhanced so that submission of completed outcome questionnaires resulted in an automatic authorization for additional sessions for the particular case. Authorization is granted regardless of the test scores or the responses on the critical items (eg, suicidal ideation), and this provides some incentive for clinicians to submit data.

### Study Design

The design of this study utilizes a cross-validation strategy. Specifically, therapists' outcomes for patients initiating treatment between January 1999 and December 2002 were used as the baseline. This baseline period corresponds to the period prior to implementing the automated authorization process. The therapist outcomes in the following period between January 2003 and June 2004 were used for cross-validation.

For the purpose of this study, a treatment episode was defined as a period of consecutive administration of the Y/LSQs with no intervals between administrations more than 90 days. Therefore, if more than 90 days has elapsed between 2 Y/LSQ scores, the former administration is considered to be the posttreatment score of an episode, and the latter administration is considered to be the intake score of a new episode.

The choice of a maximum lapse of 90 days between measurements to define an episode is of course to some extent arbitrary, although we chose this interval because it reasonably fit our collective experiences as clinicians. Different time lapses were tested, including a 180-day period. This resulted in a small decrease in the number of episodes, but otherwise no meaningful difference in the assessment of change.

To ensure independence of observations, a case was defined as the patient's first treatment episode with at least 2 Y/LSQ scores under a single therapist. This means that for patients treated multiple times, only their first episode was included in the analysis. Cases with only 1 Y/LSQ score for an episode also were excluded because calculation of the change score requires at least 2 measurements. Finally, cases with outcome data submitted by more than 1 clinician with overlapping dates of service were excluded because of the difficulty of

assigning the outcome to more than 1 clinician.

This method resulted in the inclusion of 281 therapists treating 10 812 patients. The average number of cases per therapist during the baseline period was 26.5 (SD = 12.4) with a median of 22 and a range of 15 to 78. During the cross-validation period, the average number of cases per therapist was 12.0 (SD = 8.2) with a median of 10 and a range of 5 to 73.

Therapist outcome was determined by the therapists' average residualized change score on the Y/LSQ rather than the average raw score difference between the intake and posttreatment Y/LSQ. This was done so that differences in the types of patients seen among different therapists (ie, case mix) did not confound the therapists' average outcomes.

Case mix was controlled using a multiple regression model. The residualized change score for each patient was calculated as the difference between the predicted final score (based on the case mix model) and the actual final score. Thus, if a patient's residualized score was greater than 0, that indicated that the patient improved more than what would be expected based on the particular case mix. Specifically, the following case mix variables were controlled for: intake score, age group (child, adolescent, adult), sex, diagnostic group (8 groupings based on the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [DSM-IV]* diagnostic code), and session number of the first assessment in the treatment episodes. Use of the session number when the Y/LSQ was first administered as a predictor controls for failure to collect a baseline score at the first session.

The intake score proved to be the strongest predictor of the test score at the end of the treatment, accounting for approximately 49% of the variance in the final scores. It is important to control for this variable because patients with high intake scores average more change than patients with low scores. This is in part due to regression to the mean, but the change observed with the Y/LSQ data exceeds what is expected from this statistical artifact. The other case mix variables also were predictive of outcomes, even after controlling for intake score, and thus were included as well. However, the other case mix variables (diagnosis, age, sex) only explained an additional 2% of the variance in combination.

Outcomes were assessed based on intent to treat rather than using predetermined criteria for treatment completion. In other words, all cases with intake and posttreatment scores were included in the evaluation of effectiveness, even if the patient left treatment after as few as 3 sessions. This was done to provide a conservative estimate of the therapists' effectiveness, rather than overestimating outcomes by assessing effectiveness based only on "successfully completed" cases.



The therapists included in the study were rank-ordered based on their residualized change scores during the baseline period, and they were classified as either highly effective or other. Based on the rankings, 71 therapists (25%) were classified as highly effective by averaging a residualized change score of 2.8 or greater. Therapist outcomes then were cross-validated using their separate sample of cases between January 2003 and June 2004.

RESULTS

Table 2 presents the intake and posttreatment scores during the baseline period, broken out by therapist group (ie, highly effective and others). During the baseline period, there were statistically significant differences in the residualized change scores as expected. The highly effective therapists averaged over 3-fold as much change per case based on raw scores ( $P < .001$ ) and averaged a difference of 5.9 points in mean residualized change scores ( $P < .001$ ) compared with the other therapists. The number of days between first and last test administration was similar, though the highly effective group averaged 5 days longer.

Table 3 provides the results obtained during the cross-validation period. As expected, the difference between the highly effective clinicians and the rest of the sample decreased at cross-validation due to regression toward the mean, but the remaining difference was substantial. The highly effective group continued to average greater change, with a difference of 2.8 points in the mean residualized change score ( $P < .001$ ) compared with the other therapists.

To examine therapist outcomes for patients that would be deemed similar to those observed in clinical trials, analysis of the differences between the 2 groups of therapists was restricted to patients with intake scores above the clinical cutoff. Table 4 provides the results for this restricted sample.

Consistent with the previous analysis, the highly effective therapists averaged significantly better

outcomes with patients above the clinical cutoff than the other therapists did, resulting in a difference of 2.7 points in the mean residualized change score ( $P < .001$ ). This demonstrated that the impact of therapist effectiveness remained robust even among cases with greater acuity.

DISCUSSION AND CONCLUSIONS

The purpose of this study was to assess the variability and stability of therapist outcomes. The results provide evidence that therapists in an MBHO environment vary substantially in their patient outcomes, and that these differences are robust.

Table 2. Baseline Sample: 1999 through 2002

| Sample Characteristics  | Highly Effective Therapists (n = 71) | Other Therapists (n = 210) | Difference | P     |
|---|--------------------------------------|----------------------------|------------|-------|
| Total number of cases   | 1800                                 | 5640                       |            |       |
| Mean number of cases per therapist                                      | 25.4                                 | 26.9                       |            |       |
| Mean intake score (SD)  | 50.0 (18.4)                          | 49.8 (18.2)                | 0.2        | >.75  |
| Mean change score (SD)  | 10.0 (15.0)                          | 3.0 (13.5)                 | 7.0        | <.001 |
| Mean residualized change score (SD)                                     | 4.9 (13.3)                           | -1.0 (-12.4)               | 5.9        | <.001 |
| Mean number of days in treatment episode (first to last administration) | 70.6                                 | 65.0                       | 5.5        | <.01  |

Table 3. Cross-validation Sample: 2003 Through June 2004

| Sample Characteristics  | Highly Effective Therapists (n = 71) | Other Therapists (n = 210) | Difference | P     |
|---|--------------------------------------|----------------------------|------------|-------|
| Total number of cases   | 940                                  | 2432                       |            |       |
| Mean number of episodes per therapist                                   | 13.2                                 | 11.6                       |            |       |
| Mean intake score (SD)  | 48.2 (17.8)                          | 49.5 (17.3)                | -1.3       | <.05  |
| Mean change score (SD)  | 6.9 (13.4)                           | 4.5 (13.9)                 | 2.4        | <.001 |
| Mean residualized change score (SD)                                     | 2.4 (12.4)                           | -0.4 (12.5)                | 2.8        | <.001 |
| Mean number of days in treatment episode (first to last administration) | 71.0                                 | 67.1                       | 3.9        | >.1   |

**Table 4.** Cross-validation Sample: Clinical Range Only

| Sample Characteristics  | Highly Effective Therapists (n = 71) | Other Therapists (n = 210) | Difference | P     |
|---|--------------------------------------|----------------------------|------------|-------|
| Total number of cases   | 641                                  | 1742                       |            |       |
| Mean number of episodes per therapist                                   | 9.0                                  | 8.3                        |            |       |
| Mean intake score (SD)  | 57.0 (13.4)                          | 57.4 (13.3)                | -0.4       | >.5   |
| Mean change score (SD)  | 9.9 (13.4)                           | 6.9 (13.7)                 | 3.0        | <.001 |
| Mean residualized change score (SD)                                     | 2.1 (12.9)                           | -0.6 (12.8)                | 2.7        | <.001 |
| Mean number of days in treatment episode (first to last administration) | 70.9                                 | 68.0                       | 2.9        | >.3   |

### Limitations

As with any study, limitations need to be taken into account when interpreting these results. Although well-known case mix variables were controlled for, it is always possible that other unknown variables could substantially influence therapist outcomes. Regardless of the complexity or completeness of the case mix model, without random assignment of clients to therapists, it is a logical impossibility to rule out the potential impact of other unmeasured case mix variables. However, just as it seems unlikely that the case mix adjustment model is fully adequate to account for patient differences, it is equally unlikely that all of therapist differences reported in this study are due to undetected differences in case mix, especially in light of the substantial body of published research pointing to the existence of clinician effects in controlled trials.

It is possible that the patient outcomes of a single therapist may vary significantly across different diagnoses, age groups, or some other patient characteristic. The present study did not explore this question in greater detail because of problems of cell size. Disaggregating the treatment sample into the various diagnosis groups would have meant that the sample sizes within the multiple cells for each therapist would be only a few cases. It is logical to assume that certain diagnoses may pose particular challenges and are best treated by a specialist, and it does appear likely that clinician effectiveness would vary at least across age groups. Clinician effectiveness with adults cannot be expected to necessarily translate into effectiveness with adolescents or young children.

These and related questions of therapist variability across patient types and/or treatment methods will be investigated in future studies. The problem of cell sizes

will decrease going forward because of the continued rapid accumulation of data.

A potential bias in the data collection process relates to the automatic authorization process. The rate of clinician participation and Y/LSQ submissions had increased steadily between 1999 and 2002, but the implementation of the automatic authorization of additional sessions in late 2002 resulted in a dramatic increase in Y/LSQ submissions. Therefore, there is a probability that the change in the system may have biased the results in unknown ways. In

addition, as would be expected in such a large system of care, there was significant variability in the rate of compliance with the data collection protocol, both at the therapist and patient levels. Not surprisingly, consistency of collection at the patient level covaried with the consistency of the therapist, confirming the commonsense view that the therapists have a large impact on the likelihood that the patient will complete the questionnaire.

Thus, this variability in compliance also may have influenced the results in unknown ways. Failure to complete an assessment within the first 2 sessions resulted in capturing slightly less change over the treatment episode, but this artifact was adjusted for in the case mix model used. Beyond that artifact, there was not evidence that outcomes varied systematically with compliance. It was true that the highly effective clinicians had larger sample sizes, but analysis of claims data confirmed that this was because these clinicians treated a disproportionate number of the patients in the sample. Still, the possibility that some clinicians were selectively submitting forms for patients with good outcomes can't be ruled out. Furthermore, the effects of providing therapists with feedback could not be assessed with these data.<sup>31,32</sup>

Undeniably, patient self-report measures provide only a single perspective, which is that of the patient. It is therefore likely that use of other perspectives such as outcomes measured by the therapists might have yielded different results. Therapist-rated measures are known to show more change than patient self-report measures.<sup>33</sup> However, use of therapists' assessment of their own effectiveness poses obvious problems with regards to potential for bias.

Some may argue that use of more objective but time- and labor-intensive measures (eg, school or work

attendance/performance) or batteries of measures looking at outcomes from multiple perspectives would yield more reliable assessments of outcome. On the opposite end of the spectrum, others may argue for implementing low-cost customer satisfaction surveys as outcome measures.<sup>34</sup> However, it is unlikely that either of these approaches to assessment is appropriate to assess real-time patient progress in psychotherapy, compared with outcome questionnaires that were specifically designed for its purpose.<sup>35-39</sup> Although it is ideal to obtain multiple measures of outcome so as to decrease sources of measurement bias, time pressures of a real-world practice make this problematic. Thus, brief patient self-report measures that are easy to administer and score were chosen. Self-report questionnaires also have the benefit of not burdening the therapists with extra paperwork to complete, while potentially providing the therapists with clinical information otherwise not obtained.

Lastly, the current study provided no insight into what treatments were delivered by the therapists in both groups and how the treatments were delivered. Such information was beyond the scope of the available data. To this point, the importance of conducting clinical trials to empirically support treatments should not be dismissed, as potential causal effects of treatment can be determined only through experimental designs.<sup>40,41</sup> Therefore, to claim that this study provides support for discontinuing or devaluing empirically supported treatments would be a gross misinterpretation.

In summary, there are a number of limitations to this study arising from the naturalistic nature of the data and the inherent measurement error in any attempt to measure a construct as broad as "treatment outcome." Clearly, measurement efforts across such a large system of care involving thousands of clinicians pose many challenges for both collection and interpretation of the data that are avoided entirely in a well-designed clinical trial with random assignment. Therefore, the information culled from the data must be used cautiously, with all consideration for unknown sources of measurement error while simultaneously bearing in mind that there also is a risk to the patients if the data are not used for quality-improvement purposes.

### Implications

Despite the limitations of this study, the magnitude of differences in outcome among the therapists is sufficiently large to lend credence to the proposition those outcomes could be improved by focusing on these differences. With therapists differing significantly in their effectiveness, the patients are best served if the MBHO can identify and refer to effective therapists.<sup>42,43</sup>

The role of the MBHO in the outcomes-informed environment is still evolving. What responsibility, if any, do MBHOs have with regards to offering consultation to its provider networks on how to improve outcomes? Some providers may welcome such an offer, while others may reject it as unwarranted intrusion into the patient-therapist relationship. At the very least, MBHOs have a responsibility to the therapists to provide feedback on their outcomes, while pursuing policy for publishing the outcomes for subscriber access and encouraging further analysis of the data by independent investigators.

The outcome data are important to effective providers because these data permit them to make a strong case for the value of their services. MBHOs in the future may be valued more for their ability to steer patients to therapists with demonstrated records of effectiveness than their current strategy of cost containment and utilization management.

These results also demonstrate the practical utility and benefits of utilizing patient self-report outcome data as part of a quality-improvement program. Direct measure of patients' outcomes has a greater probability of leading to improved outcomes than more commonly used quality-improvement methods that focus on the method of treatment or other process variables. The data presented in this article speak to feasibility of implementing a system of quality improvement based on use of patient self-report outcome questionnaires to identify highly effective therapists.

### REFERENCES

1. **Chambless, DL, Hollon SD.** Defining empirically supported therapies. *J Consult Clin Psychol.* 1998;66:7-18.
2. **Task Force on Promotion and Dissemination of Psychological Procedures.** Training in and dissemination of empirically-validated psychological procedures: report and recommendations. *Clin Psychol.* 1995;48(1):3-23.
3. **Hedges LV, Olkin I.** *Statistical Methods for Meta-Analysis.* San Diego, Calif: Academic Press; 1985.
4. **Raudenbush SW, Bryk AS.** *Hierarchical Linear Models: Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, Calif: Sage; 2002.
5. **Ahn H, Wampold BE.** Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *J Counseling Psychol.* 2001;48:251-257.
6. **Luborsky L, Rosenthal R, Diguier L, et al.** The dodo bird verdict is alive and well—mostly. *Clin Psychol Sci Pract.* 2002;9(1):2-12.
7. **Shapiro DA, Shapiro D.** Meta-analysis of comparative therapy outcome studies: a replication and refinement. *J Consult Clin Psychol.* 1982;92:581-604.
8. **Wampold BE.** *Great Psychotherapy Debate: Models Methods and Finding.* Mahwah, NJ: Erlbaum; 2001.
9. **Wampold BE, Mondin GW, Moody M, et al.** A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, "all must have prizes." *Psychol Bull.* 1997;122:203-215.
10. **Blatt SJ, Sanislow CA, Zuroff DC, Pilkonis PA.** Characteristics of effective therapists: further analyses of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol.* 1996;64:1276-1284.
11. **Crits-Christoph P, Baranackie K, Kurcias JS, et al.** Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychother Res.* 1991;1:81-91.
12. **Crits-Christoph P, Mintz J.** Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *J Consult Clin Psychol.* 1991;59:20-26.

13. **Elkin I.** A major dilemma in psychotherapy outcome research: disentangling therapists from therapies. *Clin Psychol Sci Pract.* 1999;6:10-32.
14. **Huppert JD, Buřka LF, Barlow DH, Gorman JM, Shear MK, Woods SW.** Therapists, therapist variables, and cognitive-behavioral therapy outcomes in a multicenter trial for panic disorder. *J Consult Clin Psychol.* 2001;69:747-755.
15. **Kim DM, Wampold BE, Bolt DM.** Therapist effects in psychotherapy: A random effects modeling of the NIMH TDCRP data. *Psychother Res.* In press.
16. **Lambert MJ, Ogles BJ.** The efficacy and effectiveness of psychotherapy. In: Lambert MJ, ed. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change.* New York: John Wiley & Sons; 2004:139-193.
17. **Luborsky L, Crits-Christoph P, McLellan T, et al.** Do therapists vary much in their success? Findings from four outcome studies. *Am J Orthopsychiatry.* 1986;56:501-512.
18. **Okiishi J, Lambert MJ, Nielsen SL, Ogles BM.** Waiting for supershrink: an empirical analysis of therapist effects. *Clin Psychol Psychother.* 2003;10:361-373.
19. **Beutler LE, Malik M, Alimohamed S, et al.** Therapist variables. In: Lambert MJ, ed. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change.* New York: John Wiley & Sons; 2004:227-306.
20. **Norcross JC.** Empirically supported therapy relationships. In: Norcross JC, ed. *Psychotherapy Relationships That Work.* New York: Oxford University Press; 2002:3-32.
21. **Brown GS, Burlingame GM, Lambert MJ, et al.** Pushing the quality envelope: a new outcomes management system. *Psychiatr Serv.* 2001;52:925-934.
22. **Lambert MJ, Hatfield DR, Vermeersch DA, et al.** *Administration and Scoring Manual for the LSQ (Life Status Questionnaire).* Salt Lake City, UT: American Professional Credentialing Services; 2001.
23. **Burlingame GM, Jasper BW, Peterson G, et al.** *Administration and Scoring Manual for the YLSQ.* Salt Lake City, UT: American Professional Credentialing Services; 2001.
24. **Vermeersch DA, Lambert MJ, Burlingame GM.** Outcome Questionnaire: item sensitivity to change. *J Pers Assess.* 2002;74:242-261.
25. **Wells MC, Burlingame GM, Lambert MJ, Hoag M.** Conceptualization and measurement of patient change during psychotherapy: development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy.* 1996;33:275-283.
26. **Jacobson NS, Truax P.** Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59:12-19.
27. **Matsumoto K, Jones E, Brown GS.** Using clinical informatics to improve outcomes: a new approach to managing behavioral healthcare. *J Information Technol Health Care.* 2003;1(2):135-150.
28. **Brown GS, Jones ER, Betts W, Wu J.** Improving suicide risk assessment in a managed-care environment. *Crisis.* 2003;24(2):49-55.
29. **Brown GS, Herman R, Jones ER, Wu J.** Improving substance abuse assessments in a managed care environment. *Jt Comm J Qual Safety.* 2004;30:448-454.
30. **Brown GS, Jones ER.** Implementation of a feedback system in a managed care environment: what are patients teaching us? *J Clin Psychol.* 2005;61(2):187-198.
31. **Lambert MJ, Whipple JL, Smart DW, et al.** The effects of providing therapists with feedback on patient progress during psychotherapy: are outcomes enhanced? *Psychother Res.* 2001;11:49-68.
32. **Lambert MJ, Whipple JL, Hawkins EJ, et al.** Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clin Psychol Sci Pract.* 2003;10:288-301.
33. **Lambert MJ, Hatch DR, Kingston MD, Edwards BC.** Zung, Beck, and Hamilton Rating scales as measures of treatment outcome: a meta-analytic comparison. *J Consult Clin Psychol.* 1986;54:54-59.
34. **Seligman MEP.** The effectiveness of psychotherapy: the *Consumer Reports* study. *Am Psychol.* 1995;50:965-974.
35. **Sechrest L, McKnight P, McKnight K.** Calibration of measures for psychotherapy outcome studies. *Am Psychol.* 1996;51:1065-1071.
36. **Brock TC, Green MC, Reich DA, Evans LM.** The *Consumer Reports* study of psychotherapy: invalid is invalid [letter]. *Am Psychol.* 1996;51:1083.
37. **Hunt E.** Errors in Seligman's "The effectiveness of psychotherapy: the *Consumer Reports* study" [letter]. *Am Psychol.* 1996;51:1082.
38. **Kotkin M, Daviet C, Gurin J.** The *Consumer Reports* mental health survey. *Am Psychol.* 1996;51:1080-1082.
39. **Mintz J, Drake RE, Crits-Christoph P.** Efficacy and effectiveness of psychotherapy: two paradigms, one science. *Am Psychol.* 1996;51:1084-1085.
40. **Shadish WR, Matt GE, Navarro AM, et al.** Evidence that therapy works in clinically representative conditions. *J Consult Clin Psychol.* 1997;65:355-365.
41. **Shadish WR, Matt GE, Navarro AM, Phillips G.** The effect of psychological therapies under clinically representative conditions: a meta-analysis. *Psychol Bull.* 2000;126:512-529.
42. **Brown GS, Dreis S, Nace D.** What really makes a difference in psychotherapy outcomes? And why does managed care want to know? In: Hubble MA, Duncan BL, Miller SD, eds. *Heart and Soul of Change.* Washington, DC: American Psychological Association Press; 1999:389-406.
43. **Wampold BE, Brown GS.** Estimating therapist variability: a naturalistic study of outcomes in private practice. *J Consult Clin Psychol.* In press.