Research setting versus clinic setting: Which produces better outcomes in cognitive therapy for depression?

Carly R. Gibbons, PhD,
Philadelphia, PA

Shannon Wiltsey Stirman, PhD,
National Center for PTSD, VA Boston Healthcare System, and Boston University

Robert J. DeRubeis, PhD,
University of Pennsylvania

Cory F. Newman, PhD, and
University of Pennsylvania

Aaron T. Beck, MD
University of Pennsylvania

Abstract

Objective—To compare the outcomes of cognitive therapy for depression under controlled and clinically representative conditions, while holding several therapist and clinical assessment factors constant.

Method—Treatment outcomes for a sample of 23 adults with a primary diagnosis of Major Depressive Disorder who received cognitive therapy in an outpatient clinic were compared with outcomes of 18 clients who were treated in the cognitive therapy condition of a large, multi-site randomized clinical trial of treatments for depression. All participants had been treated by one of two therapists who served as clinicians in both settings. Individuals in the two samples were diagnostically and demographically similar (approximately 50% Female, 83% White). A variety of client characteristics, assessed prior to treatment, as well as the outcomes of treatment, were examined.

Results—Significantly superior treatment outcomes were observed in the individuals treated in the research study, relative to clients in the outpatient clinic, and the difference was not accounted for by intake characteristics. Individuals treated by the therapists in the RCT experienced almost three times as much improvement in depressive symptoms as clients seen in the outpatient setting.

Conclusions—If replicated, the findings suggest that differences exist between treatment outcomes in research and outpatient settings and that these differences may not simply be due to therapist experience and training, or differences in patient populations. Future research should further examine the impact of fidelity monitoring, treatment expectation and motivation, and the duration and timing of treatment protocols on clinical outcomes.

Introduction

The establishment of a set of evidence-based treatments (EBTs; see Kazdin, 2008) has depended heavily on findings from randomized clinical trials (RCTs). The maximization of internal validity in RCTs are seen as great advantages by developers of treatment guidelines, but lead some to question the relevance of RCT findings to clinical practice (Kazdin, 2008). Features of RCTs include stringent inclusion and exclusion criteria, treatment manuals, and careful selection, training, and supervision of therapists. Such factors may lead to differences in treatment delivery and outcomes in RCTs compared to routine care settings.

The effectiveness and efficacy of cognitive therapy (CT) have been demonstrated for many psychiatric disorders, including depression (Beck, 2005; Westbrook & Kirk, 2005). However, studies of the effectiveness of EBTs in routine care have frequently utilized a transported version of the EBT, with levels of therapist training and supervision that are not typical of routine clinical care. In a meta-analysis of 32 studies in which EBTs were compared to usual care treatments for adolescent depression, Weisz, Jensen-Doss, and Hawley (2006), noted that none of the studies controlled for the selection of therapists, the treatment setting, or the dose of treatment. Therapists in RCTs are selected, supervised and trained carefully. They are also monitored throughout the trial, both for treatment integrity (Perepletchikova, Treat, & Kazdin, 2007) and patient outcomes (Newman & Beck, 2008). In contrast, more clinically representative conditions do not include training or monitoring (Shadish, Matt, Navarro & Phillips, 2000). Despite these differences, a review by Hunsley and Lee (2007) found comparable rates of treatment completion and symptom improvement in RCTs and effectiveness studies. Similarly, in a meta-analysis of effectiveness studies for anxiety disorders, Stewart and Chambless (2009) found that effect sizes from effectiveness studies were similar to those found in efficacy benchmarks.

In the present study we compare the outcomes of CT for depression under both controlled and clinically representative conditions. Two therapists who worked at an outpatient therapy clinic and considered CT their primary orientation each served as clinicians in an RCT for major depressive disorder (MDD) and provided CT to their own outpatients. Thus, the selection of therapists does not present a confound. They provided the treatments during the same time period, and in the same setting, so the levels of therapist training and experience, as well as the treatment environment, were equated between the two samples. Although individuals in both samples completed similar assessment procedures, treatment and client outcomes were not expressly monitored at the Center for Cognitive Therapy, as they were in the RCT, and therapists had more latitude in planning and providing treatment. Thus, we aim to provide a comparison of treatment outcomes in two different settings while holding several therapist and clinical assessment factors constant. Because these variables could be held constant in the current study, we hypothesized that differences in outcomes between the two groups would be minimal.

Method

Participants

Diagnostic and treatment outcome information was obtained from the intake evaluations and weekly self-report measures of clients at the Center for Cognitive Therapy (CCT) who consented to allow these data to be used for research. Between 1995 and 1999, 217 clients received a primary diagnosis of primary MDD. Of these, the 23 clients were selected to receive treatment from one of two therapists who provided treatment in a large RCT that was conducted during the same time period. The therapists were Ph.D. psychologists, each with at least 5 years of postgraduate experience. Therapist 1 treated 17 clients, and Therapist 2 treated six clients in the CCT sample.

The CCT is a university-affiliated outpatient clinic in Philadelphia, PA that treats individuals with a wide variety of DSM-IV Axis I and Axis II disorders. During the time period from which the sample was drawn, the CCT accepted self-payment and insurance. Clients at CCT typically received weekly CT sessions, with the frequency of sessions varying as a function of symptom severity levels, schedules, and financial considerations. Some clients in the CCT sample may have engaged in adjunctive treatments or used psychotropic medications.

Intake evaluations were conducted by Ph.D. level assessors using the Structured Clinical Interview for DSM-IV Diagnosis for Axis I and the Structured Clinical Interview for DSM-

IV Diagnosis Axis II (First & Gibbon, 2004). Assessors were trained to use the assessment instruments in workshops that totaled approximately 20 hours over 3 weeks. A Ph.D.-level supervisor regularly oversaw all assessments, and diagnoses were achieved by consensus. Clients also completed the Beck Depression Inventory (BDI II; Beck, Steer, Ball, & Ranieri, 1996). For this study, the BDI scores for each weekly session were collected from therapy charts.

Participants included in the RCT sample were those who were treated by either of the two aforementioned therapists in the RCT (n=18). Each of the therapists treated nine RCT participants. Inclusion criteria for the study included: (1) diagnosis of MDD according to *DSM-IV* criteria, (2) age 18 to 70 years, (3) English-speaking, and (4) willingness and ability to give informed consent. Consistent with Elkin and colleagues' (1989) definition of "more severely depressed," all included participants had scores of 20 or higher on the modified 17-item Hamilton Depression Rating Scale (Hamilton, 1960) at the screen and baseline visits, separated by at least 7 days. Potential participants were excluded if they had (1) a history of bipolar I disorder; (2) substance abuse or dependence judged to require treatment; (3) current or past psychosis; (4) another *DSM-IV* Axis I disorder judged to require treatment in preference to the depression; (5) 1 of the 3 excluded *DSM-IV* Axis II disorders deemed to be poorly suited to the treatments under investigation (antisocial, borderline, and schizotypal); (6) suicide risk requiring immediate hospitalization; (7) medical condition that contraindicated study medications; or (8) nonresponse to an adequate trial of paroxetine in the preceding year.

Treatment

In both settings, cognitive therapy was provided following the procedures outlined in standard texts of cognitive therapy for depression (Beck, Rush, Shaw, & Emery, 1979; Beck, 1995) and comorbid personality disorders (Beck & Freeman, 1990). The efficacy of cognitive therapy for depression has been supported in several clinical trials (Elkin et al., 1989; Hollon et al., 1992; Murphy et al., 1984). Although research suggests that cognitive behavioral treatments are effective for personality disorders (Leichsenring & Leibing, 2003), the approach of combining the specific manual for personality disorders used in this study (Beck & Freeman, 1990) with a manualized treatment for depression had not been tested prior to the current RCT (DeRubeis et al., 2005). In both settings, strategies for personality disorders could be introduced for individuals who presented with beliefs or behaviors that were consistent with cognitive conceptualizations of Axis II disorders and that were judged to interfere with treatment progress. Cognitive and behavioral interventions outlined in the Beck & Freeman (1990) manual were intended to be used to address beliefs or behaviors that were conceptualized as maintaining depressive symptoms or interfering with treatment for depressive symptoms in the RCT. In both settings, clinicians were permitted to select and sequence interventions as they believed appropriate for particular clients, rather than following a sequence of interventions set forth in the manuals. RCT clinicians met weekly for 90 minutes to review ongoing cases, however, no such provision was required at the CCT.

Measure: Beck Depression Inventory-II (BDI-II; Beck, et al., 1996)

The BDI-II contains 21 items, each rated from 0 to 3, which are summed to obtain a total score. Higher scores reflect more severe depressive symptomatology (range 0 to 63). The BDI-II's internal consistency has been shown to be high (Beck, et al., 1996).

Analytic Strategy

Participants in the two samples were compared on demographic and diagnostic factors using chi square analyses and t-tests. Next, longitudinal BDI scores across sessions from

participants treated at both the CCT and in the DeRubeis et al. (2005) study were examined using hierarchical linear modeling (HLM; also known as multilevel linear modeling and growth curve modeling). At level 1, within-subject variance is modeled from a collection of subject-specific parameters (slope and intercept), which were treated as having been randomly sampled from a population of individuals. At level 2, the subject-specific parameters are modeled in order to identify meaningful sources of between-subject variation. When the two models are combined in an HLM, the result is a mixed linear model with fixed and random effects. For all models described below, an unstructured covariance structure was used in order to model the correlation between the participant-specific slopes and intercepts. All available data were included from all participants in both treatment settings, regardless of whether the participants completed treatment or were considered dropouts. The HLM models (performed using SAS version 9.1, PROC MIXED; SAS Institute Inc., Cary, NC) were used to assess whether the two settings, research and non-research, differed in the rate of symptom reduction over time (as evidenced by a significant time-by-site interaction) and whether the two settings differed in estimated endpoint scores (as evidenced by the main effect of site at the intercept, which was centered to represent scores at the end of treatment).

The treatment protocol in the RCT called for twice weekly therapy sessions for the first four weeks of treatment, and weekly sessions thereafter. Because CCT clients typically received weekly sessions throughout treatment, the frequency of treatment sessions between the two settings was not equivalent. Additionally, therapy at CCT did not have a fixed endpoint, whereas treatment in the RCT was time-limited. No statistical correction can control for these differences and no single analysis of the data can address these differences without biasing the results in favor of one setting or the other. Therefore, to best address these issues, we examined the outcome data in several complementary ways. We compared outcomes at three different time points. Interpretation of the findings will focus on patterns that replicated or converged across analyses.

In the first set of analyses, the dependent variable was the BDI at posttreatment, using each participant's final BDI score irrespective of when it was obtained. In the second, the 15-week time point was chosen. Here, time was fixed but the number of sessions was allowed to vary within that timeframe and RCT participants tended to have received more sessions than the CCT clients (15 weeks was chosen instead of the 16 week RCT treatment endpoint because it is believed that the cost of reducing the observation period by one week was favorable to the possibility of a biasing effect of treatment termination on self-reported symptoms). Finally, data were examined at the (maximum of the) 20th session of treatment. Twenty sessions was chosen as a cutoff because it is a frequently used maximum number of sessions of cognitive therapy for depression in RCTs. In this case, none of the clients received a higher "dose" of treatment than this maximum, but the time over which the sessions occurred (and therefore the time that elapsed between intake and this end-point was allowed to vary). In the second and third sets of analyses, if a client did not provide a score at the endpoint, typically because treatment ended before that time or session was reached, a last observation carried forward (LOCF) strategy was employed. Each statistical model included intake BDI score and the interaction between setting and intake BDI as covariates, to reduce the effect of any differences in symptom severity at intake. All models also included therapist, and the interaction between therapist and setting, as variables.

Analyses of clinically significant change on the BDI were conducted according to Jacobson and colleagues' formulae (e.g. Jacobson and Truax, 1991; Jacobson, Roberts, Burns, & McGlinchey, 1999). This method evaluates two criteria for each participant. The first is whether each participant's BDI score improved such that it is unlikely to be due to chance (reliable change index, RCI). The RCI is a function of a participant's pre and posttest scores,

the standard deviation of the population prior to treatment, and the test-retest reliability of the measure (0.93; Beck et al., 1996). A participant is considered to have experienced reliable change if his or her RCI is greater than 1.96 (Jacobson et al., 1999). The second criterion evaluated, for participants shown to have reliable change, is whether their posttreatment symptom level now places them within the "normal" range for this measure. This calculation requires use of a normative sample. For this study, the normative comparison was drawn from Dozois et al. (1998). This appears to be the largest sample of its kind; it has been used in similar analyses (e.g. Westbrook and Kirk, 2005). The cutoff point for determining whether a participant "recovered" was calculated according to Jacobson's criterion 'c' (Jacobson & Truax, 1991).

Results

Chi-square analyses of pre-treatment demographic characteristics (see Table I) indicated that the two samples did not differ in the proportion of individuals who were female, white, married, or unemployed. Among the diagnostic factors we examined (see Table 1), the only significant difference was obtained in the proportion of participants with comorbid Axis II disorders $\chi^2(1, N = 41) = 3.76$, $p = 0.05$, with a higher percentage observed in the clinic sample (74%) relative to the RCT sample (39%). Mean intake BDI scores did not differ significantly between the clinic sample (mean = 27.2) and the RCT sample (mean = 30.7), $t(39) = 1.24$, $p = 0.22$, $d = 0.38$.

Descriptive statistics of treatment length and outcome can also be found in Table I. In both samples, participants were seen for approximately 18 sessions on average, with no statistically significant difference in mean number of sessions. Outcomes for participants in the two treatment settings were compared at posttreatment. This time point would presumably bias results in favor of CCT where clients could pursue a longer course of treatment than in the RCT. At posttreatment, clients at CCT were found to have significantly higher BDI scores than participants in the DeRubeis et al. (2005) RCT, ($t(39) = 3.21$, $p = 0.003$). In fact, participants in the DeRubeis et al. (2005) study experienced three times the reduction in BDI points from pre- to post-treatment, relative to the clients treated by the same therapists at CCT did, $t(39) = 3.94$, $p < 0.001$. This translates into an effect size (Cohen, 1992) for treatment setting of $d = 1.29$, a large effect. A general linear model was applied, predicting BDI score at posttreatment from treatment setting. In this model, treatment setting significantly predicted outcome, $F(1, 32) = 19.54$, $p < 0.0001$. Because there was evidence of a difference in rates of Axis II disorders between the samples at intake, the model was expanded to include the presence of Axis II comorbidity as a covariate. Treatment setting remained a significant predictor of outcome, $F(1, 33) = 15.85$, $p < 0.001$.

Two additional analyses were performed to examine the effects of treatment setting while making the best effort to address potential differences between the samples. In the first analysis, a general linear model was applied predicting outcomes from treatment setting while controlling for *all* of the diagnostic variables listed in Table I, in addition to intake BDI score, the interaction between setting and intake BDI score, therapist, and the interaction between therapist and setting. In this model, treatment setting remained a significant predictor of treatment outcome, $F(1, 29) = 13.28$, $p < 0.01$. In the second analysis, the CCT sample was diagnostically restrained to only include clients with diagnoses that were accepted in the inclusion criteria of the DeRubeis et al. (2005) RCT (n = 20). This step eliminated all clients with levels of severity that were lower than the cutoff for inclusion in the RCT, and also eliminated clients with Axis II and other comorbid diagnoses that would not have been represented in the RCT. A general linear model was applied, predicting outcome from treatment setting. Treatment setting remained a significant predictor of outcome $F(1, 31) = 15.81$, $p < 0.001$.

The same pattern of results held true when results were compared at the other time points. The model in which the dependent variable was the 15-week BDI score yielded a significant prediction by site, $F(1, 31) = 13.38$, $p < 0.001$, with RCT participants demonstrating lower BDI scores at 15 weeks than their CCT counterparts. Treatment settings remained a significant predictor of outcome when the model was expanded to include Axis II comorbidity as a covariate, $F(1, 32) = 12.15$, $p < 0.01$, when all covariates were included in the model, $F(1, 28) = 9.87$, $p < 0.01$, and when the basic model was applied using CCT clients who met the RCT inclusion criteria, $F(1, 30) = 11.03$, $p < .01$.

With the BDI after the 20th session as the dependent variable, setting was once again a significant predictor of outcome, even when intake BDI, the interaction between intake BDI and setting, therapist, and the interaction between therapist and setting were covaried, $F(1, 31) = 12.97$, $p < 0.01$. Again, RCT participants demonstrated lower BDI scores than their CCT counterparts. When the model was expanded to include Axis II comorbidity as a covariate, treatment setting remained a significant predictor of outcome, $F(1, 32) = 12.16$, $p < 0.01$. Again treatment setting remained a significant predictor of outcome when all covariates were included in the model, $F(1, 28) = 9.48$, $p < 0.01$, and when the basic model was applied excluding the three individuals with borderline personality disorder, $F(1, 30) = 10.30$, $p < 0.01$.

Analyses were conducted to determine whether this pattern of results held equally for individuals treated by both therapists (see Table I). RCT participants experienced significantly greater symptom improvement than clients treated by the same therapist at CCT, in terms of total improvement in BDI score and in terms of BDI score at the end of treatment.

Clinically significant change in depressive symptoms as measured on the BDI was assessed. Using the criterion described above, 52% of the CCT participants experienced reliable improvement over the course of treatment (14% were shown to have reliable deterioration) and 100% of the RCT participants experience reliable improvement (0% had reliable deterioration). In order to assess whether these participants would be considered "recovered," the cutoff score (Jacobson's criterion 'c'; Jacobson and Truax, 1991) was calculated to be a BDI score of 16.5 for both samples. Thus, if participants had demonstrated reliable change *and* finished treatment with a BDI score below 16.5, they were considered recovered. Only participants who began treatment with a BDI score of 16.5 or above ($n = 20$ at CCT, $n = 16$ in RCT) were included in this analysis, as those with scores lower than 16.5 did not have the opportunity to meet both criteria. Eight participants at CCT (40% of those who met the first criteria, 35% of total sample) and thirteen in the RCT (81% of those who met the first criteria, 72% of total sample) were found to be below the cutoff and therefore considered "recovered."

Discussion

This study provides a unique view of the role of treatment setting on client outcome in CT for depression. Data were collected from individuals treated by two therapists who routinely provided CT in an outpatient clinic and also served as cognitive therapists in the RCT. The results suggested a striking tendency for outcomes in the RCT to be superior to those evidenced at CCT. In fact, individuals seen by these therapists in the RCT experienced almost three times as much improvement in depressive symptoms as clients seen at CCT.

One possible explanation for these surprising findings is that the samples were not equivalent. The subsample of individuals seen by these therapists in the two treatment settings were, with one exception, similar to each other in regard to a variety of indicators

that might predict differential outcomes. Rates of Axis II comorbidity differed considerably between the two samples. However, analyses repeated with only those CCT clients who met inclusion criteria for the RCT indicated the same pattern of results that we found with the full sample. Further, when we controlled for this difference statistically, analyses indicated that it did not appear to account for the differences observed in outcomes. Despite these efforts to statistically control for Axis II diagnoses, complications related to personality disorders may have decreased the impact of treatment, or necessitated additional sessions before greater changes in depression could be realized in the CCT sample. This finding is particularly interesting because clinicians had received training in protocols that included strategies for treating individuals with personality disorders. While the findings of this study cannot necessarily be attributed to a lack of training or attention to personality factors, it is possible that the RCT therapists received feedback or advice in weekly consultation sessions that helped them treat individuals with Axis II disorders more effectively. The two groups may have also differed in terms of motivation for treatment. Although it is important to note that both groups were willing to attend preliminary appointments to undergo several hours of assessment before they received any treatment and were willing to pay for whatever expenses were not covered by insurance (e.g., a co-payment) for treatment, whereas RCT participants received free treatment and were paid to attend assessments. Another notable difference between the two groups was the possibility that CCT clients were also using psychotropic medications, whereas RCT participants were not. Data for the full sample of CCT clients indicates that 40% of the larger sample was using medications (Gibbons et al., 2010). Unfortunately, more specific data regarding the medications that individuals in the CCT group were prescribed or using was not available. It is possible that the combination of psychotherapy and medications may have a different impact on psychotherapy process and outcomes, although the combination has been shown to be advantageous in comparison to monotherapies (Friedman et al., 2006). There is also evidence that particular behavioral activities may be more predictive of change when antidepressant medications are combined with cognitive therapy, while cognitive strategies are more predictive when cognitive therapy alone is provided (Strunk et al., 2012). However, because fidelity data are not available for the CCT sample, it is not possible to determine whether clinicians adjusted their treatment strategy for individuals who were using medications.

An additional possibility that these findings raise is that differences exist between treatment outcomes in research and outpatient settings. These differences may not simply be due to therapist experience, training, or skill, as the design of the study allowed these factors to be held constant. If replicated in a study in which depressed individuals are randomized into a "research" or "clinic" setting with a larger sample of therapists, these findings could have major implications both for the design and implementation of RCTs and the execution of ESTs in clinical practice.

Factors such as supervision and monitoring of therapists and measurement of client progress may play a major role in therapist performance and client outcomes. Therapists in the RCT study had their sessions videotaped and were expected to adhere strictly to the treatment model, which may have improved client outcomes. More frequent supervision in the RCT may have also improved treatment delivery. One of the therapists was a senior clinician at CCT and did not participate in additional supervision for his non-research clinical work. Additionally, clinicians may have worked more effectively with research clients whose outcomes were being systematically monitored (Lambert, 2007) through the RCT. It is possible that one or both clinicians were less adherent to cognitive therapy protocols when they were not closely supervised or subject to fidelity monitoring, and this may have contributed to differences in outcomes between the two samples. Data from the clinical trial indicated that clinicians' average competence scores differed (Strunk et al., 2010). However, the clinician who achieved higher average competence ratings in the trial saw far more

clients in the CCT than the clinician who achieved lower competence ratings. The clients of the clinician who achieved higher average competence ratings in the clinical trial also experienced greater symptom reduction in the current study, but because the CCT did not collect fidelity data, we are unable to determine whether there is a relationship between fidelity and client outcomes. If future research indicates that differences in fidelity contribute to differences in outcomes in research and clinic settings, increased supervision and monitoring of cognitive therapy in routine clinical settings may be warranted.

The time-limited nature of treatment in the research setting also may affect outcomes. Therapists and clients may have been more likely to enact changes more rapidly or set more deliberate priorities for treatment when the limited amount of time was factored into the treatment plan (Reynolds, Stiles, Barkham, Shapiro, Hardy, & Rees, 1996). Differences in client expectations may contribute to differences in outcomes as well. As the RCT specifically recruited individuals with moderate to severe depression, RCT participants presumably expected and understood that their depression would be a major focus of the intervention. CCT clients, on the other hand, may have entered treatment with an expectation that they would simultaneously address multiple problems or symptoms, or may not have believed that depression was the most pressing issue to address in treatment. Thus, substantive differences between the focus or content of the treatment in the two settings may have impacted outcomes. CCT clients may have been less receptive to efforts to focus on one clinical problem at a time. Similarly, therapists may have pursued a narrower focus on depressive symptoms in the RCT, whereas they may have attempted to address a wider array of issues with CCT clients.

Additionally, frequency of treatment differed between the two treatment settings and the research sample had a maximum treatment length not imposed at CCT. These factors make comparisons difficult. Notably, RCT participants generally met with their therapist twice a week for the first few weeks, while CCT clients typically only met once a week. While we attempted to address this difference in our analytic strategies, we were not able to control for session frequency, and greater frequency may have contributed to a sense of "momentum" for RCT participants. Constraints in service delivery settings, client schedules and availability, and insurance reimbursement policies may limit the extent to which more frequent sessions are feasible in typical outpatient settings, and this might be an important difference between treatments in clinical trials versus routine service delivery settings. The potential impact that frequency of sessions may have on clinical outcomes warrants further study.

If the differences in outcomes that were found in this study result from contextual factors rather than differences in client characteristics, some important implications for enhancing regular practice can be gleaned from the differences between settings. Strategies such as formal outcome assessment, ongoing consultation, or fidelity monitoring may improve outcomes in clinically representative settings by increasing accountability and providing opportunities to adjust treatment strategies to address non-response or comorbid presenting problems. Formalizing agreements about the frequency and duration of treatment as well as specific target problems that will be prioritized may improve treatment efficiency and allow opportunities to address expectancies for treatment. While some of these enhancements may be relatively simple to implement, others, such as ongoing fidelity monitoring and supervision, may require resources that are unlikely to be available in many clinically representative settings. However, if such strategies are ultimately shown to result in greater improvement, they may be more cost-effective over the long term.

As our discussion above indicates, there are some limitations to the current research, many of which are inherent in efforts to compare research and naturalistic samples. While we

controlled for several factors that may threaten the internal validity of a comparison of outcomes in research and clinic settings, the study nonetheless includes a number of factors that limit our ability to explain the differences in outcomes. It will also be important for future comparisons to include larger sample sizes of clients and therapists than were available for the current study. Therapists may vary in terms of their tendency to remain highly adherent to treatment manuals in clinical practice or in other ways that could have impacted outcomes in the current study. Evaluating the impact of therapist characteristics and assessing treatment fidelity with an increased sample of therapists will allow a better understanding of the impact of such factors on treatment outcomes. Finally, it is important to note that in studying the differences in treatment outcomes between clinic and research settings, we examined only one research and one outpatient sample, for one disorder. Because the RCT included participants with Axis II comorbidity, the results of the current study may not generalize to similar comparisons for depression alone in research and clinic settings, and findings may also differ for other disorders.

Efforts to understand the extent to which treatments can be transported to non-research settings will likely include critical differences in samples, treatment fidelity and frequency, and other variables. This study identifies several factors that may be responsible for differences in outcomes, and more work remains to be done to identify those most responsible for our findings. However, the results of this study, while not conclusive, add to efforts to understand the relation between data obtained in clinical settings and those obtained in RCTs. We encourage further efforts to understand the limits of generalizability from research settings to clinically representative settings, as well as efforts to bridge the gap between practice and outcomes in research and practice.

Acknowledgments

References

Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. Cognitive therapy of depression. New York, NY: Guilford Press; 1979.

Beck AT, Steer R-A, Ball R, Ranieri WF. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. Journal of Personality Assessment. 1996; 67(3):588–597.10.1207/s15327752jpa6703_13 [PubMed: 8991972]

Beck AT. The current state of cognitive therapy: A 40-year retrospective. Archives of General Psychiatry. 2005; 62:953–959.10.1001/archpsyc.62.9.953 [PubMed: 16143727]

Beck, JS. Cognitive therapy: Basics and beyond. New York, NY: Guilford Press; 2005.

Beck, AT.; Freeman, A. Cognitive therapy of personality disorders. New York, NY: Guilford Press; 1990.

Cohen J. A power primer. Psychological Bulletin. 1992; 112:155–159.10.1037/0033-2909.112.1.155 [PubMed: 19565683]

Dozois D, Dobson K, Ahnberg J. A psychometric evaluation of the Beck Depression Inventory-II. Psychological Assessment. 1998; 10:83–89.10.1037/1040-3590.10.2.83

DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, Gallop R. Cognitive therapy vs. medications in the treatment of moderate to severe depression. Archives of General Psychiatry. 2005; 62(4):409–416.10.1001/archpsyc.62.4.409 [PubMed: 15809408]

Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Parloff MB. NIMH treatment of depression collaborative research program: General effectiveness of treatments. Archives of General Psychiatry. 1989; 46:971–982.10.1001/archpsyc.1989.01810110013002 [PubMed: 2684085]

First, MB.; Gibbon, M. The structured clinical interview for DSM-IV axis I disorders (SCID-I) and the structured clinical interview for DSM-IV axis II disorders (SCID-II). Hoboken, NJ: John Wiley & Sons Inc; 2004.

Friedman ES, Wright JH, Jarrett RB, Thase ME. Combining cognitive therapy and medication for mood disorders. Psychiatric Annals. 2006; 36:320–32.

Hamilton M. A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry. 1960; 23:56–62.

Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, Tuason VB. Cognitive therapy and pharmacotherapy for depression: Singly and in combination. Archives of General Psychiatry. 1992; 49:774–781.10.1001/archpsyc.1992.01820100018004 [PubMed: 1417429]

Hunsley J, Lee CM. Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. Professional Psychology: Research and Practice. 2007; 38(1): 21–33.10.1037/0735-7028.38.1.21

Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. Journal of Consulting and Clinical Psychology. 1999; 67:300–307.10.1037/0022-006X.67.3.300 [PubMed: 10369050]

Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology. 1991; 59:12–19.10.1037/0022-006X.59.1.12 [PubMed: 2002127]

Kazdin A. Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. American Psychologist. 2008; 63:146–159.10.1037/0003-066X.63.3.146 [PubMed: 18377105]

Lambert M. Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. Psychotherapy Research. 2007; 17(1):1–14.10.1080/10503300601032506

Leichsenring F, Leibing E. The effectiveness of psychodynamic therapy and cognitive-behavior therapy in the treatment of personality disorders: A meta-analysis. American Journal of Psychiatry. 2003; 160:1223–1232.10.1001/archpsyc.61.12.1208 [PubMed: 12832233]

Murphy GE, Simons AD, Wetzel RD, Lustman PJ. Cognitive therapy and pharmacotherapy. Archives of General Psychiatry. 1984; 41:33–41. [PubMed: 6691783]

Newman, CF.; Beck, JS. Selecting, training, and supervising therapists in randomized controlled trials. In: Nezu, AM.; Nezu, CM., editors. Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions. Oxford, UK: Oxford University Press; 2008. p. 245-262.

Perepletchikova F, Treat TA, Kazdin AE. Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. Journal of Consulting and Clinical Psychology. 2007; 75:829–841.10.1037/0022-006X.75.6.829 [PubMed: 18085901]

Reynolds S, Stiles WB, Barkham M, Shapiro DA, Hardy GE, Rees A. Acceleration of changes in session impact during contrasting time-limited psychotherapies. Journal of Consulting and Clinical Psychology. 1996; 64(3):577–586.10.1037/0022-006X.64.3.577 [PubMed: 8698952]

Shadish WR, Matt GE, Navarro AM, Phillips G. The effects of psychological therapies under clinically representative conditions: A meta-analysis. Psychological Bulletin. 2000; 126(4):512–529.10.1037/0033-2909.126.4.512 [PubMed: 10900994]

Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. Journal of Consulting and Clinical Psychology. 2009; 77(4):595–606.10.1037/a0016032 [PubMed: 19634954]

Strunk DR, Brotman MA, DeRubeis RJ, Hollon SD. Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. Journal of Consulting and Clinical Psychology. 2010; 78(3):429–437.10.1037/a0019631 [PubMed: 20515218]

Strunk DR, Cooper AA, Ryan ET, DeRubeis RJ, Hollon SD. The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early intersession symptom gains. Journal of Consulting and Clinical Psychology. 2012 Advance online publication. 10.1037/a0029281

Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. American Psychologist. 2006; 61(7):671–689.10.1037/0003-066X.61.7.671 [PubMed: 17032068]

Westbrook D, Kirk J. The clinical effectiveness of cognitive behaviour therapy: outcome for a large sample of adults treated in routine practice. Behaviour Research and Therapy. 2005; 43(10):1243–1260.10.1016/j.brat.2004.09.006 [PubMed: 16086979]

Table I

Comparison of Patients Treated by Two Therapists at Two Treatment settings

	CCT N = 23	RCT N = 18	Comparison	Effect Size
Demographics			χ^2, p	
Female, %	52	50	0.01, 0.89	0.02
White, %	83	83	0.00, 0.95	0.00
Married, %	55	44	0.10, 0.75	0.05
Unemployed, %	13	17	0.11, 0.74	0.05
Anxiety Disorders	61	44	0.54, 0.46	0.11
Substance Disorders, %	33	17	0.66, 0.42	0.13
Double Depression, %	6	4	0.03, 0.86	0.03
Recurrent Depression, %	61	70	0.06, 0.81	0.04
Axis II Comorbidity, %	74	39	**3.76, 0.05**	0.30
Symptom and Treatment Factors			t, p	d
Intake BDI, mean (SD)	27.2 (7.8)	30.7 (10.5)	t=1.24, p=0.22, d=0.38	
Number of Sessions, mean (SD)	18.4 (18.7)	18.7 (6.9)	0.06, 0.95	0.02
Posttreatment BDI, mean (SD)	20.5 (14.6)	8.2 (7.9)	**3.12, <0.01**	**1.05**
BDI Improvement, mean (SD)	7.4 (12.2)	22.6 (11.2)	**3.94, <0.001**	**1.30**
Outcomes by Therapist			t, p	d
Posttreatment BDI, Mean (SD)				
Therapist 1	16.7 (13.7)	5.6 (7.5)	**2.24, 0.04**	**1.00**
Therapist 2	29.8 (13.3)	10.6 (6.2)	**3.61, <0.01**	**1.85**
BDI Improvement, Mean (SD)				
Therapist 1	9.9 (11.8)	27.8 (8.2)	**3.99, <0.001**	**1.76**
Therapist 2	1.2 (12.1)	16.8 (11.7)	**2.44, 0.03**	**1.31**

Note: **Bold** font indicates that the samples were significantly different at α = 0.05.