Patient-Focused Research: Using Patient Outcome Data to Enhance Treatment Effects

Michael J. Lambert, Nathan B. Hansen, and Arthur E. Finch
Brigham Young University

A program of research aimed at improving the quality of psychological interventions is described. Data from over 10,000 patients were analyzed to understand the association between number of treatment sessions and clinically significant improvement. In addition to a potential dose–response relationship, typical recovery curves were generated for patients at varying levels of disturbance and were used to identify patients whose progress was less than expected ("signal" cases). The consequences of passing this information along to therapists were reported. Analyses of dose–response data showed that 50% of patients required 21 sessions of treatment before they met criteria for clinically significant improvement. Seventy-five percent of patients were predicted to improve only after receiving more than 40 treatment sessions in conjunction with other routine contacts, including medication in some cases. Identification of signal cases (potential treatment failures) shows promise as a decision support tool, although further research is needed to elucidate the nature of helpful feedback. Outgrowths of this research include its possible contribution to social policy decisions, reductions in the need for case management, use in supervision, and possible effects on theories of change.

A majority of recent psychotherapy outcome research has used a clinical trials paradigm, which typically contrasts the improvement of a sample of patients with a single diagnosis who receive a specific psychotherapy with the improvement of a sample of similar patients receiving a competing treatment, no treatment, or placebo. This research suggests that a wide variety of psychotherapies are effective (Lambert & Bergin, 1994; Smith, Glass, & Miller, 1980). In contrast with clinical trials research, a body of current inquiry on psychotherapy outcome (so-called "effectiveness research") aims to understand patient change in the context of therapy as it is actually practiced in day-to-day clinical settings (e.g., Seligman, 1995). Investigations of effectiveness range from posttherapy surveys to program evaluations, often without adequate experimental control. Although effectiveness research may be more generalizable to actual clinical practice than clinical trials research, much of it fails to affect practice because of the style in which reports are written and the considerable time that can pass before the results of such studies are available to clinicians.

Howard, Moras, Brill, Martinovich, and Lutz (1996) introduced patient-focused research as a new paradigm for evaluating psychotherapy and is aimed at monitoring an individual patient's progress over the course of therapy. This research information can serve as valuable feedback to the practitioner, supervisor, or case manager, who can make attendant treatment modifications in real time. Although clinical trials research (efficacy research) emphasizes the average response of patients to treatment in highly controlled experimental conditions, and effectiveness research focuses on the mean response of patients in naturalistic settings, patient-focused research attempts to answer the question, Is this particular treatment working for this patient? All three types of research are essential to enhancing the quality of treatment offered to patients and to placing psychotherapy on a firm empirical foundation. They are complementary to each other, often overlapping, and can inform one another, providing the synergy that may be necessary for rapid scientific progress and improved outcomes.

The purpose of this article is to present results from an ongoing program of patient-focused research ultimately endeavoring to enhance the treatment outcome of the individual patient. We introduce a conceptual framework for such research by operationally defining outcome and clinically significant or meaningful change. Efforts to examine a dose–effect relationship are described. The use of these constructs and the dose–effect data provide the foundation for identifying expected recovery curves and for building a system for tracking patient progress. Finally, we present data on the effects of giving information to clinicians about patients who are failing to show an expected pattern of improvement.

Patient outcome within our research program was operationalized through the use of a standardized self-report instrument: The Outcome Questionnaire—45 (OQ–45; Lambert, Hansen, et al., 1996). This questionnaire was developed specifically for tracking outpatients on a weekly basis. Whereas there is a clear consensus that adequate measurement of outcome requires data from multiple

sources (the patient: outside informants, such as the therapist, trained judges, and significant others; and markers of daily functioning, such as hospitalization, employment records and the like) implementation of patient-focused research often requires far simpler and less costly efforts. Treatment systems cannot tolerate expensive and time-intensive markers of change, especially when used as a start-up procedure or where patient progress is reported to therapists on a weekly schedule.

There were literally hundreds of available instruments to choose from that measure a variety of problems typically addressed in psychotherapy (Froyd, Lambert, & Froyd, 1996), yet none met all the requirements that were seen as desirable to operationalize patient improvement. Any such instrument would be brief, easily administered and scored, suitable for patients with a wide range of diagnoses, psychometrically sound, sensitive to change over a short period of time, and inexpensive (as it would be repeatedly administered to thousands of patients). In addition to these qualities, it was also considered essential to define outcome as changes in three content domains: symptomatic functioning (mainly anxiety and depression), interpersonal problems (friendship and family relations), and social role performance (work adjustment and quality of life). These areas of functioning are widely recognized as the essential ingredients of interest when assessing patient improvement (Lambert & Hill, 1994; Strupp, Horowitz & Lambert, 1997; Waskow & Parloff, 1975).

Because the methodology of patient-focused research demands repeated measurement with individual patients over time, the question of the impact of repeatedly administering a test to the same patient on several occasions is especially important. This practice has been found to produce reports of reduced symptomatology on some measures (e.g., Jorm, Duncan-Jones, & Scott, 1989). Investigation of this test–retest artifact (Durham, Burlingame, Schaalje, & Lambert, 1999) suggested that test scores decrease slightly at the second assessment but that decreases are not cumulative across further testing periods in untreated individuals. Given that the OQ–45 had the required psychometric properties for our purposes and takes about 5 min to complete, it was considered well-suited for our research.

Clinical trials, and to some degree effectiveness research, depend on observing statistically significant differences in groups of individuals following treatment; patient-focused research depends on observing noticeable changes within each individual patient. It is therefore essential to be able to define the presence or absence of meaningful change in each and every patient. Various methods have been proposed for accomplishing this task (e.g., Christensen & Mendoza, 1986; Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991; Kendall, 1999; Kendall, Marr-Garcia, Nath, & Sheldrick, 1999). Jacobson et al. (1984) operationalized "clinically significant change" as (a) patient movement from the ranks of the dysfunctional into the ranks of the functional (based on normative comparisons); and (b) movement so large that it was not likely to be the result of measurement error (reliable change). Following this standard, patients are classified as showing clinically significantly change if they meet both criteria.

In actual practice, some patients begin treatment with scores lower than that expected of the dysfunctional population, and on occasion, those with scores in the dysfunctional range may worsen rather than improve. Thus, those who begin treatment in the functional range cannot meet both criteria for clinically significant change, although they can meet the standard of reliable change. In the present program of research we consider those whose movement is reliable (meeting Marker b above) to be "improved" but not clinically significantly improved. A patient who begins treatment in the functional range and who passes the cutoff score, entering the ranks of the dysfunctional, and who reliably worsens is considered "deteriorated." Those who begin treatment in the dysfunctional range and who reliably worsen are considered to be "deteriorators" as well. A patient who does not meet the preceding criteria is considered to be in the "no change" category.

It may not be readily apparent to the reader just how essential it is to have a standard method of classifying the degree of response by an individual patient. The concept and operationalization of clinically significant change is central to patient-focused research because when tracking patients is initiated a marker is necessary to trigger decisions about the need for further treatment, termination, or referral. This methodology has its roots in single-case designs and behavior therapy but is suitable for research in other domains (Kazdin, 1993). Although the Jacobson & Truax (1991) method is just one of several possible standards, it is easily applied to standardized rating scales such as the OQ–45 (Hansen & Lambert, 1996; Lambert & Bergin, 1994; Lambert & Hill, 1994).

With this as a conceptual background, for the remainder of this article we address some of the findings associated with our ongoing program of research. In the first study, we summarize explorations of the dose–response relationship and present original analysis of a large database from across the nation. This data is relevant to third-party payers, government policymakers, and administrators who make decisions about benefit levels that are necessary to provide adequate treatment to patients with psychological disorders. It is also relevant to clinicians as they develop expectations for the length of treatment that may be necessary for a given patient to improve. This dose–response research provided a foundation and a baseline for later studies that focused on the response of specific patients to treatment. In the second study, we present the results of generating average recovery curves. These curves provided the basis for judging whether a particular patient's early progress in therapy is sufficient and likely to result in ultimate improvement or whether he or she is predicted to leave therapy before experiencing a measurable or meaningful benefit. Finally, we summarize the results of a third study. In this previously published study, we used an application of recovery-curve data to identify treatment failures, provide therapists with this information, and examine whether this feedback improved outcomes.

Study 1: The Dose–Response Relationship

Initial research efforts studied the relationship between the number of sessions of treatment patients had and their degree of improvement. This relationship is important for numerous practical reasons, the major one involving social policy—how much therapy is needed to ensure that patients have achieved an adequate benefit from undergoing psychotherapy? Howard, Kopta, Krause, and Orlinsky (1986) examined the issue and used probit analysis to estimate recovery on a session-to-session basis. The analysis suggested very rapid initial recovery followed by smaller gains over the course of treatment. It also suggested that few sessions were needed (8 sessions for 53% of patients to get a satisfactory benefit). One problem with the Howard et al. (1986) analysis was its reliance on pre- and post estimates of patient improvement rather

than session-by-session ratings of improvement. As a result, we undertook a study aimed at estimating the dose–effect relationship (Kadera, Lambert, & Andrews, 1996) and then replicated this study including a 6-month follow-up to see if patients maintained their gains (Anderson & Lambert, in press).

Anderson and Lambert (in press) combined the Kadera et al. (1996) sample with their own replication to estimate the dose–effect relationship. Using both samples, they found that improvement in therapy was much slower than the review by Howard et al. (1986) indicated. In fact, they found that 50% of patients needed 13 sessions of psychotherapy before reaching criteria for clinically significant change. When a lesser standard (reliable change) was used, 10 and 17 sessions, respectively, were needed for 50% and 75% of patients to meet the criterion of a 14-point improvement. The need for more sessions was related to higher initial levels of distress. Patients tended to maintain their improvement at follow-up or to continue improving without additional therapy.

Both of the preceding studies, although improving on the methodology of Howard et al. (1986), had significant potential limitations. The primary limitations were that the therapy was provided in a training clinic with a small number of patients ($N = 140$). Thus, it was important to discover whether the results would hold up in nontraining settings with licensed professionals and larger, more diverse patient samples.

Method

Participants. Participants were drawn from our National Database. This database was created by having provider groups, including managed care organizations, forward data to the Brigham Young University Psychotherapy Research Center in return for free use of the OQ–45. The participants took the OQ as part of receiving services from agencies and providers, with the understanding that it could be used to improve the clinical services they received and assist in their therapist's treatment efforts. Because its use was not experimental, informed consent was not obtained and results of the testing were occasionally shared with patients at the discretion of the therapists. No formal system was implemented for recording how many, how often, or which therapists shared OQ–45 results with patients.

We made the following analyses with data from five sources. One source was the national Employee Assistance Programs (EAPs) from 30 states across the nation; these programs typically treat patients for one to eight sessions and typically refer patients who appear to need longer treatment to outside providers. Patients ($N = 3,269$) in these settings were predominantly diagnosed with adjustment disorders. We also used outpatient psychotherapy settings from either national-level managed behavioral health companies that contract with independent provider groups or individual providers ($N = 536$) or from regional organizations that manage care through similar organizational programs ($N = 595$). These patients were typically diagnosed with a variety of Axis I mood and anxiety disorders. Treatment length was negotiated between the patient and therapist within the constraints of insurance carriers, who typically limit treatment to 20 sessions per year. The third source we used was a university-based counseling center ($N = 1,188$). Patients in this setting entered counseling for personal problems rather than for career concerns and were diagnosed with a wide variety of Axis I disorders. The age range of this sample of patients is more restricted than those in the other groups, and a greater proportion received adjustment disorder diagnoses than in the outpatient settings. Treatment length was negotiated between the patient and therapist with little external influence on treatment length limits. We also used a university-based training clinic. Patients from this setting ($N = 123$) are similar to other outpatient samples in terms of degree of disturbance as determined by both initial OQ scores and diagnostic make-up. Most of these patients did not have insurance and thus were willing to seek help in a training clinic where therapy sessions were recorded and sessions were under the supervision of licensed psychologists. Severe Axis II patients were excluded from treatment and referred elsewhere. Treatment lengths were negotiated between therapist and patient with no external pressure to arbitrarily set limits. Our fifth source was a state-funded, not-for-profit, community mental health clinic. This setting serves low socioeconomic-status patients ($N = 361$), who typically experience more chronic mental health concerns and Axis II problems. Treatment lengths are negotiated by patient and therapist with no arbitrary session limits.

Within Study 1, a total of 6,072 patients provided data, and the therapists providing treatment numbered in the hundreds and ranged from students in training to highly experienced professionals. Although it is hard to estimate the degree to which they represent any particular class of providers, the therapists did represent providers from clinics and agencies from over 35 states across the United States. These providers conducted therapy as usual, with a wide variety of treatment methods. Most described their practice as eclectic with cognitive–behavioral methods being most commonly used. No attempts were made to limit therapy to a specific treatment method or to monitor the quality of the therapy that was provided, and the database provided no provision for linking specific treatment methods and outcome. All of the patients received psychotherapy, and 20–30% also took medication, although the kind of medication taken and the particular patients who received it were not part of the database.

Measure. The OQ–45 is a 45-item self-report measure that takes about 5 min to complete and attempts to assess three content domains elaborated on by Lambert (1983). It is composed of three subscales that assess three aspects of the patient's life: (a) subjective discomfort (symptoms of pathology), (b) interpersonal relationships, and (c) social role functioning. These three broad domains capture two of three salient foci noted by Strupp and Hadley (1977) in their tripartite model of change, where problems of concern to the person and to society are emphasized. Their third domain, change in terms of idealized mental health, was also observed through nine items that assess mentally healthy functioning and quality of life. Each item on the OQ–45 is rated on a 5–point scale ($0 = never$, $1 = rarely$, $2 = sometimes$, $3 = frequently$, or $4 = almost always$), yielding a range of possible scores from 0 to 180, with higher values indicating poorer functioning.

The OQ–45 total score is recommended for measuring outcome because each of the three subscales has a small number of items and is highly correlated. Lambert, Hansen, et al. (1996) reported adequate internal consistency for the OQ–45 ($r = .93$). The 3-week test–retest value for the OQ–45 is also satisfactory ($r = .84$; Lambert, Burlingame, et al., 1996). Concurrent validity figures—as estimated by correlating the total score with the Symptom Checklist—90—R (Derogatis, 1983), Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961); Zung Depression Scale (Zung, 1971), Taylor Manifest Anxiety Scale (Taylor, 1953), State–Trait Anxiety Inventory (Spielberger, 1983; Spielberger, Gorsuch, & Lushene, 1970), Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988), and the Social Adjustment Scale (Weissman & Bothwell, 1976)—were all significant at the .01 level ($rs = .50–.85$). Normative information based on data collected across the country for the OQ–45 has also been reported (Lambert, Burlingame, et al., 1996; Lambert & Finch, 1999; Lambert, Hansen, et al., 1996; Umphress, Lambert, Smart, Barlow, & Clouse, 1997).

The factor structure of the OQ–45 has been examined across two samples with exploratory and confirmatory factor analyses (Mueller, Lambert, & Burlingame, 1998). The results were confirmed as being consistent with the initial conceptualization of the instrument as a three-factor solution. More important is the degree to which the OQ–45 scores are sensitive to change in treated patients yet remain stable in untreated persons. Vermeersch, Lambert, & Burlingame (2000) examined the slope of change of each item by using hierarchical linear modeling with nonpatients, high scoring nonpatients, and patients undergoing treatment over a period of 2

months. All but 2 of the items showed slopes that were significantly steeper than zero, and 30 of the items showed steeper slopes (improving scores) than the untreated controls over this same time period. The total scores of patients who repeatedly took the OQ–45 improve as a function of number of treatment sessions whereas the total scores of those who go untreated either do not (or do, but to a significantly lesser degree).

Using formulas developed by Jacobson et al. (1984) and modified by Christensen and Mendoza (1986; see Jacobson & Truax, 1991), researchers analyzed the normative data of the OQ to provide the cutoff scores for the Reliable Change Index (RCI) and clinically significant change cutoff. By using normative data from 1,353 community nonpatients gathered from communities across the United States and from 1,476 patients from similar locales who were entering treatment, the RCI was estimated to be 14 points. Patients who change in a positive or negative direction by at least 14 points are regarded as having made reliable change. Such movement is seen as exceeding measurement error and as having met one of the two criteria posited by Jacobson and colleagues (Jacobson et al., 1984; Jacobson & Truax, 1991) as indicative of clinically meaningful change. The cutoff for demarcating the point at which a person's score is more likely to come from the dysfunctional population than a functional population was estimated to be 64 according to the Jacobson and Truax (1991) formula. This score turns out to be about one standard deviation above the mean of the nonpatient sample. When patients' scores fall at 63 or below, the conclusion is that their functioning at that point in time is more similar to nonpatients than to patients (Lambert, Hansen, et al., 1996).

Data analysis. Survival analysis, as used in this study, is a nonparametric procedure used to assess longitudinal data. Survival analysis differs from other more traditional means of analyzing data in that it tracks the patient across several points in time (two thirds of patients had more than two data points) making for a robust test of meaningful clinical change. This approach is contrasted with the Howard et al. (1986) data and more typical research that uses cross-sectional and longitudinal designs, which assess either the effects of treatment at a single point in time or at baseline and termination. For instance, general linear modeling attempts to predict a single dimensional variable, such as the magnitude of change on a standardized test during the course of treatment. Use of survival analysis for psychotherapy outcome prediction differs from linear regression by predicting the outcome status of patients at any point in time along the course of treatment. Survival analysis is similar to logistic regression in that it targets a dichotomous variable, such as recovered–unchanged or success–failure. Logistic regression, however, attempts to predict the dichotomous event at a static time point, such as at treatment termination, whereas survival analysis tracks the dichotomous event across time. Traditionally, dose–response research, on the one hand, has tried to answer the question, How much? by examining the incremental effect on outcome of an increase in the dose of treatment. Survival analysis, on the other hand, attempts to answer the question, How long? by determining at what point in time an event occurs. In psychotherapy, where a typical dose is considered to be a weekly 50-min session, How much? and How long? tend to be the same question. Survival analysis has an advantage over traditional probit or logistic methods because it models outcome as a function of time by analyzing growth curves to assess dose–response.[1]

For the current study, Cox regression was applied to the entire data set, with two variables, data collection site and pretreatment level of functioning, used as covariates to determine the effect of these two variables on therapy outcome. For this analysis, time was specified as therapy session, with each session of therapy representing one unit of time. Outcome was defined as either improving or recovering during the course of treatment. A terminal event was considered to have occurred when a patient in treatment reached the outcome condition of improved or recovered and the patient's status did not change from this condition, whether or not treatment was terminated at this point. Patients were considered censored if they had not met the terminal condition by the time they left treatment or if no further data were available from them.

Results and Discussion

This analysis determined what effect the variables of treatment site and pretreatment level of functioning had on the survival function. The results for these variables are summarized in Table 1. The Cox regression analysis was based on the assumption that there was one survival curve and that this curve was modified by the variables of treatment site and pretreatment level of functioning.

For the above analysis, all of the variables were forced into the model. Procedures such as stepwise entry or backward removal of variables can be used as in logistic regression. Also, possible interactions were explored between treatment sites and level of pretreatment functioning. The overall goodness-of-fit for the model was $\chi^2(6, N = 6,072) = 298.35, p < .001$, which indicates that the model that predicted no differences in the data did not fit the observed data. Therefore, significant differences were found in the model because of the influence of the entered variables.[2]

The overall variable for treatment site had a significant effect, with a Wald $\chi^2(5, N = 6,072) = 60.40$. As noted above, the significance of the various sites depends on which other site it is being compared with in the analysis. The level of pretreatment functioning also had a significant influence on survival. This

---

[1] A key element of survival analysis is that it takes into account censored data. Data are considered censored if the event of interest is not yet reached but further data are no longer available on that participant. Working with large data sets across time creates many limitations in data collection. Therefore, methods have been developed to work with data within these challenging frameworks. Longitudinal data may be censored because of a variety of complications. Three of the most common reasons for censoring include: (a) The follow-up duration is too short, resulting in participants who have not reached the terminal event by the end of the study; (b) participants withdrawing from the study prior to reaching the terminal event, resulting in no data on these participants after their withdrawal; and (c) losing participants because of a competing event that is unrelated to the study (Feinstein, 1996). For example, this may include a patient in a psychotherapy treatment study who, outside of study protocol, obtains psychotropic medication after initiating psychotherapy within the study. From the point of receiving medication onward, the patient is under the influence of a competing event that is unrelated to the psychotherapy outcome study. The data from this patient are usable up to the point where medication treatment was initiated, but beyond this point the patient must be considered censored and further data from this patient are unusable. Survival analysis relies on the use of a defined time series that includes a definite time of origin and a specified event that when reached provides an endpoint. Survival analysis is popular in biomedical research for analyzing time-related events such as time to death, relapse of symptoms, or recovery from symptoms (Collette, 1994). For the computation of survival analysis (Cox & Oakes, 1984), one must have a clear time origin (e.g., therapy intake), a specified time interval (e.g., weekly therapy sessions), and a clearly established endpoint (e.g., reaching clinically significant improvement on a therapy outcome measure). It is also assumed that censoring (patient dropout) occurs randomly.

[2] Examination of the variables in the model revealed several significant differences. First, in comparing the sites, it should be noted that in Table 1 the line for the local HMO was selected to be the baseline value for comparison. This site was selected as the baseline simply because it was the last site entered into the model. The model was run with each site entered last, revealing that there were two clusters of sites that differed from each other, although the sites within the clusters were not significantly different. These two clusters were composed of the HMOs and (EAP) site falling together and the community mental health centers (CMHs) and counseling center site falling together.

Table 1

Summary of Cox Regression Comparing Effect of the Variables Treatment Site and Pretreatment Level of Functioning on Overall Survival

Variable	B	SE	Wald	df	R	Exp (B)
Site base (local HMO)			60.398***	5	.039	
Site 1 (training CMH)	−.415	.213	3.801	1	−.007	.660
Site 2 (counseling center)	−.394	.144	7.541**	1	−.013	.674
Site 3 (EAP)	.096	.135	0.506	1	.000	1.101
Site 4 (national HMO)	.007	.161	0.002	1	.000	1.007
Site 5 (state CMH)	−.448	.218	4.227*	1	−.008	.639
Pretreatment functioning	−.923	.261	12.475***	1	−.018	.397
Interaction: Site × Pretreatment functioning			2.417	5	.000	
Site 1 × Pretreatment functioning	.313	.423	0.547	1	.000	1.367
Site 2 × Pretreatment functioning	.258	.286	0.816	1	.000	1.294
Site 3 × Pretreatment functioning	.165	.270	0.371	1	.000	1.179
Site 4 × Pretreatment functioning	.396	.322	1.513	1	.000	1.486
Site 5 × Pretreatment functioning	.082	.436	0.035	1	.000	1.085

Note. $\chi^2(5, N = 6,072) = 298.35, p < .001$. Local HMO = intermountain area clinics providing outpatient psychotherapy; training CMH = community outpatient clinic staffed by mental health professional trainees; counseling center = outpatient clinic serving university students; EAP = employee assistance programs in companies housed across the United States; national HMO = outpatient psychotherapy settings across the United States; state CMH = community mental health center in the state of Ohio.
*$p < .05$. **$p < .01$. ***$p < .001$.

variable was entered as a categorical variable and indicated whether a patient was within or not within the functional distribution prior to treatment. This information was also entered as a continuous variable by using each patient's pretreatment OQ–45 total score, which resulted in a significant effect. There were no significant interaction effects, which simplifies interpretation of the model. The model can be summarized by saying that where a patient receives treatment has an impact on the rate of change. A patient's pretreatment functioning also influences rate of change, with higher pretreatment distress producing faster change than does lower pretreatment distress.[3]

The results of the survival analysis are also displayed graphically in Figure 1 and suggest that 50% of the patients who begin treatment in the dysfunctional range can be expected to achieve clinically significant change after 21 sessions of psychotherapy. It takes more than double this number of treatment sessions before 75% of patients reach this same criteria. Using a lesser standard (reliable improvement), we estimated 50% of patients to improve after 7 sessions and 75% to improve after 14 sessions (see Figure 2).

These results were based on an exceptionally large number of patients and therapists gathered from across the United States and suggest that benefits for mental health coverage need to extend to somewhere between 21 and 45 sessions per episode of illness, if the goal of treatment is to restore the patient to a level of functioning similar to nonpatient samples. Even this level of coverage will not be sufficient for 25–50% of patients. However, the composition of the sample makes estimation of the upper limit on the number of sessions problematic, as the samples are possibly over-represented by less-disturbed patients. At the same time, these data support the conclusions that a substantial number (a majority) of

patients can profit from brief treatment (15 sessions) by achieving a reliable—albeit insufficient—degree of success.

Interpretation of these results is somewhat limited by the unknown effects of possible violations of the assumption that censoring occurs randomly and is not systematically related to the course of treatment. In this study we had no means of testing this assumption, although in prior research (Anderson & Lambert, in press) there was no statistically significant trend for change scored to be associated with reasons for leaving the study (e.g., satisfaction with treatment, moving, end of semester, etc.). Research is needed to further examine violations of the assumptions of censoring in psychotherapy.

The preceding investigation provided a baseline from which to start evaluating expected patient outcomes in treatment. From this information, general treatment guidelines can be established for estimating time to recovery that take into account such variables as treatment setting and the severity of patient symptoms. Survival

---

[3] The values in the last two columns of Table 1, labeled *R* and Exp (B), provide means to examine the individual effects of the variables. The value *R* can be considered a partial correlation, with a value ranging from 0 to 1 (or −1 to 1, although the magnitude is equal on either side of zero, as in typical correlations). This value is useful for ranking the effect size of the individual variables, with a larger absolute value of *R* indicating predictive power. The value labeled Exp (B) contains the exponent of the beta coefficient (from the column labeled B) from the linear model for each variable. This value is analogous to an effect size and can be thought of as the amount of change that can be attributed to the variable. Feinstein (1996) refers to this value as the "hazard ratio for a one-unit change" (p. 390).
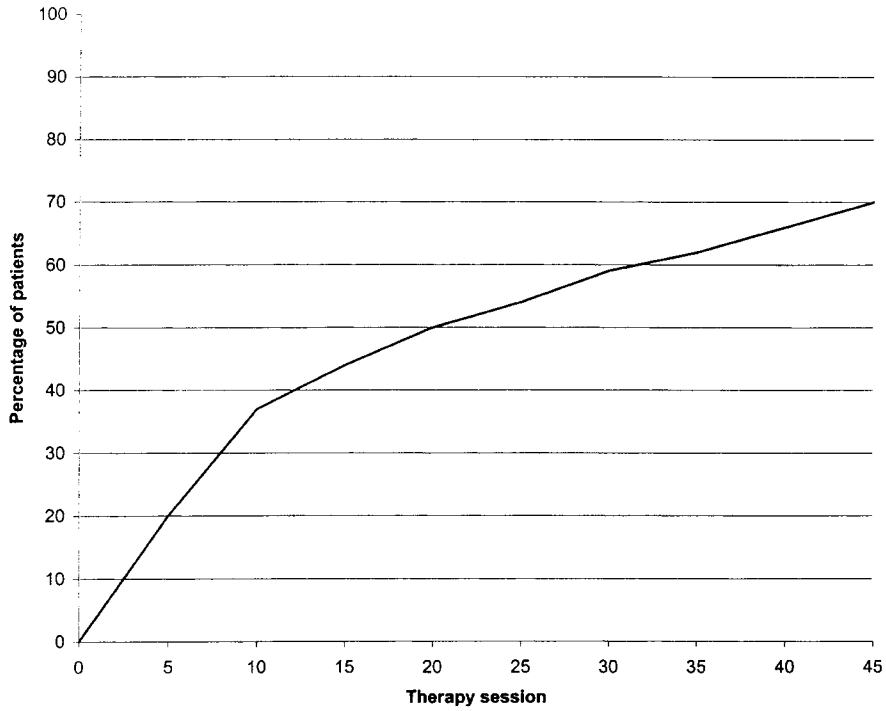
Figure 1. Time to recovery measured by clinically significantly improved criteria.

models are based on aggregates of patients, and the treatment course of an individual patient will rarely match the group as a whole, but the survival analysis results presented here suggest the possibility of establishing expected recovery curves and identify-

ing patients whose success is less than expected, at which time further data on response to treatment can be gathered and treatment modified in response to this information. Can we identify typical or expected trajectories of change?
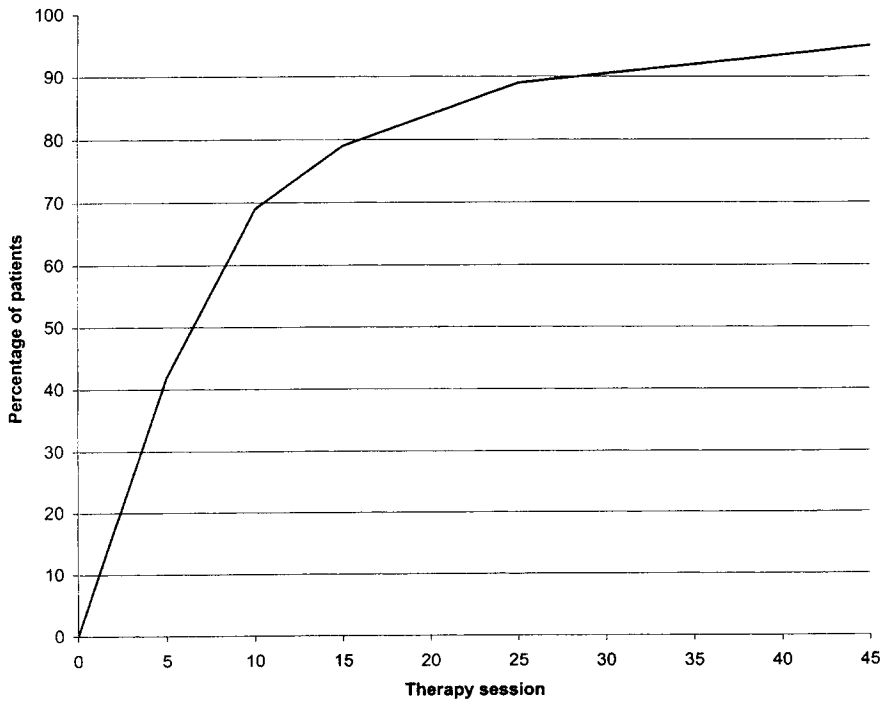

Figure 2. Time to improvement using reliable change index.

Study 2: Development of Standard Recovery Curves and Their Use for Identifying the Patient Whose Success in Therapy Is Doubtful

Our initial attempts to approach the task of predicting recovery for specific patients included using patient information, such as diagnosis, history of past treatment, symptom severity, expectations–motivation, and initial level of functioning in social roles. Multiple regression was used as a tool to predict recovery. Two variables accounted for the majority of the variance in final outcome: initial level of distress and early response to treatment (i.e., amount of change from pretest through Session three; Brown & Lambert, 1998, June). In fact, these two variables had an R^2 of .40 with final outcome in an independent sample of outpatients. The other variables made a minimal contribution to predicting change by the conclusion of treatment. Given this information, we felt it was reasonable to use these two variables to create a system for tracking and evaluating patient progress in therapy. Using just two variables also had the advantages of simplicity and efficiency because severity could be estimated with the patient's initial OQ score, whereas early response to treatment and final outcome could be estimated from change scores calculated at subsequent sessions on the same measure.

Ideally, a recovery curve should be developed for each initial score on the OQ–45, resulting in 180 possible courses of improvement. However, this proved impractical even with our large database because there were few patients with extremely low and extremely high scores. Nevertheless, we attempted to develop recovery curves that would be practical to use in everyday clinical practice to judge whether a patient at a given score-band of disturbance was responding to therapy like an average patient within that same band of disturbance. Of greatest interest was the identification of patients whose response to therapy was so slow as to suggest that they might fail to achieve final success (clinically significant or reliable change) during the course of treatment.

Method

Participants. The same National Database was used as in Study 1, but subsequently acquired data were also included. This latter data came from the same setting and from similar settings, with approximately the same proportion of patients from each type of setting. As in Study 1, the therapists were not screened and selected but represented private practitioners and agency staff who routinely performed services in their setting. The analyzed sample consisted of 11,942 patients who took the OQ–45 before treatment and at least one subsequent test period prior to or at the time of treatment termination. The majority of patients (65%) provided three or more OQ–45 test scores.

Measure. The OQ–45 was used. No other measures were collected for the purposes of this analysis.

Data-analytic procedures. In Howard et al.'s (1986) study of the dose–effect relationship, it was noted that the initial sessions of therapy produced a much more dramatic change in self-reported distress than did later sessions. This suggests that as therapy progresses, a larger number of sessions are needed to obtain the same magnitude of change as in the initial sessions. This phenomenon consistently produces a curvilinear growth curve that is best normalized by a log-linear transformation of session number. A similar finding was obtained by Howard et al. (1996) in their patient profiling project.

An initial exploration of the current data set revealed that a similar curvilinear growth curve was produced by repeating OQ–45 measurements across sessions. To allow linear modeling of these growth curves, we first transformed them using a log-linear transformation by session number. The growth curves were then generated by use of PROC MIXED in the SAS System (SAS Institute, 1990) to craft a hierarchical linear model (HLM) that first aggregates data and then estimates the recovery curve for specific OQ–45 score bands. HLM provides a method of analyzing multiwave data as a function of nesting, meaning that one analyzes the data first by computing individual growth curves and then by investigating these as a function of a group (e.g., clinical setting or clinical population; for a complete discussion, see Arnold, 1992; Bryk & Raudenbush, 1992; Rogosa & Willett, 1985).

HLM establishes individual change by modeling within-person data points for the first level of analysis. The within-person estimates then become the dependent variable in the second level of analysis or in the between-persons analysis. HLM offers several unique features that allow the flexibility needed for analysis of naturally existing data. The number of observations per person need not be constant; the timing of the observations also need not be constant within or between individuals, and missing data can be accounted for using HLM (Speer & Greenbaum, 1995). HLM makes use of all the available data in estimating change rates based on several assumptions.[4] Each band in the analysis consisted of OQ–45 total score units, each representing approximately 2% of the available patients. This resulted in 50 distinct score bands of no fewer than 220 patients within each band along the continuum from 0 to 180.

Each score band was then modeled so that an estimated recovery curve could be calculated. Given a patient's intake score on the OQ–45, his or her actual or observed progress in therapy can be plotted in relation to expected progress on the basis of the progress of the cohort with similar intake scores. To identify progress that was unexpected, tolerance intervals were set around the expected recovery curve. These were two-tailed intervals set at the 68th and 80th percentiles, meaning that we would identify patients whose improvement was so slow as to be characteristic of the slowest 10–15% of those patients on whom the model was developed (see Finch, Lambert, & Schaalje, in press, for a full description of modeling statistics). These tolerance intervals were intentionally set to capture at least 10% of potential treatment failures, in accordance with prior research (Lambert & Bergin, 1994), which found that approximately 10% of patients deteriorate in therapy as compared with 5% of untreated participants. These estimates are further supported by meta-analyses that also showed roughly 10% of effect sizes are in a negative direction (Shapiro & Shapiro, 1982; Smith et al., 1980).

Results and Discussion

Space limitations do not permit the display of all 50 recovery curves, but two are presented for the purpose of illustration. Figure 3 presents the recovery curve for patients with an intake score of 95 or 96. This score is approximately one standard deviation above the mean of persons entering outpatient treatment (i.e., somewhat more disturbed than the average outpatient). As can be seen, the estimated slope of change for an individual who starts therapy at this level of disturbance decreases (improves) over time, with an average per session improvement of just under one point.

---

[4] Arnold (1992) described HLM as a regression of regressions to which the assumptions of linear regression apply. Additionally, the data must be hierarchical, with units nested within groups. The groups must have enough within and between groups subjects to provide sufficient degrees of freedom. The base data need to be reliable and valid because the lower units of analysis form the basis for higher levels of analysis. Willett (1989) provided an equation for determining the number of participants necessary for various situations. The *ns* in the present analysis far exceed these recommendations even when the data are broken down into 50 separate score-bands.
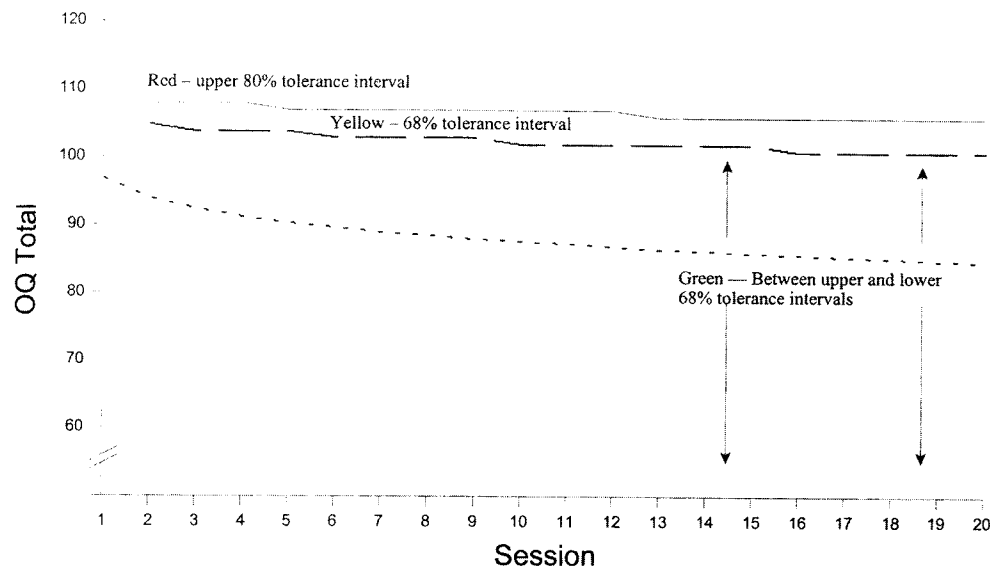
Figure 3. Expected recovery curve and prediction band for patients starting with an Outcome Questionnaire (OQ) total score of 97 or 98.

A patient is identified as an outlier if his or her score exceeds either of the confidence bands above the expected recovery curve.

Figure 4 provides the recovery curve for patients whose intake score on the OQ–45 is either 78 or 79, a score that is typical of patients who enter treatment (i.e., about the mean of outpatient samples). In this figure, we also plotted the progress of a patient from one of our samples. This patient stayed within the boundary of tolerance for the first two sessions of treatment, exceeded the boundary after three sessions of treatment, and returned within the boundary after six sessions of treatment.

The 50 recovery curves allow us to identify potential failure cases based on the patients' initial level of disturbance and the degree to which they vary from expected levels of improvement at any treatment session. This information can be provided to clinicians graphically or entered into a computer program that automatically alerts the clinician that the patient is a potential "treatment failure." The clinician can evaluate this information to see if it is consistent with his or her own observations and patient reports and then use it as seems appropriate.

Study 3: The Effects of Providing Feedback: Does Feedback on Patient Progress Improve Psychotherapy Outcomes?

The development of recovery curves with tolerance intervals naturally leads to using this information to improve clinical services by supplying feedback to providers. Such feedback, however, has rarely been the subject of study in psychological services, although it has a long history in medical practice (D. Davis, Thompson, Oxman, & Haynes, 1995). The essential element in clinical decision making is information. However, the literature on clinical decision making indicates that using the information to form accurate judgments is quite difficult (Brickman, Karver, & Schut, 1997; Rossi, Schuerman, & Budde, 1996). Dawes (1989) and Garb and Schramke (1996) suggested that clinicians may have less than optimal judgement because of a lack of feedback, making

it difficult to learn from experience. Thus, clinicians who do not receive systematic feedback from their environment may learn little about the impact of their decisions and actions. If a clinician does not accurately perceive the client's progress, he or she may assume incorrectly that no changes in treatment need to be made. Conversely, a clinician may change treatment tactics when a change is not needed. Even if perception is correct, clinicians may have difficulty obtaining follow-up information on a large portion of their patients, because this kind of data is seldom collected.

Lieberman, Yalom, and Miles (1973) were among the first to document the failure of providers to recognize casualties in encounter groups (in comparison with self and other-group-member nominations). Many therapists maintain a belief that patients get worse before they get better, thus allowing themselves to assimilate patient worsening as a necessary step in improvement rather than as a sign of harmful procedures (Canen & Lambert, 1999). This clinical lore exists despite research that has suggested that the best predictor of final outcome is an early positive response to treatment (Brown & Lambert, 1998; Haas et al., 1999) and evidence that early response is a common phenomenon in clinical trials research (e.g., Wilson, 1999). If providers are not making appropriate treatment decisions, then patients are less likely to achieve the best possible outcomes because they may not be receiving services that match their needs. Thus, valuable clinical resources may be consumed while high quality care is not received.

The continuous collection of outcome data and the use of a system to identify potential treatment failures may provide clinicians with the opportunity to improve outcomes by providing information that enables them to alter their behavior and decisions and generally to enhance the treatment process (Lambert & Brown, 1996). An important issue concerns what should constitute feedback to clinicians. Howard et al. (1996) illustrated the possible use of recovery curves that show expected and observed change with three separate cases, but they did not examine the effects of
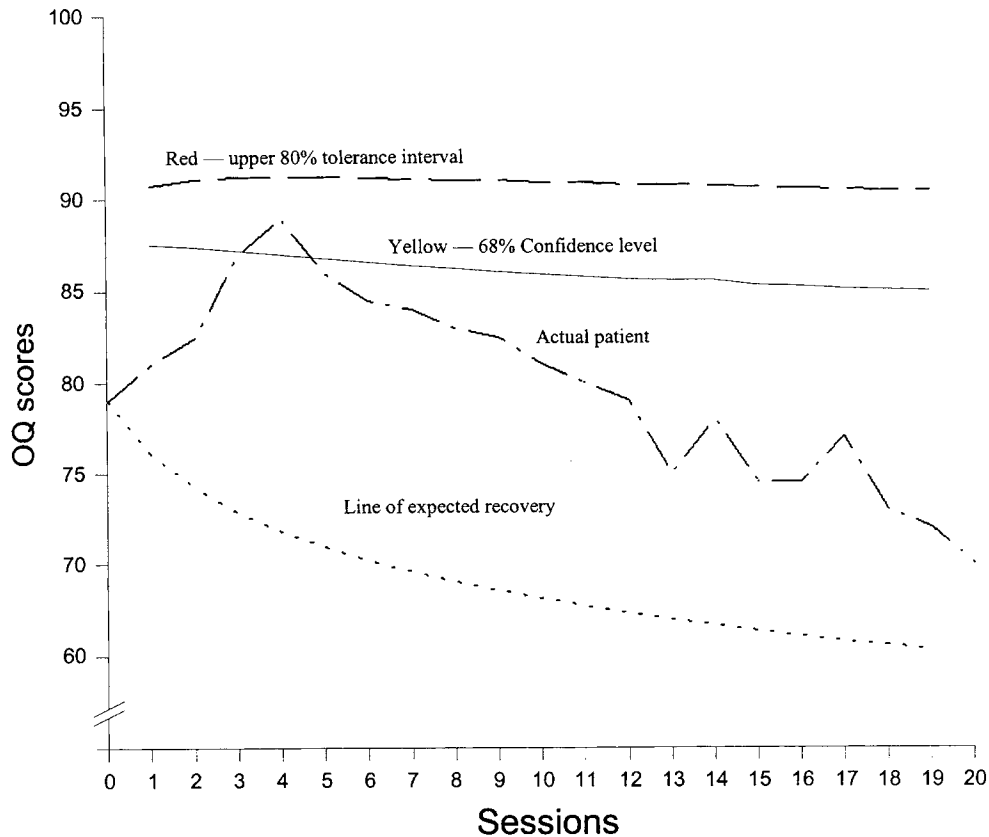
Figure 4. Expected recovery curve and tolerance band for patients starting with a score of 78 or 79 on the Outcome Questionnaire (OQ), with a specific patient's progress as an example.

feedback to clinicians on patient progress or provide information on what the clinician should do as a result of the feedback. Likewise, Kordy et al. (1999) have developed a feedback system, but it is typically applied after therapy has ended and during the follow-up period.

Kluger and DeNisi (1998) suggested that helpful feedback focuses the receiver on tasks rather than on themselves. In an attempt to enhance the mere presentation of recovery curves and prediction boundaries and to alert clinicians to actions they might take with cases that are not improving. Lambert et al. (2001) used color coding and decision suggestions along with patient graphs to test the effects of feedback on patient outcome. The research question was, does feedback on patient progress improve outcomes? The study assessed the effects of feedback on outcomes in an actual clinical setting under routine conditions. This type of design seemed appropriately suited for evaluating what actually would occur in a clinical setting rather than what could occur under conditions of tight experimental control. We wanted to create a feedback condition that could be easily used in routine clinical practice. The hypotheses tested were (a) patients of therapists who received feedback about patient progress would show better outcomes than would similar patients whose therapists did not receive feedback and (b) patients whose therapists received feedback would show better attendance (i.e. attendance consistent with cost-effective psychotherapy) than would similar patients whose therapists did not receive feedback.

Method

Procedures. To make the feedback as simple as possible, the estimated recovery curves were not provided directly to clinicians. Rather, the information from the curves was placed in a color-coded graph and a research assistant then placed a colored dot on the patient's files prior to subsequent therapy sessions. The colored dot was based on the patient's initial score, the number of treatment sessions completed, and the amount and direction of the change score at the last session. There were four color codes and corresponding feedback messages:

White-code feedback—the client is functioning in the normal range. Recommendation: Consider termination.

Green-code feedback—the rate of change the client is making is in the adequate range. Recommendation: No change in the treatment plan is recommended.

Yellow-code feedback—the rate of change the client is making is less than adequate. Recommendation: 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-code feedback—the client is not making the expected level of progress. Chances are he or she may drop out of treatment prematurely or have a negative treatment outcome. Recommendation: Steps should be taken to carefully review this case and decide on 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.

These color-coded messages corresponded with predictions regarding the patient's final treatment status. The white code indicated that a patient's score was below the cutoff for being in the dysfunctional population. The green code meant that the patient was making the expected amount of progress but had not yet achieved clinically significant change. The yellow code indicated the patient was deviating from the expected recovery curve, and there was reason to be concerned that a new direction was needed in regards to their treatment. The red code meant that there was more confidence that treatment progress suggested a poor final outcome, and a decision about a change in treatment was necessary.

Participants. The red- and yellow-coded clients combined were referred to as "signal" cases. In all, 607 clients were randomly assigned to the feedback condition or to a no-feedback condition in which the same therapists treated them but were not given a graph of the patient's progress or a color-coded dot. There were a total of 65 signal cases (34 experimental and 31 controls). (For a description of patients, see Lambert et al., 2001.)

Measure. The OQ-45, described in Study 1, was used as both the tracking and outcome measure.

Results and Discussion

The results suggested that the signal cases of therapists who received feedback improved more than the corresponding controls did, with 26% of the signal feedback cases and 16% of signal no-feedback cases achieving clinically significant or reliable change. In addition, the deterioration rate for signal patients whose therapist got feedback was 6%, whereas signal cases whose therapists did not receive feedback had a 23% deterioration rate. Patients of therapists who never received a red or yellow warning (white–green cases) improved to the same degree whether or not feedback was given. Attendance figures were also consistent with the experimental hypothesis. Signal feedback cases were given twice the number of sessions as no-feedback signal controls, whereas white–green feedback cases received significantly fewer sessions than white–green no-feedback cases. Thus, feedback resulted in more treatment for those patients whose progress was in doubt (with corresponding improvements in outcome) but in less treatment for those patients who were progressing. The net result was a reduction in treatment costs in conjunction with more efforts directed toward difficult cases who might have otherwise deteriorated.

Despite the positive effects of providing feedback, the results of the Lambert et al. (2001) study suggested the need for feedback related interventions that are more substantial but still consistent with what can be routinely offered in clinical practice. We judged the feedback as insufficient because many of the patients whose therapists got feedback did not achieve clinically significant change by the time they left therapy, and many remained symptomatic (albeit better off than the controls) despite the feedback. Because this study provided no mechanism for monitoring whether and how clinicians changed treatment in response to feedback, we are currently conducting a study of feedback that includes a method of monitoring therapist actions.

Discussion

Within the conceptual framework of patient-focused research with the goal of improving the effects of psychotherapy on patient outcomes (including symptomatic changes, interpersonal relations, social role performance, and quality of life), several advances have been made. We have operationalized outcome, defined a meaningful improvement, investigated the dose–effect relationship, identified expected trajectories of change based on initial levels of patient disturbance, and investigated the effects of providing feedback to clinicians about the status of patients in relation to their expected outcomes. We estimated that 21 sessions of psychotherapy are needed for 50% of patients to achieve clinically significant change. These results are limited by the difficulties of data collection in large samples, which are not part of an experimental protocol. Although the sample sizes in research studies that involve tight experimental controls seldom include more than 100 or 200 patients, we have been able to obtain data on over 10,000 patients. However, collection of data on such large numbers of patients precludes the use of such methodological necessities as defining and monitoring what kind of therapy is being offered or insuring the reliability of the diagnoses that are given. Patient diagnoses in the samples we have studied have unknown and even questionable reliability. Often the treatments offered have not been specified—let alone manual-guided, monitored, and rated for competence or conformity within any system of treatment. The presence or absence of medication referral or compliance and the use of collateral treatments is also often not a part of the databases we have had access to, thus it has been difficult to understand the possible effects of medication on estimates of the dose-response curve.

The number of settings from which we have collected data, their variability, and their geographical locations have provided an unusual opportunity to study change. Nevertheless, generalizability of our research may be limited by an overrepresentation of cases treated in EAP settings by therapists with limited training and in other ways be unrepresentative of routine practice. Examination of the data we do have suggests that the dose–response relationship is affected by treatment settings as well as by initial levels of severity. Research in which patient, therapist, and treatment variables are more vigorously defined and monitored will advance our understanding of the dose–effect relationship and of recovery curves within specific treatments for specific disorders. Such research may provide a more accurate basis for setting expectancies for ideal patient progress and providing feedback to clinicians about alternative treatments for patients that fail to respond to therapy as it has been offered to them. Such a task will require many studies over the next decade.

The development of recovery curves for patients with different intake scores on the OQ appeared to provide a sufficient foundation for providing feedback to clinicians about patient progress by identifying signal cases, patients for whom termination may be feasible, and patients for whom progress is satisfactory but incomplete. Several limitations can, however, be noted. The first is related to the fact that although the trend toward recovery is generally linear and most of those who ultimately recover are early responders to treatment, a patient's individual recovery trajectory shows variability. A patient may be identified as on track one week but off track another. In fact, the variability of a particular patient's scores (especially signal cases) may be diagnostic in that those patients who show the greatest fluctuations are typically the most unstable and vulnerable patients—vulnerable to environmental vicissitudes and so on. These are, perhaps, patients who are most likely to create an environment that is filled with untoward consequences, including poor social supports, limited financial resources, and unstable employment.

The estimated recovery curves are only a general guideline for predicting patient response in clinical practice and should be

treated as such. Their validity will need to be tested across an even wider range of patients and therapies. The recovery curves are limited by all the factors previously mentioned that affect estimates of the dose–response relationship and may simply identify patients who are especially difficult to treat with any known methods. In this regard, those who use the recovery curves to make decisions (such as insurance companies) should bear in mind that clinicians cannot be held solely responsible for patient progress, as it is determined by multiple sources in the patient and in the patient's environment.

Initial attempts to investigate the effects of providing feedback based on recovery curves were a partial success. As a result of giving clinicians feedback about signal cases (nonresponders), they kept those patients in therapy for longer periods of time and at termination those patients had, as a group, better outcomes than signal cases whose therapists did not receive feedback. Nevertheless, even the signal cases whose therapists got feedback showed relatively poor treatment response, with 75% rated as not changed or deteriorated when they left treatment. These findings suggested that the feedback intervention may have been inadequate to produce the desired effect and, therefore, it needs to be strengthened.

The effectiveness of feedback rests on two necessities: It must be timely and it must provide useful (i.e., action oriented) information. The timeliness of feedback would be easily solved if clinicians were more willing to administer, score, and graph patient results with readily accessible software. However, this task requires a major change in clinical practice patterns that is not likely to become a practice standard in the foreseeable future. We have depended on receptionists and research assistants to do the work of collecting and dispensing feedback information, but this has been a challenge at peak hours, at times when the patient is late, at times of emergency, and the like. Timely feedback in the context of routine practice requires all of those involved to be invested in its value. Implementation remains a limitation of this type of patient-focused research, but one that can be managed if those involved are committed and, especially, if automated methods are used to gather outcome data (e.g., telephonic, phone to fax, Web site).[5] Widespread replication of our results may increase the probability that clinicians will want to use outcome-monitoring procedures. However, further research on the effects of feedback will be needed before it will become an accepted standard like x-rays and blood pressure monitoring.

The feedback itself is another matter. Once implementation problems are overcome, it is still not certain how researchers in quality management can give therapists something they can use to improve their treatment of a specific individual: how to track alterations in treatment practices, and, thereby, help therapists alter potentially negative outcomes. How far should researchers go in trying to alter the practice of clinicians? Currently our research is aimed at developing a decision tree approach. If the patient is identified as a signal case, we will ask the therapist to go through a series of problem-solving tasks that lead to specific actions. The decision tree approach also provides an opportunity to test the effects of empirically supported therapies, with referrals for these treatments possibly increasing in the face of treatment failures that occur when they have not been the first-line treatment. Research should be directed toward developing expected recovery curves for treatments when empirically supported therapies are competently offered to appropriate patients.

Several limitations to our research efforts should be mentioned to make its boundaries clear and to foster related research endeavors. These limitations are presented in the order of the research outlined in this article. The operationalization of outcome through the use of a single self-report scale is a major limitation of the present research. Although the OQ is reliable, correlated highly with other self-report scales, and sensitive to change, no research has examined the extent to which weekly changes on the OQ are characteristic of change that would be found had other measures been used concurrently or in its place. Certainly the picture would be changed somewhat if the alternative measures focused on different constructs or different aspects of health and psychopathology. Hansen (1999), as well as Vermeersch et al. (2000), for example, have shown that the items of the OQ from the Symptomatic Distress scale change more rapidly than items that measure interpersonal functioning and social role performance. Within therapies that last less than 20 sessions, it can be seen that improvement in symptomatic states is approximately twice that observed with interpersonal difficulties and social role problems. Different scales and measurement methods place different emphasis on these factors and are likely to get differing results depending on the weighting given each factor within a scale. Research is needed to identify the degree of variance in estimating outcome that is a function of specific scales and measurement sources in relation to the length and intensity of intervention (e.g., Kopta, Howard, Lowry, & Beutler, 1994).

Another issue of importance is the consequence of using the same questionnaire repeatedly over brief periods of time. This effect has been studied on the OQ with a number of retest schedules. Durham et al. (1999), who used testing schedules of weekly, biweekly, monthly, and over 2-month administrations, found an average one point drop in scores at the second testing regardless of the time that had elapsed or the number of retests. They concluded that retesting is not a cumulative (i.e., an additive) problem and that within the limits examined (nine weekly tests) tracking patients on a weekly basis presented no more risks than did typical pretest, posttest designs. Nevertheless, this artifact needs further study and could vary considerably as a function of the time and energy that is required from patients. Researchers using other scales should be aware of this fact and keep assessment at reasonable levels.

The use of Jacobson and Truax's (1991) clinical significance and reliable change methodology or the use of a similar classification procedure is a fundamental characteristic of patient-focused research. The results of studies that have used the OQ suggest that the proportion of patients showing clinically significant and reliable change are essentially equivalent to those in other outcome studies conducted by other researchers using a variety of different scales with various samples of patients (see for example, R. Davis, Olmstead, & Rockert, 1990; Jacobson, Wilson, & Tupper, 1988; and Wollersheim & Wilson, 1991). In the only published study of concurrent validity of clinical significance cutoffs with the OQ,

---

[5] The OQ is part of a larger system that includes a Child–Adolescent scale for children 4–18 years of age (The Youth Outcome Questionnaire, Burlingame et al., 1996), an experimental Severe-Outcome Questionnaire for highly disturbed patients and additional screening instruments. These are available in software versions, with built-in graphing and feedback capabilities.

Lunnen and Ogles (1998) found those who were classified as reliably changed were significantly more satisfied with treatment. Nielsen et al. (2001) reported similar findings with satisfaction and additional data that suggested those classified as clinically significantly improved also rated themselves as having benefitted more from psychotherapy than those classified as no changers.

One unanticipated problem with clinical significance methodology in actual clinical practice is the presence of a sizable minority of patients (10–30%) who start treatment in the functional range (below the cutoff score). Some of these patients begin treatment with scores even below the mean of nonpatient samples. This phenomenon does not occur in clinical trials research because low-scoring patients are usually screened out prior to acceptance into the study, and therefore, the range of patients studied in clinical trials may be ideal for finding positive outcomes. Naturalistic studies of change do not screen out more guarded patients, more healthy patients, or both and, thereby, present a greater challenge for finding significant treatment effects. This phenomenon is probably present in many clinical settings and is not restricted to those who use the OQ to measure outcome, but it provides one reason to expect lower patient recovery rates from research on actual samples from clinics than from participants in clinical trials. In clinical settings the researcher is presented with a choice—study and report change only for those patients who begin treatment in the dysfunctional range, or use a lower standard for improvement (e.g., RCI) and analyze all patients who enter treatment regardless of their initial status (see also, Kendall et al., 1999). The consequence of looking at change across a wide range of patients requires outcome comparisons between clinics and therapists or similar groupings be adjusted for the number of patients within a comparison group who begin treatment with scores in the functional range at pretesting. Case-mix adjustments can be made on the basis of initial scores, and these can then be used to develop expectations before comparing clinic and provider outcomes (see, e.g., Brown & Lambert, 1998, June).

Another concern with adopting a lesser criterion, such as reliably improved rather than clinically significantly improved, is that the standards of care and expected outcomes from treatment will be lowered. As can be seen from the results of the survival analysis presented in this article, greatly different treatment durations are predicted depending on what level of improvement is being considered (7 sessions vs. 21 sessions for 50% improvement rate). Alternatively, use of a conservative criterion for recovery, such as clinical significance, may be unrealistic in some circumstances. Clearly, more research needs to be conducted to establish realistic and acceptable outcome criteria for specific cases, including such variables as *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) diagnoses, patient support systems, concurrent medical conditions, environmental stressors, and other patient characteristics. The issue of normal functioning is a critical issue (Kendall et al., 1999) that needs further exploration.

The possible contributions of patient-based research and tracking are just beginning to be explored in routine practice. Our program of research is seen as complementary to carefully conducted clinical trials, but it and related methodologies may also take a prominent position within empirical efforts to understand ways of helping. It embodies procedures that can be implemented in routine clinical practice and invites the curious scientist-practitioner to contribute research from personal clinical practice.

It offers multiple possibilities for influencing the manner in which managed care companies make treatment decisions and may provide an important tool for supervision of trainees. We invite the interested clinician to engage in patient-focused research, as it appears to be a research activity that is well-suited to the scientist-practitioner.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, DC: Author.

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

Arnold, C. L. (1992). Methods, plainly speaking: An introduction to hierarchical linear models. *Measurement and Evaluation in Counseling and Development, 25*, 58–90.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychology, 4*, 53–63.

Brickman, L., Karver, M. S., & Schut, L. J. A. (1997). Clinical reliability and accuracy in judging appropriate level of care. *Journal of Consulting and Clinical Psychology, 65*, 515–520.

Brown, G. S., & Lambert, M. J. (1998, June). *Tracking patient progress: Decision making for cases who are not benefitting from therapy*. Paper presented at the 29th annual meeting of the Society for Psychotherapy Research, Snowbird, UT.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Burlingame, G. M., Wells, M. G., Hoag, M., Hope, C., Nebeker, S., Konkel, K., McCollam, P., Peterson, G., Lambert, M. J., Latkowski, M., Ferre, R., & Reisinger, C. (1996). *Administration and scoring manual for the Youth Outcome Questionnaire (Y-OQ.1)*. Wilmington, DE: American Professional Credentialing Services.

Canen, E. L., & Lambert, M. J. (1999, May). *The incidence of patterned deterioration before stable improvement in psychotherapy*. Poster presented at the Western Psychological Association. Irvine, CA.

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308.

Collette, D. (1994). *Modeling survival data in medical research*. New York: Chapman & Hall.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. New York: Chapman & Hall.

Davis, D., Thompson, M. A., Oxman, A. D., & Haynes, B. (1995). Changing physician performance: A systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association, 274*, 700–705.

Davis, R., Olmsted, M. P., & Rockert, W. (1990). Brief group psychoeducation for bulimia nervosa: Assessing the clinical significance of change. *Journal of Consulting and Clinical Psychology, 58*, 882–885.

Dawes, R. M. (1989). Experience and validity of clinical judgement: The illusory correlation. *Behavioral Sciences and the Law, 1*, 457–467.

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

Durham, C., Burlingame, G. M., Schaalje, B., & Lambert, M. J. (1999). *Outcome Questionnaire: Repeated administrations, mechanical responding, and social desirability*. Manuscript submitted for publication. Brigham Young University, Provo, UT.

Feinstein, A. R. (1996). *Multivariable analysis: An introduction*. New Haven, CT: Yale University Press.

Finch, A. E., Lambert, M. J., Schaalje, B. (in press). Psychotherapy quality control: The statistical generation of recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy*.

Froyd, J. E., Lambert, M. J., & Froyd, E. (1996). A review of practices of

psychotherapy outcome measurement. *Journal of Mental Health, 5*, 11–15.

Garb, H. B., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analysis. *Psychological Bulletin, 120*, 140–153.

Haas, E., Hill, R., Lambert, M. J., & Morrell, B. (1999). *Unraveling the early response mystery: Do early responders maintain treatment gains?* Unpublished manuscript.

Hansen, N. B. (1999). *An overview of longitudinal data analysis methodologies applied to the dose response relationship in psychotherapy outcome research.* Unpublished manuscript.

Hansen, N. B., & Lambert, M. J. (1996). Clinical significance: An overview of methods. *Journal of Mental Health, 5*, 17–24.

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 Personality and Social Psychology, 61*, 68–79.

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 patient progress. *American Psychologist, 51*, 1059–1064.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352.

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.

Jacobson, N. S., Wilson, L., & Tupper, C. (1988). The clinical significance of treatment gains resulting from exposure-based interventions for agoraphobia: A reanalysis of outcome data. *Behavior Therapy, 19*, 248–251.

Jorm, A. F., Duncan-Jones, P., & Scott, R. (1989). An analysis of the re-test artifact in longitudinal studies of psychiatric symptoms and personality. *Psychological Medicine, 19*, 487–493.

Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1996). How much therapy is really enough?: A session-by-session analysis of the psychotherapy dose-effect relationship. *Journal of Psychotherapy Practice and Research, 5*, 1–22.

Kazdin, A. E. (1993). Evaluation in clinical practice: Clinically sensitive and systematic methods of treatment delivery. *Behavior Therapy, 24*, 11–45.

Kendall, P. C. (1999). Clinical significance. *Journal of Counseling and Clinical Psychology, 67*, 383–384.

Kendall, P. C., Marr-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.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science, 7*, 67–72.

Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology, 62*, 1009–1016.

Kordy, H., Richard, M., Herrmann, A., Murphy, F., Treasure, J., & Charpentier, P. (1999). A computer assisted eating disorder specific quality management system: EQUAL-TREAT. *European Eating Disorders Review, 7*, 239–258.

Lambert, M. J. (1983). Introduction to assessment of psychotherapy outcome: Historical perspective and current issues. In M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 3–32). New York: Wiley.

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143–189). New York: Wiley.

Lambert, M. J., & Brown, G. S. (1996). Data-based management for tracking outcome in private practice. *Clinical Psychology: Science and Practice, 3*, 172–178.

Lambert, M. J., Burlingame, G. L., 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., & Finch, A. E. (1999). The Outcome Questionnaire. In M. E. Maruish (Ed.) *The use of psychological testing for treatment planning and outcome assessment* (2nd ed. pp. 831–870). Mahwah, NJ: Earlbaum.

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, DE: American Professional Credentialing Services.

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72–113), New York: Wiley.

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 patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11*, 49–68.

Lieberman, M. A., Yalom, I. D., & Miles, M. B. (1973). *Encounter groups: First facts.* New York: Basic Books.

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

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

Nielsen, S. L., Smart, D. W., Isakson, R., Worthen, V., Lambert, M. J., Shively, A. P., Richard, M., Jackson, A., Schaalje, B., Okiishi, J., & Hansen, N. B. (2001). *Deciding effectiveness: Treatment outcome and satisfaction patterns at a University Counseling center.* Brigham Young University.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrica, 50*, 203–228.

Rossi, P. H., Schuerman, J. R., & Budde, S. (1996). *Understanding child maltreatment decisions and those who make them. Final report of the Understanding Placement Decisions in Child Welfare study.* Chopin Hall Center for Children, University of Chicago, IL.

SAS Institute Inc. (1990). *SAS-stat users guide. Version 6* (4th ed., Vol. 2). Cary, NC: Author.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50*, 965–974.

Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92*, 581–604.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.

Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63*, 1044–1048.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (Form Y).* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The State-Trait Anxiety Inventory Self-Evaluation Questionnaire.* Palo Alto, CA: Consulting Psychologists Press.

Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes: With special reference to negative effects in psychotherapy. *American Psychologist, 32*, 187–196.

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (1997). *Measuring*

patient changes in mood, anxiety, and personality disorders: Toward a core battery. Washington DC: American Psychological Association.

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology, 48*, 285–290.

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. (2000). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74*, 242–261.

Waskow, I. E., & Parloff, M. B. (1975). *Psychotherapy change measures.* (DHEW Publication No. ADM No. 74–1207). Washington, DC: U.S. Government Printing Office.

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

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49*, 587–602.

Wilson, G. T. (1999). Rapid response to cognitive behavior therapy. *Clinical Psychology: Science and Practice, 6*, 289–292.

Wollersheim, J. P., & Wilson, G. L. (1991). Group treatment of unipolar depression: A comparison of coping, supportive, bibliotherapy, and delayed treatment groups. *Professional Psychology: Research and Practice, 22*, 496–502.

Zung, W. W. (1971). A rating instrument for anxiety disorders. *Psychosomatics, 6*, 371–379.

AMERICAN PSYCHOLOGICAL ASSOCIATION
SUBSCRIPTION CLAIMS INFORMATION

Today's Date:_____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do NOT duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION

MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL)

ADDRESS

DATE YOUR ORDER WAS MAILED (OR PHONED)

_____PREPAID _____CHECK _____CHARGE
CHECK/CARD CLEARED DATE:_____

CITY   STATE/COUNTRY   ZIP

(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.)

YOUR NAME AND PHONE NUMBER

ISSUES: ____ MISSING ____ DAMAGED

TITLE	VOLUME OR YEAR	NUMBER OR MONTH

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.

——————————— **(TO BE FILLED OUT BY APA STAFF)** ———————————

DATE RECEIVED: _____   DATE OF ACTION: _____
ACTION TAKEN: _____   INV. NO. & DATE: _____
STAFF NAME: _____   LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.