Is It Time for Clinicians to Routinely Track Patient Outcome? A Meta-Analysis

Michael J. Lambert, Jason L. Whipple, and Eric J. Hawkins, Brigham Young University
David A. Vermeersch, Loma Linda University
Stevan L. Nielsen and David W. Smart, Brigham Young University

Empirically supported psychotherapies, treatment guidelines, best practices, and treatment manuals are methods proposed to enhance treatment outcomes in routine practice. Patient-focused research systems provide a compatible and contrasting methodology. Such systems monitor and feed back information about a patient's progress during psychotherapy for the purpose of enhancing outcomes. A meta-analytic review of three large-scale studies is summarized and suggests that formally monitoring patient progress has a significant impact on clients who show a poor initial response to treatment. Implementation of this feedback system reduced deterioration by 4% to 8% and increased positive outcomes. Our interpretation of these results suggests that it may be time for clinicians routinely and formally to monitor patient treatment response.

Key words:* quality assurance, meta-analysis, psychotherapy outcome research, clinical significance, feedback, quality management, Outcome Questionnaire. *[Clin Psychol Sci Prac 10:288–301, 2003]

The effects of psychotherapy have been found to be generally positive across a variety of studies and over substantial periods of time dating back to the 1930s (Lambert & Ogles, in press). Substantial efforts have been made to improve patient outcome. The most recent attempts have taken the form of comparative outcome studies employing clinical trial methodology and treatment manuals. Results from these studies have led to recommendations for best practices based on criteria proposed by a variety of scientific committees (e.g., Chambless & Hollon, 1998). In turn, these committees assume that clinical practice outcomes would be enhanced if psychotherapists limited their practice to the use of treatments that have substantial evidence of efficacy (Task Force, 1995). Although clinicians are unlikely to question the merit of supporting therapeutic approaches based on strong empirical foundations, current best practice recommendations that seem to favor particular treatments have engendered criticism and controversy (e.g., Garfield, 1996; Nathan, 1998). Fortunately, improving the quality of patient care and maximizing patient benefits can also be achieved by a variety of other means.

Attempts to provide quality assurance for clients receiving psychotherapy are a worldwide phenomenon and extend beyond the debate over empirically supported psychological interventions (Andrews, 2000). In a special section of the *Journal of Consulting and Clinical Psychology,* quality assurance research programs in the United States (Beutler, 2001; Lueger et al., 2001; Lambert, Hansen, & Finch, 2001), Great Britain (Barkham et al., 2001), and Germany (Kordy, Hannöver, & Richard, 2001) were described. A common feature across each of these systems is an emphasis on patient-focused research. This methodology endeavors to improve psychotherapy outcome by monitoring client progress and providing this information to clinicians in order to guide ongoing treatment, especially for the client who is not having a favorable response to treatment (signal-alarm cases). Patient-focused research, therefore, is an extension of quality assurance and represents one effort to bridge the gap between efficacy and effectiveness research and clinical practice, while enhancing patient outcome before treatment termination (Lambert, 2001). It is also well suited to models of care in which clinicians attempt to step-up or step-down treatments af-

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

ter assessing patient treatment response (Newman, 2000; Otto, Pollack, & Maki, 2000).

We report here a meta-analysis of the results of a series of studies aimed at evaluating one system of patient-focused research. Like most meta-analytic reviews, the current study was undertaken because summing data across studies allows for conclusions that provide better estimates of effects than individual studies. It also allows for analysis of treatment effects that, given small subsample sizes within the individual studies, may not have met criterion for statistical significance, but which are nonetheless reliable.

The research summarized here has been published elsewhere (Lambert et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002; Whipple et al., 2003). Each of the three studies required about one year of data collection and was based on about 1,000 clients. All three studies evaluated the effects of providing therapists with feedback about client improvement through the use of progress graphs and warnings for clients who were failing to demonstrate expected treatment responses (signal-alarm cases). The research question of interest in these studies was as follows: Does formal feedback to therapists on client progress improve psychotherapy outcomes and attendance? Two simple hypotheses were tested: (a) signal-alarm clients (those predicted to have a poor treatment response) whose therapist received feedback will show better outcome than similar clients whose therapist did not receive feedback; (b) clients of therapist receiving feedback will show better attendance (i.e., attendance representative of cost-effective psychotherapy) than similar clients of therapists not receiving feedback. Three research questions could be addressed in this meta-analysis that could not be fully addressed by the individual studies (or were addressed but provided inconsistent findings across the studies): Do trainees profit more from feedback than experienced licensed providers? Is a relatively late signal-alarm indicative of poor final response? Is the cost of providing more sessions to signal-alarm cases offset by providing fewer sessions to patients who appear to be on-track for a positive outcome?

The three studies that are combined here shared many things in common:

1. Each was conducted in the same clinic, a college counseling center.

2. Each included consecutive cases regardless of diagnosis rather than being disorder specific.

3. Random assignment of clients to experimental and control conditions was made in two of the studies (1 and 3) while study 2 assigned treatment by school semester.

4. The professional staff that provided the treatment remained relatively constant across the three studies and provided a variety of theoretically guided interventions dominated by cognitive behavioral and eclectic orientations. Professional therapists represented about 50% of the clinicians participating in each study. Graduate student trainees (practicum students and interns) represented the remainder of participating therapists and varied from study to study.

5. In each study therapists saw both experimental and control cases, thus limiting the likelihood that differences between conditions could be due to therapist effects.

6. The measure of outcome as well as rules/standards for identifying signal-alarm patients (failing cases) remained constant.

7. The length of therapy (dosage) was determined by patient and therapist rather than by research design or arbitrary insurance limits.

8. Generally, patient characteristics such as gender, age, and ethnic identification were similar across studies.

Characteristics of the three samples are presented in Table 1. As can be seen, the three samples were essentially equivalent with the exception of the smaller sample size in Study 1 and the larger treatment dosage in Studies 2 and 3 compared to Study 1. Unlike most meta-analytic reviews, the studies summed here were highly homogeneous.

One notable difference in the studies was that while the second study was a replication of the first, the third study (Whipple et al., 2003) extended the design of the first two studies by including a third experimental condition that was intended to strengthen the feedback intervention by encouraging therapists to use clinical support tools (e.g., a problem-solving decision tree and additional measures) with their signal-alarm cases. However, for the purposes of this review, the third experimental condition was collapsed into the feedback condition for the sake of parsimony and because the number of clients was relatively small.

In order to understand the meaning of this meta-analysis it is essential to discuss particular details of the methodology used in the studies. First, outcome is defined and its operational definition is provided, along with rules for categorizing each client's treatment response. Next, the method for identifying likely treatment failures (signal-

Table 1. Characteristics of clients from studies used in meta-analysis

Study	Clients/therapists N	Age M	Females %	Caucasians %	Dosage M/SD	Mood Disorder %	Intake OQ M/SD
Lambert, Whipple, et al. (2001)	609/31	22.23	70%	88%	3.66/3.87	27%	69.21/22.46
Lambert, Whipple, et al. (2002)	1020/49	22.26	70%	85%	4.53/3.34	31%	71.72/22.52
Whipple et al. (2003)	981/48	22.31	66%	86%	5.46/4.87	29.2%	70.87/22.86

alarm cases) is highlighted in order to describe the feedback that was given to therapists. Finally, the effects of feedback are summarized, and the implications of these results for research and practice are provided. We conclude with limitations of the research methodology and include an argument for routine utilization of formal feedback in clinical practice.

Defining Outcome

Patient-focused research, as applied in routine practice, demands efficient outcome assessment rather than the more ideal alternative of comprehensive assessment. Outcome measurements that are typically employed in efficacy studies often require hours of assessment from multiple research perspectives of change with a small number of homogeneous patients. In contrast, patient-focused research uses weekly assessments with a single, brief measure (Howard, Moras, Brill, Martinovitch, & Lutz, 1996). Thus, assessment in this type of research is much more frequent, with a greater diversity of patients and large final sample sizes, but it is less comprehensive and lacks multiple perspectives of change.

In each of the studies examined here, psychological dysfunction was assessed using the Outcome Questionnaire-45 (OQ-45; Lambert, Hansen, et al., 1996), which provided both a measure of weekly change on which the feedback to therapists was based, as well as the criterion measure for classification of clients into outcome groups (improvers, no-changers, and deteriorators). This self-report scale was designed to measure client progress in therapy by repeated administration during the course of treatment and at termination. 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); and, quality of life (positive aspects of life satisfaction). The OQ-45 provides a total score, based on all 45 items, as well as three subscale scores. Only the OQ-45 total score, which

provides a global assessment of patient functioning, was used in the current analysis (factor analysis of the OQ-45 is consistent with a single global level of functioning solution; Mueller, Lambert, & Burlingame, 1998).

The OQ-45 has been reported to have adequate reliability and validity across a number of settings and patient populations (both clinical and normative). Research has indicated that the OQ-45 is a psychometrically sound instrument, with adequate three-week test-retest reliability ($r = .84$) (Lambert, Burlingame, et al., 1996) and excellent internal consistency (Cronbach's ALPHA = .93) (Lambert, Hansen, et al., 1996). The OQ-45 has also been demonstrated to have strong concurrent validity coefficients. These range from .55 to .88 (all significant at $p < .01$) on the SCL-90R, BDI, Zung Depression Scale, Taylor Manifest Anxiety Scale, STAI, Inventory of Interpersonal Problems, and the Social Adjustment Scale. Furthermore, the OQ-45 has been shown to be sensitive to change in clients over short time periods while remaining stable in untreated individuals (Vermeersch, Lambert, & Burlingame, 2000). Repeated testing of persons has been found to have little systematic effect on OQ-45 test scores (Durham et al., 2002), but like other self-report measures it is subject to demand characteristics that cannot always be eliminated. Instructions call for honest reporting and clients are undergoing treatment for their own personal benefit, factors that possibly minimize conscious distortion of reported symptoms.

In short, the OQ-45 is a brief measure of psychological disturbance that is reliable, valid, and sensitive to changes patients make during psychotherapy. It is well suited for tracking treatment response during and following treatment.

Defining a Positive and Negative Outcome

An important and essential aspect of patient-focused research methodologies is the establishment of cutoff scores for indicating when the changes a patient has made are

recognizable and substantial (i.e., reliable and clinically significant). This involves evaluating when a score on an outcome measure suggests a patient has made progress sufficient to be classified as reliably improved or recovered.

Using formulas developed by Jacobson and Truax (1991), clinical and normative data for the OQ-45 were analyzed by Lambert, Hansen, et al. (1996) to provide cut-off scores for the Reliable Change Index (RCI) and for clinically significant change. Clients who change in a positive or negative direction by at least 14 points are regarded as having made "reliable change." This degree of change exceeds measurement error based on the reliability of the OQ-45 and is one of two criteria posited by Jacobson and Truax as indicative of clinically meaningful change. The second criterion requires movement from a score typical of a dysfunctional population to a score typical of a functional population (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999). The cutoff on the OQ-45 for demarking the point at which a person's score is more likely to come from the dysfunctional population than a functional population has been estimated to be 64. When a client's score falls at 63 or below, the conclusion is drawn that their functioning is more similar to nonclients than clients at that point in time. Passing this cutoff (from dysfunctional to functional) is the second criterion posited by Jacobson and Truax as an indicator of clinically significant change. Clients who show reliable change and pass the cutoff are considered recovered, while those who only show reliable changes are considered improved.

Support for the validity of the OQ-45's reliable change and clinical significance cutoff scores has been reported by Lunnen and Ogles (1998) and Beckstead et al. (2003). The validity of the Jacobson and Truax formulas for establishing change cutoff has been reported by Bauer, Lambert, and Nielsen (2002). This research suggests that patients judged to be clinically significantly improved on the OQ-45 are similarly classified based on the use of other measures such as the SCL-90R, and that other methods of calculating clinically significant and reliable change produce similar estimates of change to those of the Jacobson/Truax method.

Having a method to classify each patient's treatment response is an essential component of patient-focused research, and the Jacobson/Truax method is a valid procedure for this essential task. It has the advantage of classifying each patient's treatment response rather than basing judgments about treatment response on group statistics, a methodology that makes its use well suited to bridging the gap between research and practice (Kendall, 1999).

Prediction of Treatment Failure and Description of the Warning System

The essence of improving outcomes for poorly responding patients was the development of a signaling system that attempted to identify the failing patient before termination of services had occurred. Such a signaling system is at the core of the feedback used with therapists. It requires that the patient provide session-by-session OQ-45 data that is evaluated between sessions and judged to indicate a positive or negative sign for likely functioning at treatment termination. In patient-focused research, such a signaling system is based on the assumption that termination status can, in fact, be predicted prior to termination and that providing treatment progress information to the therapist will positively affect final outcome.

A variety of procedures have been tested to determine if treatment failures could be accurately identified. Multiple regression using a variety of predictors (such as diagnosis) suggested that the best predictors were initial level of disturbance (pretreatment OQ-45 score) and change score following separate treatment sessions (Brown & Lambert, 1998). In essence, the research determined that the best way to estimate end of treatment status was to know how disturbed the patient is prior to treatment and whether the patient's early response to treatment is positive or negative. An early positive response foretells a positive final outcome and maintenance of gains at follow-up, while an early negative response to treatment is a negative indicator.

Information about early response to treatment (dramatic response during the first three sessions; Haas, Hill, Lambert, & Morrell, 2002), the fact of a dose response relationship and its size, (Anderson & Lambert, 2001; Howard, Kopta, Krause, & Orlinsky, 1986; Kadera, Lambert, & Andrews, 1996), and the reliability of the OQ-45 were used to create algorithms for identifying patients who were predicted to leave treatment before receiving therapeutic benefit or who were thought to be at risk for having a negative treatment outcome (details provided in Lambert, Whipple, Bishop, et al., 2002; cutoff scores can be obtained from the first author). For simplicity of communication in the clinical setting, patients identified as treatment nonresponders are referred to as "signal-alarm" cases. This is a term that has precedence in other research aimed at improving the quality of patient care (Kordy, et al., 2001).

The accuracy of the algorithms has been tested and they appear to be successful at identifying patients who have negative treatment outcomes. Lambert, Whipple, Bishop, et al. (2002) examined predictive accuracy with 492 clients who were in treatment at a university counseling center. Thirty-six (7.3%) of these clients deteriorated during treatment. Twenty-nine of these deteriorators (80.6%) were identified prior to termination using the algorithms while 7 (19.4%) were missed. This level of accuracy came at the expense of misidentifying 95 (20.8%) of the clients as signal-alarm cases who did not in fact deteriorate. These rates compared favorably with identification procedures based on a purely statistical approach using Multilevel Linear Modeling (Finch, Lambert, Schaalje, 2001).

Once a patient takes the OQ-45, commences treatment, and completes a session of treatment, the decision rules are used to generate feedback. Feedback to therapists consisted of a progress graph that included all the patient's scores to that point in time and a 1/4-inch colored stick-on dot (white, green, yellow, or red) that was used to visually catch the therapist's attention and immediately convey the status of patient progress. A written message corresponding to the colored dot was also provided at each session. A brief summary of the messages follows:

White Feedback: "The client is functioning in the normal range. Consider termination."

Green Feedback: "The rate of change the client is making is in the adequate range. No change in the treatment plan is recommended."

Yellow Feedback: "The rate of change the client is making is less than adequate. Recommendations: consider altering the treatment plan by intensifying treatment, shifting intervention strategies, and monitoring progress especially carefully. This client may end up with no significant benefit from therapy."

Red Feedback: "The client is not making the expected level of progress. Chances are he/she may drop out of treatment prematurely or have a negative treatment outcome. Steps should be taken to carefully review this case and decide upon a new course of action such as referral for medication or intensification of treatment. The treatment plan should be reconsidered. Consideration should also be given to presenting this client at case conference. The client's readiness for change may need to be reassessed."

Patients completed their first pretreatment OQ-45 during the intake procedure and subsequent OQ-45s prior to each treatment session. Each time an OQ-45 was administered, an updated graph and a colored dot were given to the therapist and placed in the patient's chart. The collection of OQ-45 data was done by the secretarial staff as a routine part of receiving the patient arriving for an appointment.

META-ANALYSIS OF OUTCOME ACROSS STUDIES

Results of the three studies analyzed here were combined in order to provide the best estimate of the psychotherapy outcome effects of providing signal-alarm feedback to therapists. In this report four acronyms are used to identify the treatment conditions that were examined. Clients not progressing as expected (signal-alarm cases, with either a red or yellow code) were further categorized into two groups depending on whether their therapists were recipients of feedback. Clients whose therapists received feedback are referred to as the Not-On-Track Feedback group (NOT-Fb), whereas those clients of therapist's not receiving feedback are referred to as the Not-On-Track No Feedback group (NOT-NFb). Clients of therapists receiving only green or white coded messages (i.e., who progressed as expected) are referred to as the On-Track Feedback group (OT-Fb). Clients who were "on-track" but whose therapists were not informed are referred to as the On-Track No Feedback group (OT-NFb). The allocation of patients to groups and their final treatment status is illustrated in Figure 1. This figure includes the patients who began treatment and attended at least two sessions, but excludes about 700 clients who attended a single intake session and terminated, precluding the possibility of monitoring their progress or measuring their status following treatment. This "attrition" rate is common in routine practice given that the modal number of sessions for patients who are not participating in a research protocol is one (Garfield, 1994).

The results of combining the three studies are also presented graphically in Figure 2. As can be seen, the patients identified as NOT (signal-alarms) had a different outcome course depending on assignment to the feedback or no feedback treatment conditions. Up to the point that signal-alarm cases are first signaled (or, in the case of the no feedback condition, could have been signaled) their progress is similar. It shows an average decline in functioning from a
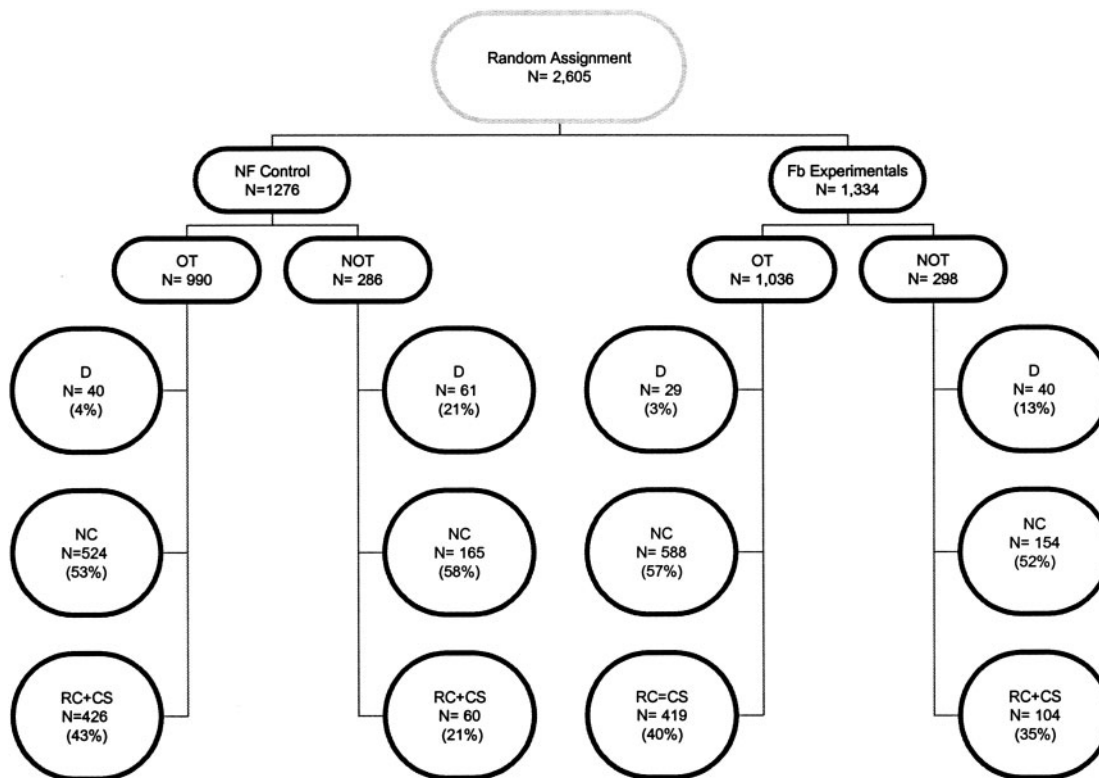
Figure 1. Diagram of design and final outcome for all patients in the meta-analysis. D = deteriorated at termination; NC= not reliably better or worse at termination; RC + CS = reliable change or clinically significant change at termination; NF Control = clients whose therapist was not informed about treatment progress; Fb Experimentals = clients whose therapist was informed about their progress.

score of 79 to a score of 89 (i.e., 10 points, about one half a standard deviation) on the OQ-45. From the point of the signal–alarm, the experimental (feedback) cases improved to a score around 72 while the control (no feedback) cases improved to an average score near 80 (i.e., they were, as a group, slightly worse off than when they entered treatment but showed improvement after being identified as a signal–alarm). The effect size for this difference was 0.39 (F $(1,581) = 26.150$, $p < .05$). In the individual studies themselves the effect sizes for this difference between NOT–NFb and NOT–Fb was 0.44, 0.34, and 0.44 in studies 1, 2 and 3, respectively (all significant at the .05 level).

Classification of Clinically Significant Change.   Table 2 presents a classification of signal–alarm patients based on their final treatment status at termination. As can be seen, 21% of the signal–alarm cases seen by therapists who received no feedback showed a negative treatment outcome at termination. In contrast, when therapists received feedback

that identified their patient as not-on-track, only 13% of the patients deteriorated. The rates for signal–alarm cases showing clinically significant or reliable change were also markedly different, with nearly a 75% increase as a result of feedback (35% versus 21% in the experimental versus control groups respectively). Classification percentages for all the patients are presented in Figure 1.

Therapy Dosage and Feedback.   Feedback was also hypothesized to affect dose of psychotherapy as well as outcome. In this context it was anticipated that those patients who were on track (i.e., who had received green or white messages without ever receiving a yellow or red warning) would receive less therapy than the control patients who were on track and whose therapists received no feedback. In contrast, patients who were not on track were expected to stay in therapy longer if feedback was given than were patients whose therapist did not receive feedback. Figure 3 summarizes evidence from the three studies on this phe-
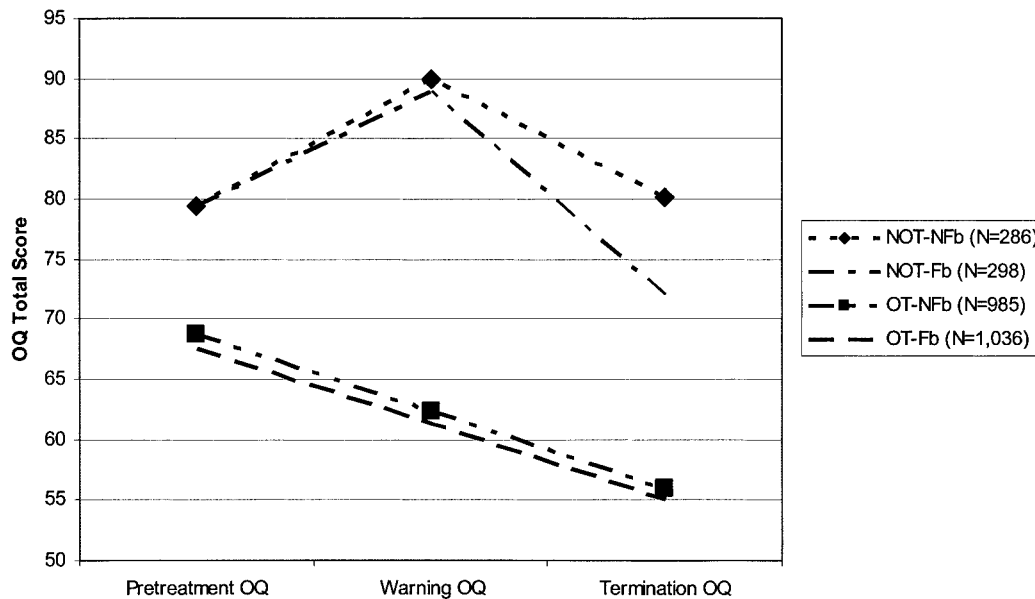
Figure 2. Change from pre- to post-testing of Not-On-Track (signal-alarm) and On-Track clients. NOT-NFb = signal-alarm cases whose therapist got no signal or message; NOT-Fb = signal-alarm cases whose therapist got a red or yellow signal, indicating they were at risk for treatment failure; OT-NFb = clients who were making satisfactory progress and whose therapist never received any information about their progress; OT-Fb = clients whose therapist got a green or white signal and message and who were predicted to have a positive outcome; Pre OQ = average client scores on the OQ-45 at intake; Warning OQ = average client score on the OQ-45 at the point at which a client qualified for a yellow or red message (the time of warning varied across clients); Post OQ = average client OQ-45 score at the session they terminated treatment (number of sessions until termination occurred varied).

Table 2. Percent of Not-On-Track (Signal-Alarm) Cases Meeting Criteria for Clinically Significant Outcome at Termination

Outcome Classification	NOT-NFb[a] n (%)	NOT-Fb[b] n (%)	χ^2
Deteriorated[c]	61 (21.3%)	40 (13.4%)	16.31*
No Change	165 (57.7%)	154 (51.7%)	
Reliable Change or Clinically Significant Changed	60 (21%)	104 (34.9%)	

[a]NOT-Fb = clients who were not on track and whose therapists were given feedback.
[b]NOT-NFb = clients who were not on track and whose therapists did not receive feedback.
[c]Worsened by at least 14 points on the OQ from pre-treatment to post-treatment.
[d]Improved by at least 14 points on the OQ or improved and passed the cutoff between dysfunctional and functional populations.
*χ^2 (2, N = 584) = 16.31, p < .001.

nomenon. As can be seen, signal–alarm cases whose therapist received feedback stayed in therapy longer (i.e., attended more sessions of therapy) than the signal–alarm patients where no feedback was given (F (1, 582) = 15.899, p < .05). On average they attend about one and one half additional sessions. This increase in attendance occurred after the warning was given.

In contrast, the OT patients of therapists who received no feedback participated in more sessions. Although only slightly greater, this session difference reached statistical significance ($F(1, 2024) = 8.136$, $p < .05$). The pattern is consistent with cost-effective psychotherapy since the decrease in attendance for the large number of OT-Fb cases offsets the additional costs of increasing dosage for the NOT–Fb feedback cases. When the average number of treatment sessions (4.69) provided to patients in the feedback condition ($n = 1334$) is contrasted with the average number of sessions (4.67) provided to patients in the No Feedback condition ($n = 1276$), there was no increase in per session costs, yet the feedback group had superior outcomes. Given the effect size difference between the treatments (i.e., decrease in deterioration and increase in reliable and clinically significant change) and no increased costs per case in the experimental group, the feedback intervention would seemingly have considerable value in providing more cost-effective treatment.

Did Outcome Vary as a Function of Therapist Experience?
Trainees were significantly more likely to have clients that became signal-alarm cases during treatment (25.5% of
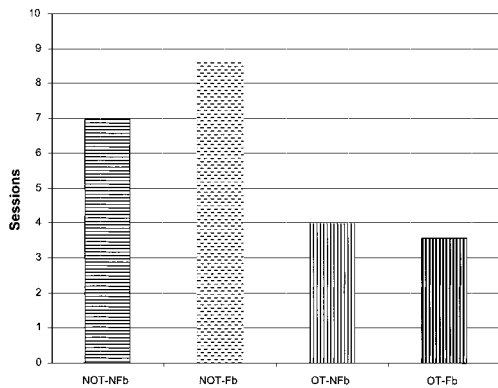
Figure 3. Mean number of sessions for experimental (feedback) and control (no feedback) groups. NOT-NFb = signal-alarm cases whose therapist got no signal or message; NOT-Fb = signal-alarm cases whose therapist got a red or yellow signal, indicating they were at risk for treatment failure; OT-NFb = clients who were making satisfactory progress and whose therapist never received any information about their progress; OT-Fb = clients whose therapist got a green or white signal and message and who were predicted to have a positive outcome.

trainees' clients versus 20.6% of staff clients; chi square (1, $n = 2610$) = 8.299, $p < .05$), a finding that supports requirements for licensure and supervision. Given their relative inexperience at helping people, it was expected that the clients ($n = 954$) of trainees ($n = 47$) would benefit more from feedback than the clients ($n = 1656$) of experienced clinicians ($n = 29$). Data from the three studies allowed for an analysis of three interaction effects between therapist experience level, feedback condition, and identification as OT/NOT. Results indicated there were no significant interactions for staff/trainee $\times$ feedback/no feedback ($F(8, 2602) = 0.853$, $p > .05$), staff/trainee $\times$ OT/NOT ($F(8, 2602) = 0.327$, $p > .05$), or staff/trainee $\times$ feedback/no feedback X OT/NOT ($F(8, 2602) = 1.778$, $p > .05$) and suggest that trainees do not profit more than experienced therapists from feedback. There was a significant main effect in this analysis that was unexpected. Overall the clients ($n = 954$) of trainees had significantly better outcome than the clients ($n = 1656$) of the professional staff. The mean change for trainee clients was -12.37 ($SD = 18.83$) while that for staff was -9.57 ($SD = 18.92$), ($F(8, 2602) = 12.509$, $p < .05$, $d = .15$). This appears to be a small but reliable difference.

These same data were also analyzed with the chi-square statistic after each patient's change was categorized in a 3 (deteriorated/no reliable change/reliable or clinically significant change) $\times$ 2 (feedback/no feedback) $\times$ 2 (staff/

trainee) frequency table to test for outcome differences between professional staff and trainees on the basis of outcome classification. Chi-squared analyses indicated patients of trainees had significantly better outcome than patients of professional staff across the entire sample (χ^2 (2, $n = 2610$) = 14.909, $p < .05$), in the feedback (χ^2 (2, $n = 2610$) = 10.422, $p < .05$) and no feedback conditions (χ^2 (2, $n = 2610$) = 9.164, $p < .05$), and both On–Track conditions (OT-Fb, χ^2 (2, $n = 2610$) = 16.122, $p < .05$) (OT-NFb, χ^2 (2, $n = 2610$) = 11.05, $p < .05$).

These data were further analyzed *within* the professional staff and trainee samples. Patients of staff in the NOT-Fb condition had significantly better outcome than patients of staff in the NOT-NFb condition (χ^2 (2, $n = 2610$) = 13.125, $p < .05$). For trainees, patients had significantly better outcome in the feedback condition than the no feedback condition (χ^2 (2, $n = 2610$) = 11.847, $p < .05$), the OT-Fb versus the OT-NFb conditions (χ^2 (2, $n = 2610$) = 8.422, $p < .05$), and the NOT-Fb versus the NOT-NFb conditions (χ^2 (2, $n = 2610$) = 5.86, $p = .05$). It appears from the foregoing analysis that the benefits of feedback to professional staff were limited to signal–alarm cases (NOT-Fb) but more broadly helpful to trainees whose clients benefited by feedback whether or not they were signal–alarm cases. In this regard they benefited more from feedback on client progress than did clients of professionals. Proportions of clients classified by individual change are presented in Table 3.

Early Versus Late Warning as an Indicator of Poor Final Outcome. Pearson's *r* was used to test the relationship between the session at which therapists received a signal-warning and final treatment outcome. The correlation between the session of first warning and outcome was $r = .073$, $p = .076$. Therefore, when the three studies were combined, the data did not support the hypothesis that later warnings would be more predictive of poor outcome. It was concluded that receiving a signal alarm at any point in the course of treatment is a poor prognostic indication.

DISCUSSION

The current meta–analysis summed three studies that tested the effects of providing therapists with feedback on patient progress with an emphasis on improving outcomes for clients predicted to have a poor treatment outcome. The control groups from these studies received treatment as usual and therefore provided a baseline for testing the

Table 3. Patient Outcome based on level of training, experimental condition, and categorization of patient change

Experimental Condition	Signal-alarm Status	Deteriorated n (%)	No-change n (%)	CS/Reliable Change n (%)
Feedback				
Professional				
	On Track	24 (3.7%)	390 (60.1%)	235 (36.2%)
	Not On Track	27 (16%)	80 (47.3%)	62 (36.7%)
	Total	51 (6.2%)	470 (57.5%)	297 (36.3%)
Trainee				
	On Track	5 (1.3%)	198 (51.2%)	184 (47.5%)
	Not On Track	13 (10.1%)	74 (57.4%)	42 (32.6%)
	Total	18 (3.5%)	272 (52.7%)	226 (43.8%)
No-Feedback				
Professional				
	On Track	25 (3.8%)	377 (56.6%)	264 (39.6%)
	Not On Track	38 (22.1%)	101 (58.7%)	33 (19.2%)
	Total	63 (7.5%)	478 (57%)	297 (35.4%)
Trainee				
	On Track	15 (4.6%)	147 (45.4%)	162 (50%)
	Not On Track	23 (20.2%)	64 (56.1%)	27 (23.7%)
	Total	38 (8.7%)	211 (48.2%)	189 (43.2%)

effects of a feedback intervention designed to alert therapists about clients at risk for treatment failure. Nine percent of the control patients deteriorated during treatment, a figure consistent with the 5–10% estimate based on past reviews of psychotherapy outcome research (Lambert & Ogles, in press; Mohr, 1995). In addition, 38% of patients achieved either reliable change or clinically significant change by the time they left treatment, a figure that approximates the estimates (but is at the lower end) of outcome in clinical trial research (Hansen, Lambert, & Forman, 2002). These figures provided a baseline with which to compare the effects of quality improvement efforts on the entire sample of patients. In contrast, the experimental groups had a deterioration rate of 5% while 39% met criteria for reliable or clinically significant change. This difference was statistically significant ($F(5, 2605) = 11.644$, $p < .05$). The effect size was 0.09, a small effect according to Lipsey's (1990) criterion.

These overall differences in outcome between the experimental and control groups, while important, mask greater differences between control and experimental group signal–alarm cases for whom the feedback intervention was hypothesized to make its contribution, (i.e., feedback was hypothesized to specifically impact the outcome of patients at risk for treatment failure, and not all patients in general). Within the control groups, the decision rules identified about 22.4% of patients as making inadequate progress (signal–alarm cases) while in the feedback groups a similar number (22.3%) was obtained. Among these

signal–alarm cases the effects of feedback were more substantial. The treatment-as-usual signal–alarm condition had a deterioration rate of 21% compared to the 13% rate found in the feedback condition. Reliable and clinically significant change rates improved from the baseline of 21% in the no-feedback control groups, to a rate of 35% in the signal–alarm feedback condition (an effect size of 0.39).

Whether treatment effects of this size warrant recommendations for widespread application of feedback systems is an open question. It should be recalled that those advocating the use of empirically supported psychotherapies do so on the basis of much smaller treatment effects. For example, estimates of average effect sizes for differences between cognitive behavior therapy and competing treatments range from zero to a maximum of 0.20 across a variety of specific treatments for specific disorders (Lambert & Ogles, 2004; Wampold, et al., 1997). Yet advocates of specific treatments find such evidence compelling and some even suggest that the failure to use specific treatments for specific disorders, based on such empirical support, makes dissemination and application an ethical responsibility (Chambless & Hollon, 1998). Given the large sample sizes of the individual studies in this meta-analysis, and a combined overall sample size of over 2,500 cases, the current findings seem substantial if not compelling. Of course one need not choose between giving feedback and using empirically supported treatments. The recognition of potential treatment failure (signal–alarm) could provide an indication for using alternative treatments

or initiation of a stepped-care model (e.g., Newman, 2000; Otto, et al. 2000).

The design and scope of the studies in this review limited testing of hypotheses that could account for the effects of feedback on signal-alarm patient outcome. However, all three studies under review, and the overall finding from combining the studies, indicate that feedback to therapists about patient status indicative of possible treatment failure results in retention of clients in treatment for an average of about 8–9 (M = 8.61, SD = 5.36) sessions. It appears that feedback results in changes in therapist behaviors that, at a minimum, keep clients in treatment longer. It seems likely that therapists become more attentive to clients when they receive a signal from the research team that the client is not progressing.

Three research questions could be addressed in this meta-analysis that could not be fully addressed by the individual studies (or was addressed but provided inconsistent findings across the studies): Do trainees profit more from feedback than experienced licensed providers? Is a relatively late signal-alarm indicative of poor final response? Is the cost of providing more sessions to signal-alarm cases offset by providing fewer sessions to patients who appear to be on-track for a positive outcome?

The results suggest that therapists in training (whose work was supervised by the professional staff who also served as therapists in the current study) had overall outcomes that surpassed those of their professional counterparts. Despite these statistically significant differences between trainees and professionals, the size of the treatment effect (d = 0.15) was not substantial. The clients seen by trainees had somewhat better outcomes (relative to professionals) regardless of their status as signal-alarm cases. The majority of past studies on the effects of training on outcome report little difference in outcome between professionals and trainee/paraprofessionals therapists with only occasional superiority for the less trained (e.g., Stein & Lambert, 1995). The finding of statistically significant superiority for trainees remains just as much a puzzle in the present studies as those of the past. In the setting in which this research was conducted, trainees had a far smaller case load than the professional staff, perhaps allowing them to make a greater investment in each client. In addition, clients were not assigned randomly to professional and trainee therapists and the small superiority in outcomes could have resulted from unstudied selection biases. It is interesting to note that the trainees had more cases that

became signal-alarms during treatment, suggesting that supervision may have played a role in promoting positive outcomes. This interpretation of the small difference in outcome needs further study.

The findings of this meta-analysis did not show an interaction between training status, signal-alarm status, and feedback condition. When change *within* training-status groups was analyzed, it appeared that feedback made a significant difference in outcome across all clients seen by trainees while it had more specific effects on professionals by improving only the outcome of signal-alarm cases. This finding suggests that both groups were able to use the feedback to improve outcomes and that feedback was helpful across levels of training (albeit more broadly helpful with trainees). Apparently, therapists in both groups have methods of using the information that was offered in a way that benefits patients who are at risk for treatment failure. Elsewhere we have provided a model for outcome-informed supervision and recommended that supervisors, colleagues, case conferences, and case managers may best serve clients by focusing supervision on potential treatment failures rather than on cases that are progressing as expected (Lambert & Hawkins, 2001).

The correlation analysis of the session during which a signal-alarm was first given and final outcome was undertaken based on the assumption that a later alarm (not until the 5th session) would be a stronger signal for ultimate treatment failure. In the independent and separate analyses, Study 1 provided evidence that early warnings rather than later warnings were more malignant, whereas Study 2 found no relationship between time of warning and outcome. Study 3 did not report findings on this topic. The meta-analytic results suggest that relatively late versus early time to a signal-alarm was not a negative indicator of treatment success. Within the constraints of the methods used in the individual studies, it appears that clinicians need to be concerned if a patient reaches the threshold for an alarm anytime during treatment. Although initially we thought that an alarm occurring later in treatment would be more malignant because it came closer to the point of termination, the meta-analytic results lead us to speculate that later alarms may be indicative of temporary, stressful life events (external to therapy), while earlier alarms may be related to aspects of the therapy itself, possibly a failure in the alliance that leads to a loss of hope for a good outcome. The manner in which these data were collected do not allow for analysis of the reasons for this phenomenon. Consequently,

these interpretations are tentative but worthy of further empirical investigation as they have implications for the meaning of failing to recover as expected.

Study 1 suggested that using feedback resulted in more efficient psychotherapy, that is, in more sessions for not-on-track feedback cases and fewer for on-track feedback cases. Studies 2 and 3 had mixed results. The meta-analysis of all three studies suggested the presence of a statistically significant interaction. These results suggest that keeping signal-alarm cases in treatment longer not only improves their outcomes, but that the cost of doing so is offset by reducing the number of sessions engaged in by on track patients.

The finding that the use of feedback reduces the average per patient cost of treatment suggests that increasing the dose of sessions for potential treatment failures can be done in a cost-effective manner. This finding has important implications for third-party payers, employers, and government agencies that have concerns about providing quality treatment while also making treatment efficient. It could be argued that a mere one-and-one-half session increase (about a 20% change), while statistically significant, is unimportant. We believe that the magnitude of the effects of feedback on participation in therapy is substantial enough to be clinically meaningful in therapies that are already very short. Clearly more research is needed on the topic of efficient but effective care.

The major limitation of this meta-analysis is the homogeneous nature of the studies under consideration. While homogeneity can be a strength in meta-analytic reviews, the data for this review came from a single treatment setting, the research was conducted by the same research team, and used the same dependent measure. Ideally, there would be enough studies under consideration that a more heterogeneous sample of patients, therapists, and settings could be summarized. Although several other signal-alarm systems exist (e.g., Barkham et al., 2001; Brown et al., 2001; Kordy et al., 2001; Lueger et al., 2001), none has reported studies of their impact on treatment outcome.

The patient samples used in the present investigation were in some ways unrepresentative of outpatient psychotherapy settings in which the methodology might be applied in the future. Counseling center clientele are at the less severely disturbed end of the mental health continuum of clinical samples (Lambert, Hansen, et al. 1996) and have a disproportionately large number of clients with adjustment disorder diagnoses. Other settings, such as community mental health centers, are likely to identify a higher

rate of cases that are predicted to fail based on the decision rules employed in this study.

Just as important, the number of signal-alarms that were delivered to therapists was only half of what it would have been had this not been an experimental study. Recall that 50% of clinicians' cases were in the control (no feedback) condition and therapists received no progress information on these clients. Thus, notification of an alarm was less frequent for the individual therapist than it would be when applied in routine practice. Therefore, red and yellow messages were possibly more likely to catch a therapist's attention than they would be in a setting with more disturbed clients that were not involved in an experiment. The actual number of clients that would be identified for management will likely vary from setting to setting. The more common such cases are, presumably the less importance giving a signal would have for therapeutic practice, so that in some settings (and in the absence of a no-feedback control group) the use of the current algorithms could diminish the impact of a signal-alarm.

It is difficult to know, based on this meta-analysis, if the results will generalize to routine practice with more disturbed samples. However, in a yet-to-be-published study (Hawkins, Lambert, Vermeersch, & Slade, in press) undertaken in a hospital-based outpatient clinic, the impact of feedback appeared to be replicated despite the fact that almost 50% of clients were identified as signal-alarm cases. Further studies of feedback across a wider sample of patients are needed before we can be confident that the results of the three reviewed studies will generalize across a wide range of patient populations.

Another limitation of the studies analyzed in this meta-analysis is that the researchers made no attempts to govern how feedback was used by therapists. While this methodology heightened the likelihood that the results reflect what would happen in actual clinical applications, clinicians may have chosen a variety of ways to respond to the feedback. These include discussing the feedback with the client during sessions as a means of stimulating dialogue and exploration, using the feedback to inform case conceptualization and subsequent interventions, or essentially ignoring the feedback provided. Greater understanding of what makes feedback effective is needed.

Patient-focused research has a number of inherent limitations for applications in practice. Some other limitations grow out of methodological considerations that are a product of conducting this type of research. The current re-

search (and patient-focused research in general) is limited to a single self-report measure of improvement and therefore only provides one view of the presumed impact of therapy on clients. Furthermore, decisions regarding the continued provision of treatment should never be made on the basis of questionnaire-based expected treatment response curves alone. The decision rules used in this study, and the recovery curves on which they are based, can best be seen as one source of information, as fallible indicators of the more comprehensive assessments necessary to make individualized treatment decisions (Meyer, 1998; Strupp, 1996). Clinical trials and effectiveness studies can be, and usually are, much more comprehensive, thereby providing a more complex, complete, and perhaps accurate portrayal of change. However, if patient-focused outcome research is to have any applicability, it must remain simple and easy to implement in day-to-day clinical practice. The measure of outcome used in these studies correlates well with other measures commonly used in outcome studies but these measures also rely on self-report methodology.

Since the use of feedback depends on frequent albeit brief assessments, measuring client status requires clinicians, practicing outside of a research protocol, to make substantial changes in the way they practice. Generally speaking, clinicians do not see the value of frequent assessments based on standardized scales (Hatfield & Ogles, 2002) because they are confident in their ability to accurately observe client worsening and provide an appropriate response. Despite evidence that suggests therapists are not alert to treatment failure (Yalom & Lieberman, 1971; Meyer & Schulte, 2002), therapists' confidence in their own clinical judgment stands as a barrier to implementation of monitoring and feedback systems. The idea that expected recovery curves and decision tools based on standardized measures can have a positive patient benefit is novel and in some ways may be seen as undermining clinical judgment, despite overwhelming evidence that clinical judgments are usually found to be inferior to actuarial methods across a wide variety of predictive tasks (Meehl, 1954; Sines, 1970).

It remains to be seen if the evidence on the effects of feedback will, in fact, have widespread impact in routine practice because implementation of such systems remains a serious obstacle. This problem parallels the problem found in clinical trials research and meta-analytic reviews that support one theoretically based treatment over another. If the results of meta-analytic reviews consistently favored one treatment over another, it might require ther-

apists to modify their practices and possibly modify long held beliefs about the nature of psychopathology and psychotherapy. Tracking patient treatment response in a systematic way may require more openness to advances in the field than clinicians are able to accept or implement.

The results of the research summarized in this meta-analysis suggest the value of continuing to study these promising methods and their potential clinical utility through widespread application in routine practice. While future research is recommended to understand the limits of applicability of these methods, several studies are already underway that extend the research in new directions. As already noted, the third study in this series attempted to use clinical support tools for helping therapists with signal-alarm cases. These tools include a decision tree for guiding inquiry about why a treatment may be failing, additional assessment scales (therapeutic alliance, social support, and readiness for change) for assistance in identifying new courses of action, and recommendations for alternative courses of action depending on what is determined to be the problem. Initial results (Whipple et al., 2003) suggest that these tools increase the impact of feedback alone. Attempts to replicate and extend these findings are underway. In a separate line of research, the effects of providing patients with feedback on their progress using similar messages to those given to therapists is being conducted (Hawkins, et al., 2002) with preliminary results indicating that patient feedback adds to the benefits of therapist feedback. Finally, clinical trials and process research are necessary if we hope to better understand how therapists use information from feedback and clinical support tools to improve outcome for their signal-alarm clients. We encourage further research on these promising methods.

REFERENCES

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

Andrews, G. (2000). A focus on empirically supported outcomes: A commentary on the search for empirically supported treatments. *Clinical Psychology: Science and Practice, 7,* 264–268.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE_OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69,* 184–196.

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2002). *Clinical significance methods: A comparison of statistical techniques.* Manuscript submitted for publication.

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. *Behavior Analyst Today, 4,* 79–90.

Beutler, L. E. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. *Journal of Consulting and Clinical Psychology, 69,* 197–204.

Brown, G., Burlingame, G. M., Lambert, M. J., Jones, E., & Vaccaro, J. (2001). Pushing the quality envelope: A new outcomes management system. *Psychiatric Services, 52,* 925–934.

Brown, G., & Lambert, M. J. (1998, June). *Tracking patient progress: Decision making for cases who are not benefiting from psychotherapy.* Paper presented at the annual meeting of the Society for Psychotherapy Research, Snowbird, Utah.

Chambless, D. L., & Hollon, S.D. (1998). In defense of dissemination of empirically supported psychological interventions. *Journal of Clinical and Consulting Psychology, 66,* 7–18.

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 Research, 20,* 240–257.

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

Garfield, S. L. (1994). Research on client variables in psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 190–228). New York: Wiley.

Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy. *Clinical Psychology: Science and Practice, 3,* 218–229.

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.*

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.

Hatfield, D. R., & Ogles, B. M. (2002). The current climate of outcome measures use in clinical practice. Unpublished

Manuscript. Department of Psychology. Ohio University, Athens, OH.

Hawkins, E. J., Lambert, M. J., Vermeersch, D., & Slade, K. (in press). The effects of providing patient progress information to therapists and patients. *Psychotherapy Research.*

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

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Efficacy, effectiveness, and client progress. *American Psychologist, 51,* 1059–1064.

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.

Kendall, P. C. (1999). Clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 283–285.

Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 285–299.

Kordy, H., Hannöver, W., & Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg model. *Journal of Consulting and Clinical Psychology, 69,* 173–183.

Lambert, M. J. (2001). Psychotherapy outcome and quality improvement: Introduction to the special section on client-focused research. *Journal of Consulting and Clinical Psychology, 69,* 147–149.

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., Hansen, N. B., & Finch, A. E. (2001). Client-focused research: Using client outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172.

Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G., Huefner, J. C., & Reisinger, C. W. (1996). *Administration and scoring manual for the Outcome Questionnaire (OQ 45.2).* Wilmington, DL: American Professional Credentialing Services.

Lambert, M. J., & Hawkins, E. J. (2001). Using information about client treatment progress in supervision: Are outcomes enhanced? *Australian Psychologist, 36,* 131–138.

Lambert, M. J., & Ogles, B. M. (in press). The efficacy and effec-

tiveness 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, 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., 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.

Lipsey, M. W. (1990). *Design sensitivity.* Newbury Park: Sage Publications.

Lueger, R. J., Howard, K. I., Martinovich Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual clients using expected treatment response models. *Journal of Consulting and Clinical Psychology, 69,* 150–158.

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

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

Meyer, G. J. (1998, August). A consumer's perspective on outcome assessment. In K. L. Moreland (Chair), *Different perspectives on the measurement of treatment outcomes.* Presented at the 106th annual convention of the American Psychological Association, San Francisco, CA.

Meyer, F. & Schulte, D. (2002). Zur validitat der beurteilung des therapieerfolgs durch therapeutin. *Zeitschift fur Klinische Psychology und Psychotherpie, 31,* 53–61.

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice, 2,* 1–27.

Mueller, R., Lambert, M. J., & Burlingame, B. M. (1998). The Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70,* 246–262.

Nathan, P. E. (1998). Practice guidelines: Not yet ideal. *American Psychologist, 53,* 290–299.

Newman, M. G. (2000). Recommendations for a cost-offset model of psychotherapy allocation using generalized anxiety disorder as an example. *Journal of Consulting and Clinical Psychology, 68,* 549–555.

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.

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63,* 179–183.

Sines, J. O. (1970). Actuarial versus clinical prediction in psychopathology. *British Journal of Psychiatry, 116,* 129–144.

Strupp, H. H. (1996). The tripartite model and the Consumer Reports study. *American Psychologist, 51,* 1017–1024.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically validated psychologist treatments: Report and recommendations. *Clinical Psychologist, 48,* 3–23.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000).Outcome Questionnaire-45: Item sensitivity to change. *Journal of Personality Assessment, 74,* 242–261.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122,* 203–215.

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.

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