

Naturalistic Examination of a Training Clinic: Is There a Relationship Between Therapist Perception and Client Self-Report of Treatment Outcomes?

Scott H. Waltman, Katherine H. Rex, and Alyson Williams
Pacific University

Numerous psychometrically sound measures of psychotherapy outcome have been developed; however, few clinicians administer such tools, and when they do so, they rarely utilize these results to inform treatment. Moreover, studies have reported a discrepancy between therapists' judgment and clients' self-report of treatment outcome. The current study aimed to further investigate the relationship between therapists' judgment and clients' self-report of treatment outcome in a psychological training clinic, using the criteria of reliable change (RC) and clinically significant change (CSC). One hundred and forty-four clients receiving therapy at a psychological training clinic in the Pacific Northwest were administered the Outcome Questionnaire 45.2 (OQ) at the beginning and end of treatment. Study participants were primarily Caucasian (88%) and female (56%), with a mean age of 32 years. Changes in outcome scores were compared to therapist judgments regarding treatment effectiveness. Results indicated a significant relationship between therapists' judgment of treatment outcome and whether clients achieved RC or CSC. A large effect size ($\Phi = .56$) was found between CSC and therapist judgment whereas a medium effect size ($\Phi = .32$) was found between RC and therapist judgment. These results suggest that there is lack of agreement between the RC and CSC criteria, and that outcome measure alone cannot account for therapist judgment about treatment outcomes. Suggestions are made for maximizing the accuracy and practicality of client outcome assessment.

A number of psychometrically sound instruments assessing psychotherapy outcome exist; however, the majority of clinicians do not use such tools (Lambert & Hawkins, 2004). Moreover, even when clinicians do administer measures that track therapeutic outcomes, many do not utilize the results to inform treatment (Garland, Kruse, & Aarons, 2003). Research findings show discrepancies between therapists' judgment and clients' self-report of treatment outcomes (Swift, Callahan, & Levine, 2009); it appears that many clinicians evaluate treatment outcomes based on their own perceptions, rather than based on standardized measures intended to accurately assess clients' clinical symptoms and functioning. This raises important clinical questions, given that the accurate evaluation of treatment outcomes informs whether treatment was effective and is a critical component of improving treatment (American Psychological Association [APA], 2006).

In a recent study of a psychological training clinic, Swift and colleagues (2009) compared two methods of evaluating treatment outcome: therapists' judgment of treatment outcome and clients' scores on standardized outcome tracking measures. The authors reported a low level of agreement

between these two methods, suggesting that therapists who rely on their judgment alone are more likely to describe treatment as being successful compared to those using collateral information from objective outcome measures. Drawing on these research findings by Swift and colleagues (2009), the current study aims to investigate whether there is a lack of agreement between therapists' judgment of treatment outcomes and clients' self-reported changes in other psychological training clinics. The present study uses a naturalistic design (i.e., absence of experimental manipulations) that provides a realistic evaluation of how a training clinic functions.

Clinicians' Use of Outcome Measurements

Given that therapists are subject to the same judgment errors as all other humans (Ruscio, 2007), numerous treatment outcome measures have been developed to help clinicians assess the effectiveness and quality of psychosocial interventions (Garland et al., 2003; Hatfield & Ogles, 2004; Mours, Campbell, Gathercoal, & Peterson, 2009). Usually, these are client self-report measures that assess on a weekly timeframe the impact of presenting problems on functioning. However, it is estimated that less than one-third of all clinicians use outcome tracking measures in their clinical practice (Lambert & Hawkins, 2004). Underuse of outcome measures was reported by Garland and colleagues (2003) in a survey of clinicians who participated in a state program that mandated the use of outcome measures. The authors found that as many as 92% of the clinicians had never referenced standardized outcome measures in their practice (Garland et

Scott H. Waltman, M.S., Katherine H. Rex, M.S., and Alyson Williams, Ph.D., School of Professional Psychology, Pacific University.

Scott Waltman would like to thank Dr. Alyson Williams for her patience, guidance, and assistance in that endeavor.

Correspondence concerning this article should be addressed to Scott Waltman, School of Professional Psychology, Pacific University, 190 SE 8th Ave. Hillsboro, OR 97123. Email: walt2155@pacificu.edu.

al., 2003). In a different survey of a national sample of licensed psychologists, only 37% of respondents reported using some sort of standardized assessment in their practice (Hatfield & Ogles, 2004).

It is necessary to examine therapist attitudes about outcome measures, as this may illuminate why these measures are so infrequently utilized. In the study by Garland and colleagues (2003), clinicians reported that outcome measures are cumbersome or intrusive, can be expensive to practitioners, are burdensome to clients, and do not provide new information (Garland et al., 2003). Rather than using outcome measures that have been shown to lead to improvements in treatment outcomes (Reese, Norsworthy, & Rowlands, 2009) and increased cost-effectiveness (Slade et al., 2006), many of these clinicians reported using “real world functional indicators” (e.g., school grades), clinical intuition, and client-reported satisfaction in order to evaluate treatment effectiveness (Garland et al., 2003). It is clear that some therapists may view their clinical judgment as being more accurate or useful than the information obtained from outcome measures. It is unclear, however, whether this trend also exists in a training clinic setting, as there is limited research on the relationship between student clinician and licensed professional judgment. Understanding the relationship between therapist judgment about treatment outcomes and client changes in scores on outcome measures will be informative as to whether some student therapists view their developing clinical judgment as being more accurate and useful than the information generated from standardized outcome measures.

Measurement of Reliable Change and Clinically Significant Change

The regular use of standardized outcome measures is a recommended routine practice for tracking client progress (Hatfield & Ogles, 2004). Advantages of using these measures include their established reliability and validity and their sensitivity in determining clinically meaningful changes in client functioning. The concepts of reliable change (RC) and clinically significant change (CSC) were introduced as a statistical approach of determining when meaningful change has occurred (Jacobson, Follette, & Revenstorf, 1984). RC indicates that a change in scores on a measure is sufficiently larger than the standard error of the difference between the two scores, and thus is not due to chance (Jacobson & Truax, 1991). CSC is more a more stringent criterion for determining treatment outcome than RC, as CSC requires that RC has occurred and that a client’s score on an outcome measure has moved from the dysfunctional to the functional range (Jacobson & Truax, 1991). In order for clients to obtain CSC, they must begin treatment in the clinical range of functioning and move into the nonclinical range of functioning (Wise, 2004). The use of the criteria of RC and CSC has advantages over other statistical methods. A traditional statistical method of comparing different groups is comparing group means (e.g., analysis of variance; ANOVA), which has two major limitations (Jacobson, Roberts, Berns, &

McGlinchey, 1999). First, comparing group means provides little information about the variability within those groups (i.e., the proportion of group members who improved or recovered). Second, a large effect size does not mean that the effect was clinically meaningful (Jacobson et al., 1999).

The criteria of RC and CSC have been used to evaluate the effectiveness of treatment in psychological training clinics. Callahan and Hynan (2005), for example, examined the treatment outcomes of 61 clients who were treated in a psychological training clinic by comparing clients’ changes in scores on a routinely administered measure, the Outcome Questionnaire 45.2 (OQ; Lambert et al., 2004), from the beginning to the end of treatment. They determined that over the course of treatment, 18% of participants achieved CSC, 33% demonstrated RC, and 67% showed no RC (Callahan & Hynan, 2005). This study illustrated that fewer clients achieve CSC than RC, which is to be expected as CSC is a more stringent criterion to meet. In another study of a training clinic, Swift and colleagues (2009) found a CSC rate of 23% and a RC rate of 37%. The authors concluded that therapists often labeled treatment as successful when CSC had not occurred, and sometimes therapists labeled treatment as unsuccessful when CSC had happened (Swift et al., 2009).

Therapist Judgment of Treatment Outcome

Borrowing from the premature termination literature, it has been suggested that therapist judgment may be the best method of determining when a client has prematurely terminated (Wierzbicki & Pekarik, 1993); however, a problem with using therapist judgment to determine outcome is the issue of reliability (Wierzbicki & Pekarik, 1993). Not all therapists may decide a client has prematurely terminated from therapy based on the same criteria (Hatchett & Park, 2003; Swift et al., 2009), and research has found that therapists are likely to use self-serving attributions when explaining why a client prematurely terminated from treatment (Murdock, Edwards, & Murdock, 2010). Additionally, therapists’ and client’s descriptions of treatment termination may differ (Hunsley, Aubry, Verstervelt, & Vito 1999). In fact, it has been proposed that therapist perception of progress and client self-report of satisfaction have a low level of agreement (Mours et al., 2009). Several studies have failed to find a significant relationship between client satisfaction and symptomatic improvement (Lambert, Okiishi, Finch, & Johnson, 1998; Lunnen & Olges, 1998; Lunnen, Ogle, & Pappas, 2008; Pekarik & Wolff, 1996), and yet, multiple surveys of practicing clinicians have found that many clinicians are likely to rely on their own intuition and the client’s self-report to determine if treatment has been successful (Garland et al., 2003; Mours et al., 2009).

Given potential discrepancies between therapists’ and clients’ reports, treatment evaluation based on therapists’ judgment alone may be insufficient. Research has demonstrated that therapists’ judgments about treatment outcome are more likely to be accurate only when it is positive (Hunsley et al., 1999). This may be due to biases in

TRAINING CLINIC EXAMINATION

the judgment of the clinicians. Specifically, therapists may be more likely to attend to positive information about treatment outcomes than to negative information (Ruscio, 2007). These biases in therapist judgment are problematic, especially since it is likely that they decrease therapists' ability to recognize the occurrence of a negative treatment outcome.

The current study is an investigation of the relationship between clients' changes in Outcome Questionnaire-45.2 scores (OQ; Lambert et al., 2004) and student therapist judgment about treatment success. This study is particularly relevant because the student clinicians had access to the OQ data as they rated the successfulness of treatment. Thus, the level of agreement between therapist judgment and outcome scores may be interpreted as an indicator of how influential outcome measurement is to student therapist judgment. The following research questions guided the present examination (a) how often do clients achieve RC and CSC?; and (b) what is the relationship between therapist judgment of treatment outcome and client changes in OQ scores? It was hypothesized that therapist judgment about treatment outcomes would be related to whether RC and CSC occur on client OQ scores.

Method

Participants

Clients. Archival data were used from 144 clients receiving therapy during the 2007-2008 academic year at a university-based psychological training clinic located on a small campus in a downtown urban area of the Pacific Northwest. This training clinic provides outpatient services to the general public. The clinic operates on a sliding fee scale, with session fees ranging from \$20 to \$85 depending on client income level. The clinic additionally offers a discounted rate of \$20 per session for college students. Although data concerning socioeconomic status were not collected as a part of this study, the clients receiving therapy from this clinic are often unemployed or *working poor* (i.e., underemployed) (Brooks, 2007). Clients may receive individual or couples therapy, and may present with a variety of presenting concerns. Typical presenting problems include depressive symptoms, anxiety symptoms, and relationship concerns. Individuals interested in treatment participate in a standardized screening procedure. Those who endorse particular risk factors, such as active suicidal ideation or excessive substance use concerns, are referred out to treatment settings that may provide higher levels of care.

Of the 144 clients, 31% ($n = 45$) were excluded from analysis due to either missing OQ test score data or the client having only attended an intake session. Clients who were identified as only attending an intake session were not included in this evaluation, as it has been suggested that clients who did not return for therapy after the intake did not start treatment (Hatchett & Park, 2003). Furthermore, at least two administrations of an outcome measure are required to calculate RC and CSC. Of the remaining clients, demographic data was available for 67% ($n = 66$). This

sample of 66 clients had more women (56.1%) than men (43.9%), had an average age of 32 years ($SD = 8.87$), with a range of 17 to 57 years. In regards to ethnicity, 87.9% of clients were Caucasian, 4.5% were of Asian descent, 1.5% were African-American, 4.5% were multi-ethnic, and 1.5% were unknown. Inclusion in the current study depended upon the presence of both a therapist rating of the successfulness of treatment and test scores from the first and last administrations of the OQ.

A Pearson's Chi-Square (χ^2) was utilized to determine if there was a relationship between therapist description of treatment outcome and whether or not demographic data were available. Additionally, an ANOVA was used to evaluate whether group differences existed in the average amount of change in OQ scores depending on the availability of demographic data. No significant differences were found in these analyses, which suggest that it was appropriate to include clients with missing demographic information in further analyses.

Therapists. Clients received therapy from doctoral students who were enrolled in a clinical psychology program and supervised by licensed clinical or counseling psychologists. These clinicians were trained in treatment for presenting problems ranging from anxiety and depression to relationship problems. Supervisors were licensed psychologists who were either faculty members or psychologists from the community with private practices. This training program follows a practitioner-scholar model, and student therapists were either in their second, third, or fourth year in the program. Three of these therapists were pre-doctoral interns who were either in their fifth or sixth year of training. In total, approximately 45 therapists recorded data for this study. At the time this data were collected, information about individual therapists was not collected due to concerns about student privacy and confidentiality. This will be discussed later as a limitation of the study.

Measures

Outcome Questionnaire. The OQ is a 45-item self-report measure that assesses general psychological distress (Lambert et al., 2004). An overall score may be calculated after administration, along with subscale scores related to symptom distress, interpersonal functioning, and social role. The OQ is typically administered to every adult client at every session in this training clinic. For this study, only the first and last administrations of the OQ were analyzed. The average number of sessions completed by clients was 13. Other studies have used the OQ to measure RC and CSC, and to identify when individual treatment has been successful (Callahan & Hynan, 2005; Callahan, Swift, & Hynan, 2006). On the OQ, an overall score above 63 is indicative of symptoms in the clinical range (Anderson & Lambert, 2001). The OQ manual states that no gender differences exist between male and female scores. The manual also specifies that no significant differences have been found between the total scores of Caucasians, Hispanics, and African-Americans. The OQ has adequate psychometric properties,

with a test-retest reliability of .87 and internal consistency of .93 (Lambert et al., 2004). The internal consistency for this sample could not be calculated, as only the total OQ scores from the first and last administration were analyzed in this study.

Closing File Outcome Tracking Form. The Closing File Outcome Tracking Form (interested readers may contact the corresponding author for a copy; Brown, Williams, Waltman, & Sutton, 2010) was created for use in this particular training clinic. It is a brief form that is completed by the client’s assigned therapist as a part of the regular file closing procedure. This form contains information about diagnosis, client scores on the OQ, therapist description of therapy outcome, therapy duration, and demographic information. For this study, only a subset of information from the form was analyzed: data related to OQ scores from the first and last administration of the measure, therapist description of the outcome of therapy, and demographic information.

Procedure

Clients were informed that their demographic and treatment related information may be used for research purposes during the informed consent process at the beginning of therapy. Approval for the study was obtained from Pacific University’s Institutional Review Board (IRB). Data obtained from the Closing File Outcome Tracking Form were de-identified and coded into a database. Data were coded into the following three variables: therapist judgment, reliable change, and clinically significant change. Pearson’s Chi-Square (χ^2) was utilized to determine if there was a relationship between therapist description of treatment outcome and client changes in OQ scores. A phi correlation (Φ) was conducted to measure the relationship between changes in OQ scores and therapist rating of treatment successes.

Therapist judgment. The information about therapist judgment of the treatment outcome was recorded on the Closing File Outcome Tracking Form by each student therapist. Therapist judgment was coded as it was recorded on the form: “successful completion of treatment,” “substantial progress without successful completion of

treatment,” “incomplete or moderate progress,” or “no progress.”

Reliable change and clinically significant change. Based on the recommendations of the test publisher (Lambert et al., 2004), a decrease in total OQ scores of 14 or more was coded as “reliable change.” A change in total OQ scores of 13 or less was coded as “no reliable change.” Clinically significant change requires that RC has occurred and that a predetermined cut-off point has been crossed. A drop in total OQ score of 14 or more, and a decrease from a total score of above 64, to a total score below 63, was coded as “clinically significant change.” A failure to attain a drop in OQ scores of at least 14 points or a failure to cross the cutoff point of 63 was coded as “no clinically significant change.” The 47 clients who began treatment with OQ scores below 64 were unable to achieve CSC and were thus excluded from the examination of the relationship between therapist judgment and CSC.

Results

Descriptive Statistics

Descriptive statistics were used to compute the frequency of RC, CSC, and therapist description of treatment outcomes. Overall, from the beginning to the end of treatment, the mean change in total OQ score for the entire sample was a 9.67 point decrease. Only 35.4% of clients achieved RC. In determining CSC, it was found that only 18.2% of the clients had achieved CSC. It is worth noting that 47.5% of the clients began treatment with an OQ score in the nonclinical range, so they could not achieve CSC; of the clients who began treatment in the clinical range of functioning, 34.6% achieved CSC. It was found that therapists described 10.1% of treatment outcomes as “no progress,” 33.3% of treatment outcomes as “incomplete or moderate progress,” 21.2% of treatment outcomes as “substantial progress without successful completion of treatment,” and 35.4% of treatment outcomes as “successful completion of treatment.” Table 1 provides the percentages of treatment outcomes that achieved RC or CSC and the percentage of treatment outcomes according to therapist description.

Table 1

Comparison of Methods for Determining Client Treatment Outcome

Therapist Judgment of Treatment Outcome	Total ^a (%)	Reliable Change		Clinically Significant Change ^b	
		Yes (%)	No (%)	Yes (%)	No (%)
Successful Completion	35.4	18.2	17.2	8.1	27.3
Substantial Progress	21.2	9.1	12.1	6.1	15.2
Incomplete or Moderate Progress	33.3	7.1	26.3	4.0	29.3
No Progress	10.1	1.0	9.1	0	10.1
Total		35.4	64.6	18.2	81.8

^an = 99. ^b47.5% of clients began treatment with OQ scores in the nonclinical range, and subsequently could not achieve clinically significant change.

TRAINING CLINIC EXAMINATION

Therapist Judgment, Reliable Change, and Clinically Significant Change

In order to evaluate if therapist judgment was related to RC, a two-way contingency table analysis was conducted. The two variables were therapist judgment with four levels (i.e., “successful completion of treatment,” “substantial progress without successful completion of treatment,” “incomplete or moderate progress,” or “no progress”) and RC with two levels (i.e., RC and no RC). The two variables, therapist judgment and reliable change were found to be significantly related, $\chi^2(3, N = 99) = 10.18, p = .017, \Phi = .321$, consistent with a medium effect size (Cohen, 1988). These results support the conclusion that therapist judgment of treatment outcome is related to whether RC occurs.

In order to determine if therapist judgment was related to whether CSC occurred, a second two-way contingency table analysis was completed. The two variables were therapist judgment with four levels (i.e., “successful completion of treatment,” “substantial progress without successful completion of treatment,” “incomplete or moderate progress,” or “no progress”) and CSC with two levels (i.e., CSC and no CSC). The two variables of therapist judgment and CSC were found to be significantly related, Pearson $\chi^2(3, N = 52) = 16.20, p = .001, \Phi = .558$, consistent with a large effect size (Cohen, 1988). These results support the conclusion that a relationship exists between therapist judgment of treatment outcome and whether CSC occurs.

Post-Hoc Analysis

After it was found that a sizeable portion of the sample was not capable of achieving CSC, the researchers decided to conduct an additional post-hoc analysis, a one-way analysis of variance (ANOVA), to evaluate the relationship between therapist judgment of treatment outcomes and the change in OQ scores from the first to last administration of the measure. It was hypothesized that significant differences would be found between the average amount of change in total OQ scores and therapists’ judgment of treatment outcomes. The independent variable was therapist judgment and the dependent variable was the change in OQ scores from first to last administration of the measure. The results of the ANOVA were significant, $F(3, 95) = 4.41, p = .001$. The strength of the relationship between therapist judgment and the change in OQ scores, assessed by η^2 , was moderate, with the therapist judgment accounting for 12% of the variance in the dependent variable.

Follow up tests were conducted to evaluate pairwise differences among the means. Based on Levene’s test of the homogeneity of variance, it can be assumed that the variances were homogenous among the three groups. Post hoc comparisons were therefore conducted using Tukey’s HSD

There was a significant difference in the means between the group that was judged to have successfully completed treatment ($M = -18.80, SD = 23.13$) and the group that was judged to have made no progress ($M = 1.50, SD = 10.64$), as well as between the group that was judged to have

Table 2

Mean Change in Total OQ score by Therapist Judgment of Treatment Outcome

Therapist Judgment	Total Score	M (SD)	95% CI ^a
Successful Completion	35	-18.80 (23.13)	-25.87, -11.73
Substantial Progress	21	-10.95 (20.87)	-20.08, -1.82
Incomplete or Moderate Progress	33	-2.55 (21.12)	-9.83, 4.73
No Progress	10	1.50 (10.64)	-11.7, 14.73
Total	99	-9.67 (22.15)	

^aCI = confidence interval.

successfully completed treatment and the group that was judged to have incomplete or moderate progress ($M = -2.55, SD = 21.12$). No significant difference was found between the substantial progress without completion of treatment group ($M = -10.95, SD = 20.87$), and the no progress group. There was not a significant difference found between the other pairwise comparisons.

Discussion

The current study investigated the relationship between therapists’ judgment and clients’ self-report of treatment outcome; the latter was operationalized as changes in outcome scores using both RC and CSC criteria. The study is of particular relevance to those interested in understanding how the use of outcome measures affects therapist judgment, as the therapists had access to the OQ data when they were rating the overall successfulness of treatment.

The results of this study demonstrate that a relationship exists between therapists’ judgment of treatment outcome and whether a client achieves RC or CSC on an outcome measure. Additionally, a large effect size was found between CSC and therapist judgment, indicating a strong relationship between these two variables. These findings suggest that training therapists often labeled treatment as successful when CSC did occur, which is commensurate with findings from a prior study by Swift and colleagues (2009). Although these results denote that therapist judgment and RC are related, the results also suggest that there is a low level of agreement between these two methods of determining treatment outcomes as 17% of therapists labeled treatment as successful when RC had not occurred, and 8% labeled treatment as unsuccessful when RC had occurred. Further research is needed to explore why this low level of agreement exists. It is possible that this low level of agreement is due to clinicians trusting their own judgment more than the outcome tracking measure.

The post hoc analysis revealed that changes in OQ scores significantly differed between the different client groups depending on whether the therapist judged treatment to be

successful or incomplete (See Table 2). There were no significant differences between the successful completion and substantial progress groups, suggesting that the distinction between these different groups may not be meaningful. Within the subset of clients who were judged to have successfully completed treatment, the mean change in OQ score was a decrease of 18.8 points. This supports the conclusion that a client who has been judged to have successfully completed treatment will, on average, have had a decrease in OQ score by 18.8 points, which is a larger score decrease than the one necessary for obtaining RC. However, this analysis does not provide information about the level of agreement between therapist judgment and RC as it applies to individual cases.

The rates of RC and CSC found in this study seem to be commensurate with those reported in other psychological training clinics. In this study, it was found that 35.4% of clients achieved RC. Prior reported rates of RC in other training clinics include 37% (Swift et al., 2009) and 33% (Callahan & Hynan, 2005). In this study, it was established that 18.2% of clients achieved CSC; it is worth noting, however, that 47.5% of the clients included in this study could not achieve CSC because they started treatment with a score in the nonclinical range. Prior reported rates of CSC in other training clinics include 23% (Swift et al., 2009) and 18% (Callahan & Hynan, 2005). As such, the rates of RC and CSC found in this study appear consistent with the rates reported by other training clinics.

Rates of Reliable and Clinically Significant Change

It has been suggested that therapist perception of treatment progress and client self-report of treatment satisfaction are not accurate indicators of treatment effectiveness (Mours et al., 2009). By contrast, this study showed that therapist perception of treatment outcome was significantly related to the criteria of RC and CSC on a standardized measure. In this study, therapists judged that 56.6% of clients achieved successful completion of treatment or made substantial progress, whereas 35.4% of clients were found to have met the criteria for RC. When comparing treatment success rates, there was a discrepancy of over 20%, which was potentially due to errors in therapists' perception (Ruscio, 2007).

It is also possible that in some cases, treatment was successful despite the client not achieving RC on the total OQ score. For this particular subgroup, the mean change in total OQ scores was a decrease by 9.67 points. This decrease in overall score could indicate RC on one of the subscales of the OQ, but this could not be further investigated since subscale data were not collected for the database. The OQ has three subscales: Symptom Distress, Interpersonal Relations, and Social Role. On these subscales, a respective decrease in score of 10, 8, and 7 points is considered RC (Lambert et al., 2004). It is possible that if subscale data were collected, then individuals who did not demonstrate RC on the OQ may demonstrate RC on one of these subscales. The therapists who rated the successfulness of treatment were cognizant of

this subscale data, and that knowledge may, in part, account for the disparity between therapist judgment and RC rates. As previously stated, OQ scores were the only client outcome data collected in this study. It is therefore possible that if these clients were administered a standardized measure specific to their presenting difficulties, then perhaps RC may have been found. In many instances in this clinic, additional standardized measures were utilized. For example, clients being treated for an anxiety disorder are often administered an inventory specific to anxiety, such as the Beck Anxiety Inventory (BAI; Beck & Steer, 1990). If a client were repeatedly administered an additional measure, then scores from that measure may have influenced the therapists' description of the outcome of therapy; however, data from any additional measures were not collected as a part of this investigation. The knowledge gained from additional assessments may account for some of the disparity in treatment success rates.

Implications for the Relationship between Therapist Judgment and Outcome Measures

The study's findings of the low level of agreement between changes in scores on the OQ and therapist judgment have important implications for clinical training and practice. Student therapists should be trained to recognize that their perception of client progress may not accurately reflect the client's self report. Psychologists in training may, in some cases, be selectively attending to that information which confirms their beliefs that the client is making progress, thereby ignoring any conflicting information (Ruscio, 2007). For example, if a therapist feels that a client is making progress, then he or she may focus on how the client consistently attends sessions and ignore the client's self-report of increasing symptomatology on an outcome measure. Clinicians would benefit from seeking information that is both confirmatory and disconfirmatory to their perceptions (Ridley, 2005).

Student clinicians' supervisors may play a role in helping students learn how to obtain such information and to navigate the process of developing a comprehensive view of clients' progress in treatment. The use of standardized outcome measures can and should be a preferred method to evaluate the accuracy of a clinician's intuition, and to inform practice by tracking client progress. Psychologists in training are forming the habits which will guide their future practice, and it is imperative that they form habits consistent with best and ethical practice guidelines (APA, 2006). Such habits would include regularly using outcome tracking measures to track client progress, evaluating if changes in course of treatment are necessary, and determining if therapy was successful. This study utilized a measure of general distress, but other outcome tracking measures which quantify symptom distress, life satisfaction, and personality functioning are also valuable sources of information. In addition to seeking guidance from supervisors, student clinicians may benefit from participation in trainings designed to introduce the concept of using

TRAINING CLINIC EXAMINATION

multiple sources of information to evaluate treatment progress.

Limitations

The present study used archival data, which therefore limited the type of data available for analysis. The information about therapist judgment of outcome was categorical, limiting the types of analyses which could be conducted. Also, this study utilized nonparametric statistics, and it can therefore be difficult to estimate how the findings of this study relate to the general population. This study took place in the Pacific Northwest with a sample of clients that may not represent the cultural or ethnic diversity of other regions. Of the 99 clients whose information was included for analysis, demographic data was only available for 66 of those clients; this limits generalizability of the findings. Additionally, no demographic data about the therapists was collected, and it is possible that some therapist characteristics (e.g., years of training) could reveal interesting information. For example, a recent study found that particular student therapist variables, such as the number of client contact hours and days in doctoral training, were predictors of treatment outcome (Powell, Hunter, Beasley, & Vernberg, 2010).

Future Directions

The results of this study were mixed regarding the agreement between different ways for measuring outcomes. RC sometimes occurred when the therapist did not describe treatment as being successful. In other instances, RC did not occur, but the therapist did describe treatment as being successful. It remains unclear what a clinician should do when these two methods produce different descriptions for the same outcome. Should clinicians disregard their own judgment? Or should clinicians trust their judgment and intuition? The answer to these questions is beyond the scope of this study, but it may be important to recognize the discrepancies in the reports regarding therapy outcome. To account for the low level of agreement between the two methods, it is recommended that therapists draw information about the disposition of treatment from a number of sources, such as the client's self-report, scores on a routinely administered outcome measure, verbal report from a collateral source (e.g., a client's spouse), clinical observation, and progress towards treatment goals. Additionally, a Multitrait Multimethod type approach (MTMM; Campbell & Fiske, 1959) may be appropriate in determining whether treatment was successful. Gathering information about the outcome of treatment from multiple sources may provide a more comprehensive and therefore clearer description of the outcome of therapy.

One method of determining treatment outcomes that could incorporate both therapist perception and standardized measures is the use of the client's treatment plan as the standard for evaluating the successfulness of treatment. An appropriate treatment plan includes realistic and measurable goals (Leahy & Holland, 2000). It should also include goals reflecting what the client expects to gain from treatment. If

clients meet their treatment goals, then treatment has been successful. If clients do not meet their treatment goals, then treatment has not yet been successful. Research has found that many clinicians prefer using "real world functional indicators" instead of standardized measures in the evaluation of the effectiveness of treatment (Garland et al., 2003). Using the client's treatment plan to evaluate the successfulness of treatment may be a valuable addition to the use of therapist judgment and standardized outcome measures.

It is possible to construct treatment plans that measure progress based on real world functional indicators (e.g., improvements in school or work attendance), scores on outcome measures, and client reported improvement (e.g., subjective units of distress scales). Research on the use of treatment plans to evaluate when treatment has been successful will demonstrate whether a client's treatment plan is an appropriate method for determining the outcome of treatment. It is possible that research may find the essential components needed for a treatment plan to be an appropriate means of determining when treatment is successful. Another future direction for research would be to study the attitudes of student clinicians in regards to outcome measures and assessment, as it is unknown how their attitudes correspond to the previously studied attitudes of practicing professionals.

References

- American Psychological Association. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271-285. doi:10.1037/0003-066X.61.4.271
- Anderson, E. A., & Lambert, M. J. (2001). A survival analysis of clinically significant change in outpatient psychotherapy. *Journal of Clinical Psychology*, *57*, 875-888. doi:10.1002/jclp.1056
- Beck, A. T., & Steer, R. A. (1990). *Manual for the Beck Anxiety Inventory*. San Antonio, TX: Psychological Corporation.
- Brooks, F. (2007). The living wage movement: Potential implications for the working poor. *Families in Society*, *88*, 437-442. doi: 10.1606/1044-3894.3653
- Brown, A., Williams, A., Waltman, S., & Sutton, C. (May, 2010). *Using more than intuition: Understanding client outcomes in a psychology training clinic*. Poster session presented at the meeting of the Oregon Psychological Association, Portland, Oregon.
- Callahan, J. L., & Hynan, M. T. (2005). Models of psychotherapy outcome: Are they applicable in training clinics? *Psychological Services*, *2*, 65-69. doi: 10.1037/1541-1559.2.1.65
- Callahan, J. L., Swift, J. K., & Hynan, M. T. (2006). Test of phase model of psychotherapy in a training clinic. *Psychological Services*, *3*, 129-136. doi: 10.1037/1541-1559.3.2.129
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *The Journal of Behavioral Health Services & Research, 30*, 393-405. doi: 10.1007/BF02287427
- Hatchett, G. T., & Park, H. L. (2003). Comparison of four operational definitions of premature termination. *Psychotherapy: Theory, Research, Training, 40*, 226-231. doi:10.1037/0033-3204.40.3.226
- Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35*, 485-491. doi:10.1037/0735-7028.35.5.485
- Hunsley, J., Aubry, T. D., Verstervelt, C. M., & Vito, D. (1999). Comparing therapist and client perspectives on reasons for psychotherapy termination. *Psychotherapy: Theory, Research, Practice, Training, 36*, 380-388. doi: 10.1037/0033-3204.36.4.380
- 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. doi: 10.1016/S0005-7894(84)80002-7
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67*, 300-307. doi:10.1037/0022-006X.67.3.300
- 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. doi: 10.1037/0022-006X.59.1.12
- Lambert, M. J., & Hawkins, E. J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice, 35*, 492-499. doi: 10.1037/0735-7028.35.5.492
- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C.,...Burlingame, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.2)*. Orem, UT: American Professional Credentialing Service.
- Lambert, M. J., Okiishi, J. C., Finch, A. E., & Johnson, L. D. (1998). Outcome assessment: From conceptualization to implementation. *Professional Psychology: Research and Practice, 29*, 63-70.
- Leahy, R. L., & Holland, S. J. (2000). *Treatment plans and interventions for depression and anxiety disorders*. New York: The Guilford Press.
- Lunnen, K. M., & Ogles, B. M. (1998). A multi-perspective, multi-variable evaluation of reliable change. *Journal of Consulting and Clinical Psychology, 66*, 400-410.
- Lunnen, K. M., Ogles, B. M., & Pappas, L. N. (2008). A multiperspective comparison of satisfaction, symptomatic change, perceived change, and end-point functioning. *Professional Psychology: Research and Practice, 39*(2), 145-152.
- Mours, J. M., Campbell, C. D., Gathercoal, K. A., & Peterson, M. (2009). Training in the use of psychotherapy outcome assessment measures at psychology internship sites. *Training and Education in Professional Psychology, 3*, 169-176. doi: 10.1037/a0016135
- Murdock, N. L., Edwards, C., & Murdock, T. B. (2010). Therapists' attributions for client premature termination: Are they self-serving? *Psychotherapy Theory, Research, Practice, Training, 47*(2), 221-234.
- Pekarik, G., & Wolff, C. B. (1996). Relationship of satisfaction to symptom change, follow-up adjustment, and clinical significance. *Professional Psychology: Research and Practice, 27*, 202-208.
- Powell, J. L., Hunter, H. L., Beasley, L. O., & Vernberg, E. M. (2010). Using fine-grained indexes of therapists' experience and training to predict treatment outcome in a university-based training clinic for children and families. *Training and Education in Professional Psychology, 4*(2), 138-144.
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy Theory, Research, Practice, Training, 46*, 418-431. doi: 10.1037/a0017901
- Ridley, C. R. (2005). *Overcoming unintentional racism in counseling and therapy: A practitioner's guide to intentional intervention* (2nd ed.). Thousand Oaks, CA: Sage.
- Ruscio, J. (2007). The clinician as subject: Practitioners are prone to the same judgment errors as everyone else. In S. O. Lilienfeld & W. T. O'Donohue (Eds.), *Great ideas of clinical science: 17 principles that every mental health professional should understand* (pp. 29-47). New York: Routledge.
- Slade, M., McCrone, P., Kuipers, E., Leese, M., Cahill, S., Parabiaghi, A.,...Thornicroft, G. (2006). Use of standardized outcome measures in adult mental health services. *British Journal of Psychiatry, 189*, 330-336. doi: 10.1192/bjp.bp.105.015412
- Swift, J. K., Callahan, J., & Levine, J. C. (2009). Using clinically significant change to identify premature termination. *Psychotherapy: Theory, Research, Training, 46*, 328-335. doi:10.1037/a0017003
- Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24*, 190-195. doi: 10.1037/0735-7028.24.2.190
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82*, 50-59. doi: 10.1207/s15327752jpa8201_10