Psychotherapy Research, January 2007; 17(1): 1–14

Routledge
Taylor & Francis Group

PRESIDENTIAL ADDRESS

Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care

MICHAEL LAMBERT

Department of Psychology, Brigham Young University

(Received 18 August 2006; revised 23 September 2006; accepted 24 September 2006)

Abstract
The findings of a decade-long program of research aimed at tracking patient treatment response are summarized. Outcome measures were developed and applied on a weekly basis with patients undergoing therapy in routine practice. Treatment response was defined as reliable and clinically significant change, and longitudinal statistical models for identifying poor progress were applied. The author was able to estimate the amount of therapy necessary for recovery, study early dramatic improvers, and assess the consequences of providing practitioners with alarm signals when patients were off track for a positive outcome. Results indicate that integrating treatment response research into routine mental health care reliably improved positive outcomes and reduced negative outcomes. Future directions for this line of action research are suggested.

The effects of psychotherapy have been found to be generally positive across a variety of studies and over a substantial period of time dating back to the 1930s (Lambert & Ogles, 2004). Considerable effort has been made to further improve patient outcomes; the most touted attempts take the form of controlled comparative outcome studies using clinical trials methodology and treatment manuals that carefully define and prescribe therapist actions. Results from these studies have led to recommendations for best practices (specific treatments for specific disorders) based on criteria proposed by a variety of scientific committees. Such recommendations have met with considerable controversy and stirred widespread debate about the wisdom of assuming that effective treatments can be so simply defined and delivered.

Even if one were to accept the idea that the delivery of the right treatment for the right disorder had an appreciable consequence for patients seen in routine care, it is also clear that these same treatments are occasionally harmful or result in no detectable change in some patients. In randomized clinical trials (RCTs), in which treatments are offered by carefully selected and closely supervised therapists to carefully screened patients with a specific disorder, about 35% to 40% of patients experience no benefit and a small group of patients, perhaps between 5% and 10%, deteriorate (Hansen, Lambert, & Forman, 2002).

In our program of research, we have endeavored to improve psychotherapy outcome in routine care through monitoring patient progress and providing this information to clinicians in order to guide ongoing treatment, especially for the patient who is not having a favorable response to treatment (not-on-track [NOT] cases). This research is an extension of quality-assurance action research and represents one effort to bridge the gap between research and practice while enhancing patient outcome before treatment termination. It is also well suited to models of care in which clinicians attempt to step up or step down the intensity of treatments after assessing a patient's treatment response (Otto, Pollack, & Maki, 2000). I have documented here our progress in enhancing patient outcome by describing how we have operationally defined outcome, treatment success and failure, and our experience with monitoring treatment response and feeding back this information to therapists and patients before treatment termination. In addition,

Correspondence: Michael Lambert, Department of Psychology, Brigham Young University, 272 TLRB, Provo, UT 84602. E-mail: michael_lambert@byu.edu

ISSN 1050-3307 print/ISSN 1468-4381 online
DOI: 10.1080/10503300601032506

some important incidental findings that have relevance for practice are presented.

Defining Outcome

Our program of research began with the development of a suitable but brief (5–10 min) outcome measure, the Outcome Questionnaire-45 (OQ-45; Lambert, Morton, et al., 2004), which provides both a measure of weekly change as well as the criterion measure for classification of a patient into outcome categories (improvers, no changers, and deteriorators) after treatment. It was designed to assess four domains of functioning: symptoms of psychological disturbance (mainly depression and anxiety); interpersonal problems; social role functioning (e.g., problems at work or school); and quality of life (positive aspects of life satisfaction). Factor analysis of the OQ-45 has confirmed the presence of symptom (depression, anxiety, and somatic), interpersonal, and social role subscales but also that a general mental health factor is described by the total score (de Jong et al., in press; Mueller, Lambert, & Burlingame, 1998). Even though the diverse elements of outcome measured by the scale have been validated, we have focused our outcome studies on the total score because of the small number of items that have been used to measure some factors.

Two large-scale studies have shown that the OQ-45 items, subscales, and total score are sensitive to change in patients over short time periods (9 weeks) but remain stable in untreated individuals (Vermeersch, Lambert, & Burlingame, 2000; Vermeersch et al., 2004). We consider these sensitivity studies as good examples of one of our unique contributions to methodological advancements in scale development. Through the collection of weekly measurements of patient progress, we were able to assess the speed of change on each item of the OQ-45 and its value as a scale for change measurement (in contrast to usual procedures that emphasize diagnostic value or usefulness in treatment planning). Based on these studies, the OQ-45 appears to be well suited for tracking patient treatment response and can be used on a weekly basis with limited test–retest artifacts (Durham et al., 2002). Both these studies suggest that changes on the OQ-45 subscales do not occur at the same speed; changes in interpersonal functioning lag behind changes in symptoms and social role functioning.

A series of studies have provided information about the psychometric properties of the OQ-45. The internal consistency of the OQ-45 is persistently high ($\alpha = .93$), and the test–retest reliability typically produces coefficients of about .84 (Lambert, Morton, et al., 2004) in untreated individuals. Concurrent validity of the OQ-45 has been demonstrated through correlates with a variety of scales (e.g., Symptom Checklist-90-R [Derogatis, 1983], $r = .78$; Beck Depression Inventory [Beck, Ward, Mendelson, Mock, & Erbaugh, 1961], $r = .80$; Inventory of Interpersonal Problems [Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988], $r = .55$; Social Adjustment Rating Scale [Weissman & Bothwell, 1976], $r = .56$). These psychometrics have held up across a variety of cultures (e.g., Mexico, Chile, Puerto Rico, Germany, Netherlands, and Italy) and languages other than English.

The norms for the OQ-45 are based on data collected nationally (Lambert et al., 1996; Umphress, Lambert, Smart, Barlow, & Clouse, 1997) and indicate that it discriminates well between client and nonclient samples. This is important because an essential aspect of our program of research involves using these normative data to understand the clinical significance (recognizable, substantial, and practically important) of change for each patient. Using formulas developed by Jacobson and Truax (1991), clinical and normative data for the OQ-45 were analyzed by Lambert et al. (1996) to provide cutoff scores for the reliable change index (14 points) and clinically significant (CS) change (dysfunctional/functional cut-off: 64/63). This latter cut-off score requires that the patient score within 1 *SD* of the mean of normal functioning ($M = 45$). Individuals are defined as recovered when they leave therapy with an OQ-45 score that has improved by at least 14 points and is below 64. Patients who improve by 14 points are considered improved and those who worsen by at least 14 points are regarded as deteriorated if they leave treatment in the dysfunctional range. These cut-off scores provide a means for individualizing change at the end of treatment and follow-up and can also be used to raise questions about the necessity of further treatment. They have the advantage of providing absolute standards for classification that apply across all studies we conduct rather than having standards of improvement and deterioration that are sample specific and vary from study to study.

Support for the validity of the OQ-45's reliable change and clinical significance cut-off scores have been reported by Lunnen and Ogles (1998), Beckstead et al. (2003), and Bauer, Lambert, and Nielsen (2004). This research suggests that the Jacobson–Truax formulas provide a sound basis for estimating cut-off scores and that classification of change based on other measures results in considerable consensus with OQ-45 classifications. Despite this encouraging evidence, much additional work is necessary to understand the relationship between these cut-off

scores and their practical meaning for overall patient health and well-being.

In the course of collecting weekly outcome ratings from patients undergoing treatment in routine clinical practice, a number of findings have emerged that have clear implications for social policy and theoretical concerns. Before highlighting the research aimed at improving outcomes for specific clients, I address several important findings.

Treatment Length

The amount of psychotherapy necessary to bring about positive outcomes has been a concern for decades and remains a topic of serious debate (Lambert, Bergin, & Garfield, 2004). Early concerns about excessively long, mainly psychodynamic therapy resulted in rejections of this form of care in the community mental health movement that emerged in the 1950s and 1960s because long treatment lengths were associated with unacceptably long waiting lists. Traditional long-term psychodynamic treatments were shortened, with numerous planned short-term therapies (25 or fewer sessions) being developed and advocated in the 1960s. More recent concern about treatment length has been associated with the management of skyrocketing medical costs and based on the assumption that positive outcomes can be achieved with few sessions. The usual method of cutting costs in the United States has been to limit treatment to four to eight sessions, providing standard limits for all patients.

It should be noted that many organizations claim to be delivering evidence-based services while ignoring the fact that the clinical trial treatments that they are attempting to duplicate were based on 12 to 14 sessions, not four to eight sessions. Even a casual observer of the worldwide diversity of policy and practices concerning treatment length might be quite puzzled at the absence of the use of empirical data to shape important policy decisions concerning treatment length. In our mind, treatment length is most appropriately driven by patient treatment response rather than theoretical or cost-based decisions. How much therapy is enough?

Unfortunately, empirical data have not been particularly clear on this issue. For example, of the 156 studies published from 1950 to 1992 on this topic, 100 indicated a positive relationship between therapy duration and outcome and six indicated a negative relationship (Orlinsky, Grawe, & Parks, 1994). This correlational evidence is often used to support the general clinical belief that longer term therapies provide more benefit than short-term therapies and offers some assurance that longer therapy is better than shorter therapy; however, it provides little guidance about how much therapy is enough. In addition, the actual size of the correlations, although statistically significant, are very small ($r < .15$) and could support a conclusion that treatment length is relatively unimportant to positive patient outcome (see also Shapiro et al., 2003). Correlating the number of sessions patients have with eventual treatment outcome does not prove to be very informative for establishing policies for setting session limits.

Howard, Kopta, Krause, and Orlinsky (1986) made an early attempt to address the question, "How many sessions are required to produce meaningful change?" and suggested that 50% of patients derive such a benefit by the ninth session and that the dose–effect relationship in psychotherapy could best be characterized as negatively accelerating, a finding that is usually interpreted as indicating that each successive session of therapy has less impact on a patient's well-being. Following the lead of these researchers, we addressed the issue of dosage in an attempt to shed further light on the relationship between number of sessions received and patient benefit, with the advantage that we used a standard definition of good outcome and session-by-session data rather than statistical modeling based on pre–post data.

Survival Analysis

Using data ($N > 6{,}000$ patients) from a variety of U.S. clinical samples that received routine clinical care (ranging from employee assistance programs to community mental health centers), we have published five studies that estimated the dose–effect relationship (Anderson & Lambert, 2001; Hansen & Lambert, 2003; Kadera, Lambert, & Andrews, 1996; Snell, Malllinckrodt, Hill, & Lambert, 2001; Wolgast, Lambert, & Puschner, 2003). In each of these studies, patients rated their functioning at each session of treatment on the OQ-45. Our standard definitions of reliable and CS change were used as dependent variables (the binomial event of interest), and the Kaplan–Meier survival analysis statistic (or related technique) was used to provide a population estimate of the distribution model of sessions needed to reach the event of interest (normal functioning and reliable improvement). Across these studies, the session at which 50% and 75% of patients reached criteria for recovery ranged from Session 11 to 21 and Session 25 to 45, respectively.

These data provide an important source of information for policy decisions aimed at limiting the amount of psychotherapy a patient may receive. These data also allow government agencies, insurance companies, employers, clinics, and consumers

to understand the consequences of such policies. The findings suggest that limiting treatment to four sessions, a common practice, will result in insufficient treatment for the vast majority of patients. Even policy decisions that permit more than twice this amount of treatment cannot be justified on the basis of empirical findings. For example, data from the directors of counseling centers at 73 U.S. colleges and universities show that more than half of such centers have a self-imposed session limit of 10 sessions (Stone & McMichael, 1996). Our findings suggest that limiting psychotherapy to this amount is likely to be appropriate for fewer than half of the clients who seek treatment, even when the least rigorous criterion, reliable change, is used to define a positive treatment response. It is also clear that such a policy is particularly disadvantageous to the most disturbed clients, whose time to recovery is slower and who are most at risk for personal failures, self-harm, and hospitalization. The most important determinant of treatment outcome is initial patient functioning; variables such as severity, complexity of disorder, somatization, dual diagnosis, personality pathology, and the like are very important (Clarkin & Levy, 2004; Lambert & Anderson, 1996).

It is also worth noting that the estimated time to recovery for patients in routine care produces recovery estimates that are similar to actual CS outcomes of patients treated in RCTs (50% in 14 sessions; Hansen & Lambert, 2003). Actual CS outcome (rather than modeled outcome) of patients in routine care is closer to 30%, with an average dosage of about four sessions. These data suggest the need for providers and service managers to give high priority to keeping patients in treatment longer rather than setting session limits. In college counseling centers, about 90% of patients have terminated treatment by Session 15 (Nielsen & Lambert, 2006; Wolgast et al., 2003); the remaining 10% of clients use 30% to 40% of the total sessions delivered. Survival curves also show the same tendency to flatten as the number of sessions increases, with few patients meeting criteria for reliable change and CS change at subsequent sessions after Session 25, which is equivalent to 6 months of weekly treatment. We interpret such data as indicating that if a general session limit is set it would be more reasonable to think in terms of 20 to 25 sessions for most patients, depending on initial severity and treatment setting.

Although setting general session limits has great advantages for managing costs and making the life of administrators easier, it is not to the advantage of most patients and diminishes the overall value of services, especially for the most disturbed patients. Our data suggest that the highest quality of care will be provided by monitoring each client's treatment response and making treatment length a function of treatment response and mental health status rather than arbitrary or theory-driven limits. Limiting treatment in such a circumstance requires frequent assessment of patient well-being and would be instigated only when it becomes clear that no further progress can be expected if a patient continues to receive the same treatment.

Managing Treatment by Measuring Treatment Response and Outcome

An alternative to setting general session limits is to make the length of treatment responsive to each individual patient. A primary focus of our research has been to understand and apply information about usual treatment response in routine care through the use of longitudinal data analysis

Expected Treatment Response

Survival statistics are ideal for providing estimates of an event of interest (e.g., recovery) that can be used to guide policy decisions. Different statistical procedures are necessary to model the shape and speed of change over time for the purpose of understanding expected treatment response for the individual patient. For this purpose, we have used hierarchical linear modeling (HLM) of session-by-session ratings of functioning produced by patients. HLM is especially useful in analyzing longitudinal data from routine care, a situation in which patients begin treatment at diverse levels of initial disturbance and have treatment lengths that are highly varied. The results of such modeling can be examined for subsets of patients who meet criteria for an event such as clinically significant or reliable change (or deterioration) but are most valuable as a means of establishing session-by-session expectations for a course of psychotherapy in relation to unique levels of initial disturbance.

Finch, Lambert, and Schaalje (2001) applied HLM to a large database consisting of 11,492 patients treated in a variety of settings, including employee assistance programs, university counseling centers, outpatient clinics, private practice, and a clinical psychology training clinic. An initial graphic analysis of the data revealed decelerating growth curves similar to those identified in dose–response studies, a lawful linear relationship between the log of the number of sessions and the normalized probability of patient improvement, again illuminating the tendency for larger and larger doses (number of sessions) in order to find a higher percentage of recovered patients.

Although the data set used to model treatment response was large, it was not of sufficient size to be able to establish an individual recovery curve for each intake score because the statistical techniques require a larger number of cases for reliable modeling. OQ-45 scores falling at the extremes of the continuum are quite rare. Therefore, the full range of scores was divided into distinct groups by percentiles. This yielded 50 groups, identified by intake score, with no fewer than 220 patients in each band, representing approximately 2% of the total sample with intake score increments as small as 1 point at the group average and a larger spread between intake scores at the two ends of the continuum. The resulting groups of data were analyzed using the PROC MIXED functions of the statistical analysis system to generate a linear model for recovery curves. This linear model allowed comparisons of individuals even when OQ-45 scores were missing at different session numbers and the ultimate number of sessions, length of time between sessions, and overall length of therapy were different among clients. Essentially, a separate regression line and error estimate was generated for each patient in the analysis.

Tolerance intervals calculated in this model allowed for the identification of the 10% and 15% of patients in a given sample whose rate and trajectory of progress, respectively, deviated significantly from what might be the predicted course of recovery for others entering therapy with a similar intake score. The tolerance intervals created were primarily aimed at identifying this 10% of the patient population who deviate from the recovery track and formed the core of an empirical warning system by providing table values and charts of predicted therapeutic gains against which the data of any given patient can be compared. If at any session after intake the OQ-45 total score for a patient does not exceed the tolerance interval, then therapy is judged as proceeding as anticipated for this particular patient, and a green light message can be given as feedback for therapists to proceed as they have been. If the same OQ-45 score falls outside of the upper 15% and does not surpass the upper bound of the 10% tolerance interval, the patient is considered to be deviating by more than 1 *SD* from what is expected of a typical person at this point in therapy, and the therapist would receive a yellow message as a warning to attend to this patient's progress. This 1 *SD* unit approximates a 14-point increase in the OQ score (the marker for reliable change) but varies depending on the patient's initial level of disturbance. If this same OQ-45 score falls above the upper limits of the 10% interval, then the patient is deviating significantly in a negative direction from what is predicted at this point in therapy. The 10% boundary is consistent with the estimate that about 10% of patients deteriorate after psychotherapy (Lambert & Ogles, 2004). At this point, the therapist would receive a red warning message that therapy may be heading toward an unsuccessful conclusion and that the therapist needs to consider an alternative course of action. In our program of research, we consider patients whose treatment response crosses either the yellow or red boundary to be alarm-signal or NOT cases.

An opposite pattern can be described for scores falling on the lower side of the low tolerance bands, indicating that the individual is deviating significantly in a positive direction from what is normally expected at this point in therapy, and the therapist would receive a blue message indicating confidence that the patient will leave treatment with a positive outcome and maintain gains for at least 1 year (Haas, Hill, Lambert, & Morrell, 2002).

Early, Dramatic Treatment Response

Before turning to the effects of providing therapists with alarm signals for patients who are deviating from expected recovery, presentation of findings concerning extraordinary positive courses of treatment is in order. Few studies have investigated the timing of response in psychotherapy. Among those studies that have addressed this issue, the apparent finding is that early dramatic (as opposed to delayed or worsening) response to psychotherapy is positively related to better intermediate and long-term outcomes. For example, Renaud et al. (1998) examined rapid response in adolescents treated for major depressive disorder and found that early responders, compared with initial nonresponders, showed better outcomes at the end of treatment and after 1- and 2-year follow-up assessments. Ilardi and Craighead (1994) examined several studies of cognitive–behavioral treatment for depression. They argue that temporal changes (early rapid response) observed during the first 4 weeks of therapy support the notion that most changes occur before specific techniques have been administered in any relevant dose, a conclusion that has been challenged by Tang and DeRubeis (1999) in their analysis of sudden gains in depression. The findings in this area are not limited to depression; similar findings have been found in addictive disorders, panic, and bulimia nervosa. For example, numerous authors have suggested that early reductions in bingeing and purging are excellent predictors of final treatment response, with early response occurring within the first 4 weeks (five to six sessions) of treatment (e.g., Agras et al., 2000).

Two significant problems in the study of early response to treatment are apparent in the literature. There is a lack of agreement regarding how to define early treatment response. Among the various definitions of early response, Fennel and Teasdale (1987) utilized a median split to determine response rate, whereas others have used a single item from a questionnaire, the extent to which change occurred in the clinician's rating of improvement from one week to the next, determining whether the clinician rated the client as having minimal or no psychopathology by Week 1 or 2 of treatment, improvement of greater than 50% over a prespecified number of sessions and the like. An equally important problem is that few studies have actually measured response on a weekly basis; therefore, most studies have been unable to accurately identify the point in time at which a client responded to treatment. Typically, studies have relied on reassessment conducted at midtreatment points in clinical trials.

Haas et al. (2002) used a unique method of quantifying early dramatic treatment response by using session-by-session ratings of client symptomatology in relation to expected recovery curves, as just described (blue signals). This strategy allowed for the creation of a standardized method for determining the timing of treatment response and its size in relation to a typical response, thereby providing a data-based standard rather than a sample-relative standard of early response. Data analytic techniques allowed for examination of early dramatic response as both a continuous and a discrete variable aimed at evaluating the clinical significance of change.

Participants were 147 clients who completed the OQ-45 (Lambert, Morton et al., 2004) at intake, before each counseling session, and at a follow-up 6 to 24 months after termination of services. The method of identifying patients who had an early positive response to treatment was based on calculating a difference score between the expected and actual change for each participant; that is, each patient's response to treatment was judged against that of patients who showed similar levels of initial disturbance (i.e., had the same initial score on the OQ-45). Higher average difference scores indicate an early positive response, and negative scores indicate that a given participant responded to treatment more slowly than was expected. Clients in this study were classified into outcome categories at termination and follow-up based on clinical significance criteria (Jacobson & Truax, 1991).

Hierarchical regression procedures, controlling for initial severity, were used to address the question of whether response rate is predictive of long-term outcomes. The dependent variable in each instance was termination OQ-45 scores. Because it has been shown that the severity of a disorder is related to outcome as well as rate of response (Taylor & McLean, 1993), the client's intake OQ-45 score was entered first, and as expected, initial OQ-45 scores were highly predictive of termination OQ-45 scores. Results also indicated a significant negative effect for rate of response as measured by difference scores that followed after entry of the intake OQ-45. Thus, clients with faster rates of response to psychotherapy reported lower end-of-treatment OQ-45 scores. In the second hierarchical regression, which used the same format but utilized OQ-45 scores at follow-up as the dependent variable, an effect for early response and follow-up scores was also found. Clients with faster rates of response reported lower OQ-45 scores at follow-up an average of 1 year posttherapy.

An examination of response rate and the clinical outcome categories showed that early responders were more likely to benefit to a clinically significant degree at both termination and follow-up. These results suggest that participants in the improved and recovered groups evidenced a more rapid dramatic response to treatment than those in the casualty, deteriorated, and no-change groups. There were no rapid responders who, at the end of treatment, were categorized as casualty or deteriorated. Only 16% of the rapid responders made no reliable change by the end of treatment. The remaining 84% were either categorized as improved or recovered. In an analysis of the tendency to maintain gains from termination to follow-up, participants who showed a rapid dramatic response to psychotherapy maintained their treatment gains after treatment termination. The findings on early dramatic response to psychotherapy shed some light on the failure to find high correlations between number of treatment sessions and outcome; patients seem to depart from routine care close to the time they experience improvement rather than in relation to planned endings.

At present, the active mechanisms linking early dramatic response to long-term outcomes are unknown. They might include a "flight into health" or other unrealistic reflections of resistance, medications beginning to work, significant reductions in psychosocial stressors, or simply a very rapid response to psychotherapeutic interventions coming from insight or important realizations. Whatever the active ingredients are, they appear to work quickly in many cases. The timing of improvements during psychotherapy has theoretical implications beyond placebo explanations for change. If response to therapy precedes introduction of theoretically important techniques, then it is difficult to attribute central importance to these techniques in the healing

process. Early dramatic responders to psychotherapy may be more resilient, better prepared, more motivated, and thus more receptive to therapeutic influences of any kind. Early responders may be more likely to proceed from one problem to the next in therapy (a sequential pattern), whereas delayed responders continue to revisit the same therapy topic across sessions (Fennel & Teasdale, 1987). Early response may also indicate a better fit between client and therapist and reflect the positive effects of the working alliance, which often can be detected by Session 3 of treatment (Krupnick et al., 1996).

Whatever the mechanism of change, the emerging findings suggest that early dramatic treatment response is relatively common (perhaps 25% of patients) and foretells recovery and maintenance of gains. It suggests that these patients, were their progress monitored, could be discharged from treatment after just a few treatment sessions. We could then focus on maximizing the outcome of patients whose early positive response cannot be detected and, even more important, those patients who outright worsen during treatment.

Accuracy of Predicting Treatment Failure

If treatment failure is to be minimized, it is necessary to predict poor outcomes before they occur. We have conducted six studies aimed at examining the accuracy of our predictive methods (Ellsworth, Lambert, & Johnson, 2006; Hannan et al., 2005; Lambert, Whipple, Bishop, et al., 2002; Lutz et al., 2006; Percevic, Lambert, & Kordy, 2006; Spielmans, Masters, & Lambert, 2006). These studies vary with regard to patient populations, methods, and findings. The general finding is that deterioration can be predicted before it occurs through the use of information about patients' initial level of disturbance and their treatment response after one or more sessions of psychotherapy (empirical algorithms). Once these variables are taken into account, little improvement in predictive accuracy is achieved by adding such variables as diagnosis (i.e., what matters most is how disturbed a patient is, not which disorder they have or if they have comorbid conditions, because both these variables correlate with degree of disturbance), sex, age, ethnicity, type of treatment, and experience of therapist.

The hit rate (overall correct classification rate) for OQ-45 feedback appears to be about .80. The sensitivity of the predictive system is defined as the proportion of people who ultimately deteriorate in treatment who were predicted to deteriorate by the feedback system. Sensitivity in our studies typically is about .88. This means that if 100 individuals deteriorated in treatment, the OQ-45 feedback system would initiate an alarm signal for 88. Predictive accuracy is always a function of base rates of the phenomenon of interest as well as the predictive system used; more rare events such as deterioration and suicide are difficult to detect. Because deterioration rates for psychotherapy usually hover at about 5% to 10%, hit rates for predicting failure must exceed the baseline judgment that deterioration will not occur (e.g., always predicting no deterioration will lead to accuracy 90% to 95% of the time).

In this context, the algorithms we have developed show that if a patient never receives red or yellow signals it is almost certain that he or she will not deteriorate. The negative predictive power (the proportion of people who are predicted not to deteriorate in treatment who, in fact, do not deteriorate) is typically about .98. The positive predictive power values are much lower, somewhere about .20. This ratio can be interpreted as indicating that there is about a 20% increase (over baseline rates of 8%) in ability to predict treatment failure by using the feedback system. These results are consistent with validity coefficients that can be expected when base rates are very low (Steiner, 2003).

The accuracy of empirical systems for predicting whether a patient will deteriorate has been contrasted with the accuracy of therapist judgment. Hannan et al. (2005) examined therapist accuracy by asking 40 therapists (20 trainees and 20 experienced professions), at the end of each session, whether they believed a patient would leave treatment in a deteriorated state and whether a patient was worse off at this particular session compared with functioning at therapy entry. Therapists were aware of the purpose of the study, the dependent measure (OQ-45), and cut-offs for judging deterioration and that the base rate for deterioration was likely to be 8%. During the 3-week period, predictions were made for 550 clients and compared with our empirical prediction method. Only three of 550 clients were predicted to deteriorate, one of whom actually deteriorated. The therapists did not identify 39 additional clients who deteriorated during treatment. Despite being armed with base-rate information and having familiarity with the outcome measure used in the study, therapists showed an inability to accurately forecast negative outcome. Their predictions of deterioration would have improved markedly if they had simply used their judgment that a patient was worse off at any particular session. Although correctly identifying 16 clients who had worsened during treatment, they did not interpret this information as a predictor of final patient outcome. In contrast, the statistically based alarm system, applied to patient scores during the same 3-week time

period, identified 77% of the deteriorators but generated numerous false-positive results.

Unfortunately, when we designed the foregoing study, practical constraints dictated that we use algorithms based on the OQ-45 to predict deterioration based on the OQ-45 (as the standard for predictive accuracy). It would be desirable to use a battery of outcome measures coming from multiple measures and sources as the predicted final outcome. This limitation suggests the need for additional research on predictive accuracy of our statistically based methods. Undoubtedly, multiple measures and measurement sources would result in varied estimates of negative outcome and varied estimates of predictive accuracy. At the same time, Beckstead et al. (2003) found considerable consensus between different self-report measures in classifying individuals' clinically significant change based on diverse self-report measures, including the Symptom Checklist-90, Inventory of Interpersonal Problems, Social Adjustment Rating Scale, and Quality of Life Inventory. Whatever differences in predictive accuracy may exist as a result of different operational definitions of patient deterioration (based on patient self-report), we doubt that therapist ratings are the most suitable means of judging this phenomenon. The reader may recall the early work of Yalom and Leiberman (1971), in which group leaders proved to be very poor at judging which group members were casualties, and group members were much more aware of the negative impacts of the group and leaders on some patients.

Methods of Preventing Treatment Failure

We have applied the foregoing statistically based alarm system as an intervention for preventing deterioration and enhancing positive outcomes in patients. In addition, we have created, and provided therapists with, problem-solving guidelines (clinical support tools [CSTs]; Lambert, Whipple, Harmon, et al., 2004) for helping such patients and provided direct feedback about progress to patients themselves. In our current research, we have incorporated administration of the OQ-45 (by means of hand-held computers), scoring, and application of algorithms into software called OQ-Analyst. A screen shot of the software output is presented in Figure 1, which illustrates the progress of a fictional patient, Brad News. The OQ-45 is typically administered before each session, making the Session 1 score a reflection of the patient's status before Session 1 and the Session 8 score reflecting his status before Session 8. Brad was predicted to have a negative treatment outcome after undergoing a single session of treatment (at Session 2). He again passed the cut-off score for a red warning after seven sessions of treatment. Messages consistent with his progress at any session of interest are provided by OQ-Analyst. At Session 8, another red alarm was sent. The therapist could look below the graph and read the message that is provided. Messages vary depending on the size of the deviation from expected treatment response (the dark sloping line) and amount of therapy. Patient scores are displayed in the graph in relation to the horizontal line at a score of 63, which demarcates normal functioning. As can be seen, the feedback report provides information about Brad's answer to five critical items, as they were answered just before Session 8.

Five controlled studies have been published that examine the effects of providing progress feedback to patients (Harmon et al., 2006; Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Lambert et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002; Whipple et al., 2003). Each study required about 1 year of daily data collection and evaluated the effects of providing feedback about a patient's improvement through the use of progress graphs and warnings about patients who were not demonstrating expected treatment responses (alarm-signal cases).

Our primary question was, Does formal feedback to therapists (or patients) on patient progress improve psychotherapy outcomes? We hypothesized that patients identified as alarm-signal cases (those predicted to have a poor final treatment response) whose therapist received feedback will show better outcomes than similar patients whose therapist did not receive feedback (See also, Lambert, Whipple, et al. 2003).

The studies shared many things in common: (a) Each included consecutive cases seen in routine care regardless of patient diagnosis or comorbid conditions (rather than being disorder specific); (b) random assignment of patient to experimental (feedback) and treatment-as-usual (TAU) conditions (no feedback) was made in all but one of the studies; (c) psychotherapists provided a variety of theoretically guided treatments, with more clinicians adhering to cognitive–behavioral and eclectic orientations followed by psychodynamic and experiential orientations; (d) a variety of clinicians were involved (postgraduate psychotherapists and graduate students each represented about 50% of the therapists); (e) therapists saw both experimental (feedback) and TAU (no feedback) cases, thus limiting the likelihood that outcome differences between conditions could be due to therapist effects; (f) the outcome measure as well as the methodology rules or standards for identifying alarm-signal patients (failing cases) remained constant; (g) the length of therapy (dosage) was determined by patient and therapist rather than by research design or arbitrary insurance

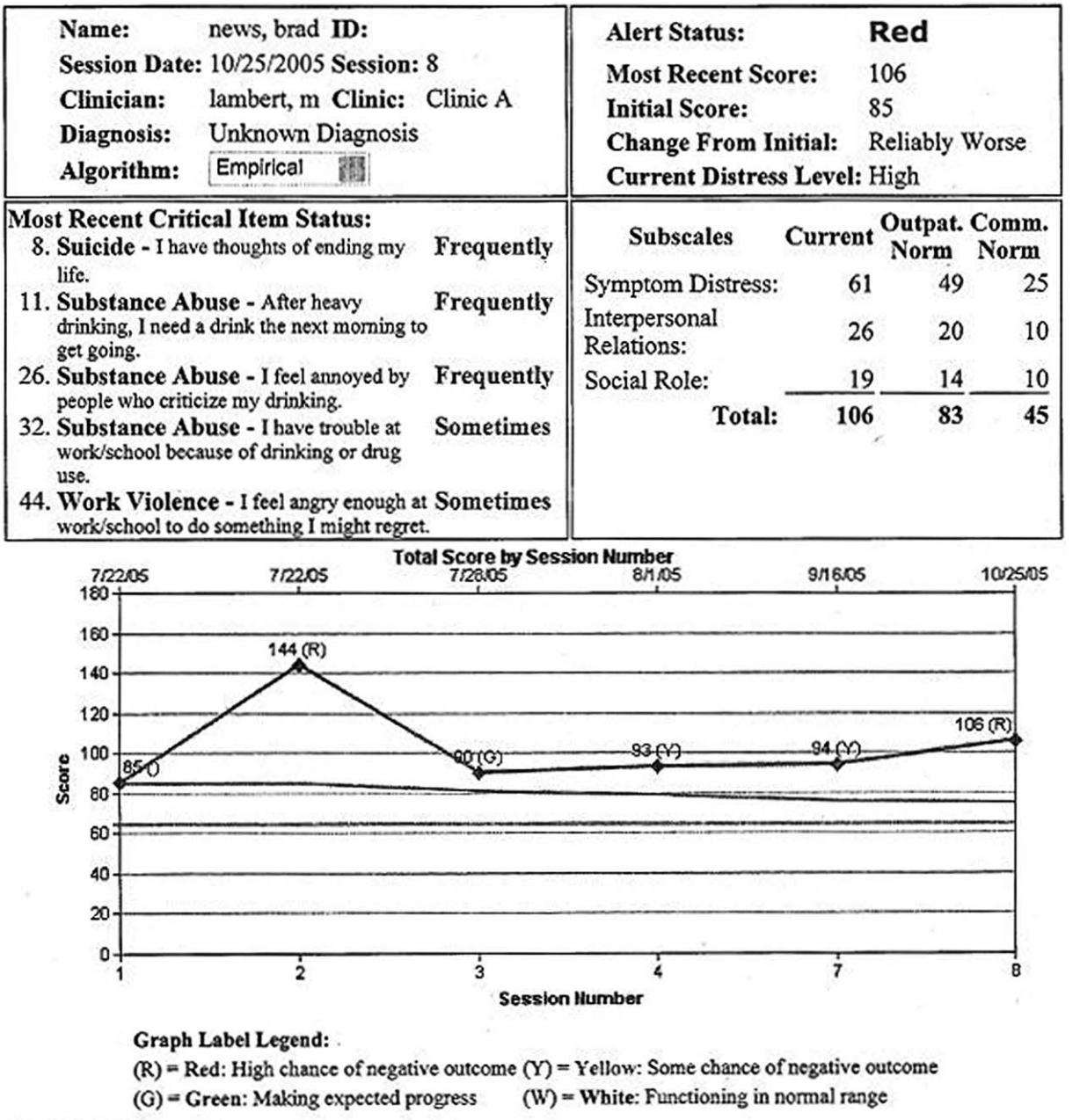

Figure 1. OQ-Analyst screen shot illustrating feedback graph and report of patient progress.

limits; and (h) patient characteristics such as gender, age, and ethnicity were generally similar across four of the studies and came from the same clinic, whereas the fifth sample (Hawkins et al., 2004) was older, more disturbed, and treated in a hospital-based outpatient clinic.

Another notable difference in the studies was that two of the studies (Harmon et al., 2006; Whipple

et al., 2003) extended the design of the first two studies by including a second experimental condition. This was intended to strengthen the feedback intervention by encouraging therapists to use Clinical Support Tools (CSTs) (a problem-solving decision tree and additional measures) with alarm-signal cases. The Harmon et al. (2006) and Hawkins et al. (2004) studies also included two experimental conditions aimed at comparing TAU with feedback to therapists and feedback to therapists and patients.

Results from the combined studies are presented graphically in Figure 2. As can be seen, the patients identified as NOT (alarm-signal patients) had a different outcome course depending on assignment to the no-feedback or feedback treatment conditions. Up to the point at which these alarm-signal cases were first signaled or, in the case of the no-feedback treatment-as-usual condition, could have been signaled, the graph illustrates an average worsening of about 10 points (about 0.5 *SD* on the OQ-45). From the point of the alarm signal, all the experimental (feedback [Fb]) groups (NOT-Fb, NOT-Fb + CST, and NOT-Th & PT) improved, whereas the no-feedback control (NOT-TAU) cases improved to an average score near 80 and were, as a group, slightly worse off than when they entered treatment.

Also displayed is the outcome for on-track (OT) cases that did (OT-FB) and did not (OT-TAU) receive feedback. As can be seen, these patients made steady progress and left treatment, as a group, well within the ranks of normal functioning. It appeared to make little difference in outcome for feedback (green or white messages) to have been given.

In the individual studies themselves, the effect sizes (standardized mean differences) for the difference between various feedback conditions for the NOT patients and NOT-TAU controls ranged from .34 to .92. Such effect sizes are surprisingly large when one considers that an average effect for comparative studies (active treatments) typically falls between .00 and .20 (Lambert & Ogles, 2004) and is widely considered important enough to lead to a recommendation of a best practice. Across the five studies, some inconsistent results have been found. Usually the provision of NOT feedback increases the number of sessions that patients attend by about two to three sessions and decreases sessions for OT cases by 0.66 session, but not in Hawkins et al. (2004). In about half the studies, feedback to OT cases improved outcomes despite reducing treatment length. Direct feedback to patients in the form of a written message improved outcomes dramatically in Hawkins et al. (2004) but had no impact in the Harmon et al. (2006) replication. A second replication study is in the final stage of data analysis.

Table I presents a classification of alarm-signal patients based on their final treatment status at termination. As shown, 20% of the alarm-signal cases seen by therapists who received no feedback (NOT-TAU) showed a negative treatment outcome at termination. In contrast, when therapists received feedback that identified their patient as NOT, only 15%, 12%, and 8% of T-Fb, T/P-Fb patients and T-Fb+CS, respectively, deteriorated. The rates for alarm-signal cases showing clinically significant or

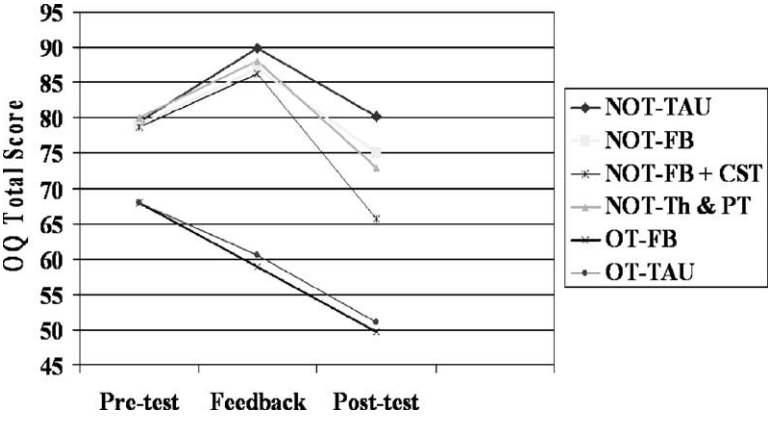

Figure 2. Change from pre- to posttesting of not-on-track (alarm-signal) and on-track patients. (NOT-NFb = alarm-signal patients whose therapist received no signal or message; NOT-FB = alarm-signal patients whose therapist received a red or yellow signal indicating they were at risk for treatment failure; OT-TAU = patients who were making satisfactory progress and whose therapist never received any information about their progress; OT-FB = patients whose therapist received a green or white signal and message and who were predicted to have a positive outcome; NOT-FB + CST = alarm-signal patients whose therapist received feedback and used the clinical support tools; NOT-Th & PT all patients who entered treatment and were assigned to receive feedback and whose therapist also received feedback; Pre-test = average patient scores on the Outcome Questionnaire-45 at intake; Feedback = average patient score on the Outcome Questionnaire-45 at the point at which a patient qualified for a yellow or red message (the time of warning varied across patients); Post-test = average patient Outcome Questionnaire-45 score at the session at which treatment was terminated (number of sessions until termination occurred varied).

Table I. Percent of Not-On-Track (Signal-Alarm) Cases Meeting Criteria for Clinically Significant Change at Termination Summed Across Five Studies.

	NOT-TAU[a]	NOT-T-FB[b]	NOT-FB+CST[c]	NOT-Th & PT
Outcome Classification	n (%)	n (%)	n (%)	n (%)
Deteriorated[e]	64 (20%)	90 (15%)	12 (8%)	19 (13%)
No Change	184 (58%)	316 (53%)	73 (47%)	71 (48%)
Reliable/or Clinically Significant Change[f]	70 (22%)	196 (33%)	169 (45%)	57 (39%)

[a]NOT-TAU = patients who were not on track and whose therapist was not given feedback.
[b] NOT-FB = patients who were not on track and whose therapist received feedback.
[c] NOT-T-FB+CST = patients who were not on track and whose therapist received feedback and used clinical support tools.
[d] T/P-Fb = therapist feedback plus written direct feedback to patients.
[e] Worsened by at least 14 points on the OQ from pre-treatment to post-treatment.
[f] Improved by at least 14 points on the OQ or improved and passed the cut-off between dysfunctional and functional populations.

reliable change were also markedly different; the highest rates of improvement were in the T-Fb+ CST condition (45%) compared with 22% in the NOT-TAU condition. These data suggest that the improved outcomes for patients in the experimental conditions are not only statistically significant but also of considerable clinical meaning for the individual client.

Improving Outcomes Through Examination of Therapists' Effects

Because of the rather large database that has accumulated over the years, and with the permission of therapists, we have been able to evaluate the outcome of patients as it relates to specific therapists in two overlapping studies (Okiishi et al., 2003, 2006). After testing for the effects of therapeutic orientation, therapist experience, gender, and age, and finding that these variables were unrelated to patient outcome, we sought to compare outcome across 71 therapists who treated clients with equivalent levels of initial severity. Each therapist saw a minimum of 30 patients (range = 30–350+). As might be expected, there was considerable variability in patient outcome as a function of which therapist treated a patient. Patient outcome was normally distributed across therapists. In general, therapists in the second and third quartiles could not be distinguished from one another based on their patients' outcome, but comparisons between the top and bottom quartiles revealed statistically significant differences.

To give an example of the differences in outcomes for extreme groups, the upper 10% ($n = 7$) and lower 10% ($n = 7$) of therapists are contrasted here. The top-rated therapists' clients had an improved or recovery rate of 44% and a deterioration rate of 5%, whereas the clients seen by the bottom-rated therapists improved or recovered at a rate of 28%, with 11% deteriorating. One therapist who saw more than 160 patients had a 19% deterioration rate, whereas another saw more than 300 patients, with less than 1% deteriorating. For the purposes of quality improvement, such dramatic differences are well worth exploring with regard to how these individuals practice, an ongoing investigation. Such differences also provide an immediate opportunity for a clinic to enhance patient treatment response by decreasing referrals to therapists with the poorest outcomes.

Because therapists were given anonymity in return for participation in the data analysis, this information could not be directly and immediately used to enhance patient outcome, but further understanding and use of this information for scientific and humanitarian purposes are highly desirable. One is reminded of the progress made in treating cystic fibrosis as described by Gawande (2004) in an article for the *New Yorker*: "The Bell Curve: What Happens When Patients Find Out How Good Their Doctors Really Are?" In the case of cystic fibrosis, life expectancy was dramatically increased, presumably as a function of examining and sharing (making public) outcome data showing the variability in time to death across specialty clinics and subsequent examination of practices engaged in by the most successful clinics. Certainly examining psychotherapy outcome as a function of provider holds considerable promise for patients, but little has been done to facilitate such research or use the research to enhance patient outcome.

Limitations and Implementation Issues

Given the large sample sizes in our clinical trials and a combined overall sample size of more than 4,000 cases, the current findings on feedback seem compelling. However, several limitations to our research program should be noted. First, most of the results summarized on feedback were based on data collected in a university outpatient clinic. Studies in other settings are needed. In addition to the Hawkins et al. (2004) study, which was conducted in a hospital-based outpatient clinic, a 2006 study of

30-day inpatient treatment in Switzerland (Berking, Orth, & Lutz, 2006) using similar methods has replicated the effects of progress feedback, but many more studies are needed to determine the limitations and generalizability of such interventions.

Second, we made no attempts to dictate how feedback was used by therapists (e.g., therapists were free to share the information with patients or not, depending on therapist preferences). Although this methodology increases the likelihood that the results reflect what will happen in other clinical settings, clinicians' actions with regard to reviewing feedback, sharing it with clients, and or modifying treatment remain largely unknown (with the exception of our two studies in which we delivered feedback directly to patients as well as therapists).

Third, our research is limited to a single self-report measure of improvement and, therefore, provides only one view of the impact of therapy on patients. We are well aware of the fact that decisions regarding the continued provision of treatment, modification of ongoing treatment, and the like cannot be made on the basis of a single questionnaire or independent from clinical judgment. We envision our feedback system as supplementing clinical decision making, a lab test result, rather than a replacement for the clinician's judgment.

A few comments on the practicality of implementing a feedback system in routine care are in order. Generally, clinicians do not see the value of frequent assessments based on standardized scales (Hatfield & Ogles, 2004), possibly because they are confident in their ability to accurately observe patient worsening and provide an appropriate response. Despite evidence suggesting that psychotherapists are not alert to treatment failure (Hannan et al., 2005; Yalom & Lieberman, 1971) and strong evidence that clinical judgments are usually found to be inferior to actuarial methods across a wide variety of predictive tasks (Meehl, 1954), therapist confidence in their own clinical judgment stands as a barrier to implementation of monitoring and feedback systems. In addition, clinicians are used to practicing in private and with considerable autonomy. Monitoring patient treatment response makes the effects of practice somewhat public and transparent. Such transparency inevitably raises evaluation anxiety and fears of losing control. Implementation requires the cooperation of therapists and takes time before it is apparent to clinicians that the feedback is helpful.

The practical difficulties of adding monitoring activities to busy practices can be an important barrier to implementation. Fortunately, recent developments in software programs make the possibility of instantaneous feedback to clinicians easy to implement. If the patient takes the OQ-45 immediately before the scheduled psychotherapy session, feedback through the OQ-Analyst is available to the therapist before beginning that session.

We continue to work on implementation issues in combination with further research studies. With regard to implementation, we are incorporating our problem-solving decision tree and CSTs into the OQ-Analyst software. This will enable alarm-signal patients to readily take assessments that facilitate clinician problem solving and will provide interventions strategies for therapist consideration more rapidly. In addition to making implementation automatic, we continue to modify our CSTs and test their effects to facilitate problem solving with the failing case.

Whereas most practitioners see themselves as well above average, a phenomenon not restricted to mental health professionals and psychotherapy, patient outcome is normally distributed across therapists, with a small but important minority producing dramatically positive and negative outcomes. As noted earlier, we are recording therapy sessions of therapists whose patients have unusually positive outcomes and hope that studying these therapists will reveal highly salient aspects of the healing process and thereby enhance training and practice.

We have expanded our attempts to measure outcomes by developing a shortened version of the OQ-45 (the OQ-30) and a version to be used with the severely and persistently mentally disabled. In addition, we have developed measures for children 17 years of age and younger, the Youth Outcome Questionnaire and Youth Outcome Questionnaire-30, and tested algorithms for identifying off-track cases.

We hope that the results of our feedback studies convinces practitioners that systematically monitoring their patients with the methods described here is in the best interest of patients and that researchers will consider replication, improvement, and expansion of our methods and findings.

Acknowledgements

The following list of individuals made significant contributions to various aspects of the research reported in this summary. Without their help the findings reported here would never have emerged. In addition to their intellectual contributions and tireless efforts I have been lucky enough to have enjoyed their company and emotional support. They are listed in alphabetic order: Ed Anderson, Steffi Bauer, Matthew Bishop, Jeb Brown, Gary Burlingame, Taige Bybee, Kara Cattani-Thompson, Arthur Finch, Eric Haas, Corinne Hannan, Wolf-

gang Hannover, Nathan Hansen, Cory Harmon, Derek Hatfield, Eric Hawkins, Scott Kadera, Hans Kordy, Wolfgang Lutz, Kevin Masters, Stevan Nielsen, Ben Ogles, John Okiishi, Robert Percevic, Bernd Puschner, Matthias Richard, Karstin Slade, David Smart, Glen Spielmans, Richard Tingey, Val Umphress, David Vermeersch, Jason Whipple, Bob Wong.

I would also like to thank Brigham Young University, the College of Social Home and Family Life, and the Counseling and Career Center for their generous financial contributions in support of this program of research. An atmosphere has been created in which success is encouraged. I would also like to thank the therapists and clients at the counseling center, and in many other settings, who joined with us in exploring new methods for enhancing patient outcome.

References

Agras, W. S., Crow, S. J., Halami, K. A., Mitchell, J. E., Wilson, G. T., & Kraemer, H. C. (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *American Journal of Psychiatry*, *157*, 1302–1308.

Anderson, E. M., & Lambert, M. J. (2001). A survival analysis of clinically significant change in outpatient psychotherapy. *Journal of Clinical Psychology*, *57*, 875–888.

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment*, *82*, 60–70.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goates, M. K., & Vermeersch, D. A. (2003). Clinical significance of the Outcome Questionnaire (OQ-45.2). *The Behavior Analyst Today*, *4*, 79–90.

Berking, M., Orth, U., & Lutz, W. (2006). Wie effekiv sind systematische Ruckmeldungen des therapieverlaufs an den therapeuten? Eine empirishe studie in einem stationar-verhaltenstherapeutischen setting. [How affective is systematic feedback of treatment progress to therapist? An empirical study in a cognitive behavioural inpatient setting] *Zeitschrift fur Klinishe Psychologie und Psychotherapie*, *35*, 21–29.

Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (pp. 194–226). New York: Wiley.

de Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (in press). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology and Psychotherapy*.

Derogatis, L. R. (1983). *The SCL-90: Administration, scoring, and procedures for the SCL-90*. Baltimore, MD: Clinical Psychometric Research.

Durham, C. J., McGrath, L. D., Burlingame, G. M., Schaalje, G. B., Lambert, M. J., & Davies, D. R. (2002). The effects of repeated administrations on self report and parent report scales. *Journal of Psychoeducational Assessment*, *20*, 240–257.

Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A comparison of the Outcome Questionnaire-45 and Outcome Questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology and Psychotherapy*, *13*, xxx–xxx.

Fennel, M. J. V., & Teasdale, J. D. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research*, *11*, 253–271.

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy*, *8*, 231–242.

Gawande, A. (2004). The bell curve: What happens when patients find out how good their doctors really are? *The New Yorker*, *12*, 82–91.

Haas, E., Hill, R., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*, *58*, 1157–1172.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology: In Session*, *61*, 155–163.

Hansen, N. B., & Lambert, M. J. (2003). An evaluation of the dose-response relationship in naturalistic treatment settings using survival analysis. *Mental Health Services Research*, *5*, 1–12.

Hansen, N. B., Lambert, M. J., & Forman, E. V. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, *9*, 329–343.

Harmon, S. C., Lambert, M. J., Smart, D. W., Hawkins, E. J., Nielsen, S. L., Slade, K., & Lutz, W. (2006). Enhancing outcome for potential treatment failures: Therapist/client feedback and clinical support tools. *Psychotherapy Research*.

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures use in clinical practice by psychologists. *Research & Practice*, *35*, 485–491.

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K., & Tuttle, K. (2004). The effects of providing patient progress information to therapists and patients. *Psychotherapy Research*, *14*, 308–327.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, *56*, 885–892.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist*, *41*, 159–164.

Ilardi, H. H., & Craighead, W. E. (1994). The role of non-specific factors in cognitive-behavioral therapy for depression. *Clinical Psychology: Science and Practice*, *1*, 138–156.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1996). A session-by-session analysis of the psychotherapy dose-effect relationship. *Journal of Psychotherapy Practice and Research*, *5*, 132–151.

Krupnick, J. L., Sotsky, S. M., Elkin, I., Simmens, S., Moyer, J., Watkins, J., & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome. Findings in the National Institute of Mental Health Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, *64*, 532–539.

Lambert, M. J., & Anderson, E. M. (1996). Assessment for time-limited psychotherapies. *Annual Review of Psychiatry*, *15*, 23–47.

Lambert, M. J., Bergin, A. E., & Garfield, S. L. (2004). Introduction and overview. In M. J. Lambert (Ed.), *Bergin &*

Garfield's handbook of psychotherapy and behavior change (5th ed., pp. 3–15). New York: Wiley.

Lambert, M. J., Burlingame, G. M., Umphress, V. J., Hansen, N. B., Vermeersch, D., Clouse, G., & Yanchar, S. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, *3*, 106–116.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., et al. (2004). *Administration and scoring manual for the Outcome Questionnaire–45*. Salt Lake City, UT: OQ Measures.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York: Wiley.

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically derived and rationally derived methods for identifying clients at risk for treatment failure. *Clinical Psychology and Psychotherapy*, *9*, 149–164.

Lambert, M. J., Whipple, J. L., Harmon, C., Shimokawa, K., Slade, K., & Christofferson, C. (2004). *Clinical support tools manual*. Provo, UT: Department of Psychology, Brigham Young University.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice*, *10*, 288–301.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on client progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, *11*, 49–68.

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. K. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy*, *9*, 91–103.

Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology*, *66*, 400–410.

Lutz, W., Lambert, M. J., Harmon, S. C., Stulz, N., Tschitsaz, A., & Schürch, E. (2006). The probability of treatment success, failure and duration—What can be learned from empirical data to support decision making in clinical practice? *Clinical Psychology and Psychotherapy*, *13*, 223–232.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, *70*, 248–262.

Nielsen, S. L., & Lambert, M. J. (2006). *Psychotherapeutic treatment as usual at a large university counseling center: Dose effect or Goldilocks response?* Manuscript submitted for publication.

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, S. L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their patients' psychotherapy outcome. *Journal of Clinical Psychology*, *62*, 1157–1172.

Okiishi, J. C., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). In search of supershrink: Using patient outcome to identify effective and ineffective therapists. *Clinical Psychology and Psychotherapy*, *10*, 361–373.

Orlinsky, D. E., Grawe, K., & Parks, B. K. (1994). Process and outcome in psychotherapy—Noch einmal. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy & behavior change* (4th ed., pp. 270–376). New York: Wiley

Otto, M. W., Pollack, M. H., & Maki, K. M. (2000). Empirically supported treatments for panic disorder: Costs, benefits, and stepped care. *Journal of Consulting and Clinical Psychology*, *68*, 556–563.

Percevic, R., Lambert, M. J., & Kordy, H. (2006). What is the predictive value of responses to psychotherapy for its future course? Empirical explorations and consequences for outcome monitoring. *Psychotherapy Research*, *16*, 364–373.

Renaud, J., Brent, D. A., Baugher, M., Birmaher, B., Kolko, D. J., & Bridge, J. (1998). Rapid response to psychosocial treatment for adolescent depression: A two-year follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1184–1191.

Shapiro, D. A., Barkham, M., Stiles, W. B., Hardy, G. E., Rees, A., Reynolds, S., & Stantup, M. (2003). Time is of the essence: A selective review of the fall and rise of brief therapy research. *Psychology & Psychotherapy: Theory, Research & Practice*, *76*, 211–235.

Snell, M. N., Mallinckrodt, B., Hill, R. D., & Lambert, M. J. (2001). Predicting counseling center clients' response to counseling: A 1-year follow-up. *Journal of Counseling Psychology*, *48*, 463–473.

Spielmans, G. I., Masters, K. S., & Lambert, M. J. (2006). A comparison of rational versus empirical methods in prediction of negative psychotherapy outcome. *Clinical Psychology and Psychotherapy*, *13*, 202–214.

Stone, G. L., & McMichael, J. (1996). Thinking about mental health policy in university and college counseling centers. *Journal of College Student Psychotherapy*, *10*, 3–27.

Streiner, D. L. (2003). Diagnostic testing: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, *81*, 209–219.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive–behavioral psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, *67*, 894–904.

Taylor, S., & McLean, P. (1993). Outcome profiles in the treatment of unipolar depression. *Behavioral Research Therapy*, *31*, 325–330.

Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the Outcome Questionnaire. *Journal of Psychoeducational Assessment*, *15*, 40–55.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2002). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment*, *74*, 242–261.

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome Questionnaire: Is it sensitive to changes in counseling center clients? *Journal of Counseling Psychology*, *51*, 38–49.

Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychology*, *33*, 1111–1115.

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem solving strategies in routine practice. *Journal of Counseling Psychology*, *58*, 59–68.

Wolgast, B. M., Lambert, M. J., & Puschner, B. (2003). The dose-response relationship in a college counseling center: Implications for setting session limits. *Journal of College Student Psychotherapy*, *8*, 15–29.

Yalom, I. D., & Lieberman, M. A. (1971). A study of encounter group casualties. *Archives of General Psychiatry*, *25*, 16–30.