NIH Public Access
Author Manuscript
J Affect Disord. Author manuscript; available in PMC 2011 September 1.

Published in final edited form as:
J Affect Disord. 2010 September ; 125(1-3): 169–176. doi:10.1016/j.jad.2009.12.030.

The Clinical Effectiveness of Cognitive Therapy for Depression in an Outpatient Clinic

Carly J. Gibbons, Ph.D.[1,2], **Jay C. Fournier**[3], **Shannon Wiltsey Stirman**[5], **Robert J. DeRubeis, Ph.D.**[3], **Paul Crits-Christoph, Ph.D.**[6], and **Aaron T. Beck, M.D.**[6]

[1]Yale University School of Medicine

[2]VA Connecticut Healthcare System

[3]Department of Psychology, University of Pennsylvania

[4]National Center for PTSD, VA Boston Healthcare System

[5]Department of Psychiatry, Boston University

[6]Department of Psychiatry, University of Pennsylvania

Abstract

Background—Cognitive therapy (CT) has been shown to be efficacious in the treatment of depression in numerous randomized controlled trials (RCTs). However, little evidence is available that speaks to the effectiveness of this treatment under routine clinical conditions.

Method—This paper examines outcomes of depressed individuals seeking cognitive therapy at an outpatient clinic (N = 217, Center for Cognitive Therapy; CCT). Outcomes were then compared to those of participants in a large NIMH-funded RCT of cognitive therapy and medications as treatments for depression.

Results—The CCT is shown to be a clinically representative setting, and 61% of participants experienced reliable change in symptoms over the course of treatment; of those, 45% (36% of the total sample) met criteria for recovery by the end of treatment. Participants at CCT had similar outcomes to participants treated in the RCT, but there was some evidence that those with more severe symptoms at intake demonstrated greater improvement in the RCT than their counterparts at CCT.

Limitations—The CCT may not be representative of all outpatient settings, and the structure of treatment there was considerably different from that in the RCT. Treatment fidelity was not assessed at CCT.

Conclusions—Depressed individuals treated with cognitive therapy in a routine clinical care setting showed a significant improvement in symptoms. When compared with outcomes

Corresponding author: Carly J. Gibbons, Ph.D., Yale University School of Medicine and VA Connecticut Healthcare System, 950 Campbell Avenue (151D), West Haven, CT 06516, Tel: (203) 937-3486 x3924, Fax: (203) 937-3472, carly.gibbons@yale.edu.

Conflict of Interests: None of the authors had any conflicts of interest.

Contributors: Carly Gibbons, Robert DeRubeis, and Paul Crits-Christoph designed the study. Aaron Beck provided access to the CCT data and Robert DeRubeis provided access to the RCT data. All authors contributed to planning data analyses and interpretation of data. Jay Fournier performed the majority of statistical analysis. Carly Gibbons and Shannon Wiltsey Stirman wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

evidenced in RCTs, there was little evidence of superior outcomes in either setting. However, for more severe participants, outcomes were found to be superior when treatment was delivered within an RCT than in an outpatient setting. Clinicians treating such patients in non-research settings may thus benefit from making modifications to treatment protocols to more closely resemble research settings.

Keywords

Cognitive therapy; depression; effectiveness; randomized controlled trials

Introduction

The efficacy of cognitive therapy (CT) in the treatment of many psychiatric disorders, including depression, has been established in a large number of randomized clinical trials (RCTs; Cuijpers et al., 2008; Butler et al., 2006; Strunk and DeRubeis, 2001, Gloaguen et al., 1998). The RCT methodology is well accepted as the "gold standard" for establishing efficacy (Seligman, 1995, p. 966). Based on RCT evidence, a number of psychosocial treatments have been designated as "empirically supported therapies" (ESTs) (Chambless and Hollon, 1998). Often, however, RCTs are conducted in academic settings using procedures that are discrepant from how psychotherapy is most commonly delivered. This includes the use of specially trained therapists, extensive supervision, random assignment to treatment, and restrictive participant inclusion/exclusion criteria. Some have argued that these features of RCTs compromise their clinical validity (Goldfried and Wolfe, 1998; Westen et al., 2004). This has led to an increase in focus on questions about the extent to which results of such efficacy trials can be translated into clinical effectiveness, as was captured by a question posed by Carroll and Rounsaville (2007, p. 851): "Do they [ESTs] work in the real world as well as the ivory tower?"

Researchers have begun to assess the effectiveness of ESTs in non-research settings. One area of concern with such evaluations, however, is the extent to which treatment in the "real-world" has been assessed under clinically representative conditions (Shadish et al., 1997; Shadish et al., 2000). Studies of clinical effectiveness range on a continuum of how clinically representative conditions are. Shadish and colleagues have shown that, often, representativeness has been associated with a lower degree of control over research methodology (Shadish et al., 2000).

To date, methodological issues have limited the available research knowledge about outcomes in clinical settings. Benchmarking is a method for assessing clinical effectiveness whereby outcomes observed in outpatient samples are compared to those obtained in research studies (McFall, 1996). Wade and colleagues (1998), who reported one of the first uses of benchmarking in the psychotherapy literature, transported cognitive therapy for panic disorder to a community mental health center. Participants seen in this mental health center were shown to have similar treatment outcomes to participants in two efficacy studies. In the depression literature, several studies have employed similar methodologies and suggested few differences in outcomes for depressed participants treated in RCTs versus clinical practice. Persons, Bostrom, and Bertagnolli (1999) reported that the outcomes of participants with depressive symptoms in a private practice who received cognitive therapy, alone or in conjunction with medication, did not differ from the outcomes of participants receiving those treatments within the context of two randomized controlled trials. One potentially important (and limiting) difference between benchmark samples and clinic samples is method of diagnosis. In this study, participants in the RCTs were diagnosed as having depression via structured diagnostic interviews whereas the private practice

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

participants were diagnosed by unstructured means, with a minimum BDI score substituting a diagnosis of major depressive disorder.

In another study of cognitive therapy in the clinic, Merrill and colleagues (2003) transported cognitive therapy to a community mental health center. They found that clinic participants evidenced rates of symptom improvement similar to participants in two published RCTs for depression. This provided an excellent indicator of the ability to transport an EST into clinical care, but as clinicians were receiving ongoing intensive supervision in the EST, the findings may not generalize to many practice settings, in which supervision is often not provided.

Minami and colleagues (Minami et al., 2007; Minami et al., 2008) compared outcomes from 35 research studies of the treatment of depression to outcomes of depressed participants treated with usual-care psychotherapy treatments within a managed care setting (the exact type of therapy was not reported). Overall, their results indicated that the average outcome in treatment-as-usual settings was similar to those observed in clinical trials. One complicating factor in interpreting these results, however, was that the authors relied on pre- to post-treatment effect sizes to draw their conclusions. Comparisons of this index between samples are affected by differences in the sample variances, such that higher indexes are achieved when pre-treatment variance is constrained, as happens when minimum severity criteria are employed in research studies.

In a large sample of participants with a range of psychiatric disorders in routine clinical practice under the National Health Service in England, Westbrook and Kirk (2005) reported that participants treated with cognitive behavior therapy responded well, on average. Comparisons were made between the outcomes observed in subsets of depressed and anxious participants and those obtained in relevant clinical trials for those disorders. For depressed participants, the mean post-treatment BDI score for participants in routine care was not different from those in an RCT, but the participants in routine care evidenced a lower recovery rate than participants in the RCT. Unfortunately, participants in the sample were not formally diagnosed, so such subset analyses must be interpreted with caution.

In order to address some of the methodological limitations in the literature to date, this study examines outcomes of CT in a non-protocol outpatient clinic setting at the Center for Cognitive Therapy (CCT) in Philadelphia. It aims to provide evidence of the effectiveness of cognitive therapy under routine clinical conditions and to expand on previous research by including a sample of participants diagnosed using structured clinical evaluations and treated with one treatment modality, cognitive therapy. Treatment in this setting met all of the conditions set forth for clinical representativeness (Shadish et al., 2000): participants present at the CCT with common psychiatric problems, and are routinely treated with CT. Clients are referred to the center by treatment providers and other health professionals as well as by word of mouth, and clinicians at the center are clinical staff (not researchers). The structure of treatment is weekly therapy sessions, as seen in typical clinical practice. Although participants complete weekly symptom measures as standard procedure and as a clinical tool, treatment and participant outcomes are not expressly monitored, and therapists are free to proceed with treatment idiosyncratically. Participants present with a variety of symptoms and disorders and are not excluded on the grounds of comorbid symptom features. Therapists are not specifically trained immediately prior to treating the participants in this sample, and research is not conducted at the CCT. Therapists are free to apply treatment interventions as they see fit – they do not adhere to a specific manual or treatment guideline, and the length of treatment is open-ended (within the constraints imposed by clients' financial concerns and by the policies of managed care and insurance companies).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In addition to providing data on clinical effectiveness for future research, a second goal of the current study is to compare the outcomes of participants treated in this clinical setting (the CCT) to a benchmark of outcomes of participants treated in a large NIMH-funded RCT of cognitive therapy versus medications for depression (DeRubeis et al., 2005). Differences between the two samples in baseline participant characteristics will be examined and controlled statistically in outcome comparisons. The DeRubeis et al. study is an ideal comparison group because it was conducted in the same setting in which the CCT operates in, adjacent to the same university campus, receiving referrals from many of the same external treatment providers. The DeRubeis et al. study was also conducted around the same time that participants in the CCT sample were seen for treatment and therapists in the study had prior training from A.T. Beck and colleagues in cognitive therapy. The number of similarities between the CCT and the DeRubeis et al. study enhances the ability to compare the two settings and provides the greatest degree of ability to make inferences about similarities or differences in treatment outcomes.

In terms of specific hypotheses, we predicted that, based on previous research, cognitive therapy under routine conditions at the CCT would be an effective treatment for depression. Further, since results of benchmarking comparisons to date have yielded little evidence of a difference between outcomes in clinical practice and RCTs, we hypothesized that there would be no difference in outcomes between the CCT and RCT sample.

Method

Participants

Diagnostic and treatment outcome information were obtained from the structured intake evaluations and subsequent weekly self-report measures of 217 participants at the CCT who had been given a primary diagnosis of Major Depressive Disorder between 1995 and 1999. This period was chosen because participants in the clinic at that time were given systematic intake evaluations that were routinely and carefully collected, and because participant charts detailing session-by-session depression severity were available. The CCT is a university-affiliated outpatient practice in Philadelphia, PA. The therapists at the CCT are Ph.D. psychologists, licensed clinical social workers, and medical residents with varying levels of training, experience and years of service at the CCT. The clinic treats individuals with a wide variety of DSM-IV Axis I and Axis II disorders. In the time period from which the sample was drawn, the CCT accepted self-payment and insurance from several private insurance companies. It also accepted Medicaid, which allowed many participants to receive treatment at no cost to them. Cognitive therapy is the primary treatment modality at the CCT.

Intake evaluations were conducted by Ph.D. level assessors using the Structured Clinical Interview for DSM-IV Diagnosis for Axis I (First and Gibbon, 2004) and the Structured Clinical Interview for DSM-IV Diagnosis Axis II (First and Gibbon, 2004). Participants also completed the Beck Depression Inventory (BDI II; Beck et al., 1996). Assessors were trained to use the assessment instruments through workshops whose cumulative duration totaled approximately 20 hours over 3 weeks and included training in administering measurements as well as consensus-rating sessions. A Ph.D.-level supervisor regularly oversaw all assessments and diagnostic procedures.

Participants at the CCT typically received weekly cognitive therapy sessions throughout treatment, with the frequency of sessions varying as a function of symptom severity levels, schedules, and financial or insurance-related issues. Prior to the start of each session, all participants were asked to complete weekly symptom measures including the BDI. For this study, the BDI scores for each session were collected from all therapy charts. Because of the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

routine nature of data collection, therapists in this study did not conduct treatment with the idea that their outcomes would be monitored as they are in a clinical trial. Therapists could note in a participant's chart whether they were taking medication or receiving concurrent treatment while in therapy at CCT, but there was not a systematic method for doing so. Such data were therefore collected when they were available.

Upon intake, participants entering the CCT provided informed consent allowing their de-identified medical records to be used for research purposes in a standard IRB-approved procedure for all participants within the University's healthcare system. Participant data in the current study was collected and examined within a protocol approved by the same IRB. Data for the current study were collected from therapy charts by the first author as well as two research assistants who were blind to study hypotheses.

RCT Benchmark: DeRubeis et al. (2005) RCT

Data was obtained from the DeRubeis et al. (2005) study (Total $N = 240$, Cognitive Therapy $n = 60$). In order to maximize the ability to detect differences, a comparison of intake characteristics between participants at the CCT and in the DeRubeis et al. study utilized the full sample of participants from the DeRubeis et al. study ($N = 240$). In comparing treatment outcomes between the two settings, only participants assigned to the cognitive therapy arm of the study were considered ($n = 60$). These participants had been randomly assigned to cognitive therapy and received 16 weeks of treatment. Sessions were conducted twice weekly for four weeks of treatment, once or twice weekly for the next eight weeks and then once weekly for the final four weeks. Participants in this study completed the BDI at intake and at each treatment session. Their scores on this measure could therefore be compared to those of participants at CCT across treatment. This was primarily achieved using hierarchical linear modeling (HLM) as described below.

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

The BDI-II is a revision of the original self-report Beck Depression Inventory (BDI; Beck et al., 1961). It contains 21 items and the total score for the measure is found by summing the item scores, with higher scores reflecting more severe depressive symptomatology. The internal consistency of the BDI-II has been shown to be high (Beck et al., 1996; Whisman et al., 2000).

Analytic Strategy

Clinical significance—Analyses of clinically significant change on the BDI were conducted according to Jacobson and colleagues' formulae (e.g. Jacobson and Truax, 1991; Jacobson, et al., 1999). This method evaluates two criteria for each participant. The first is whether the 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), cited in Kendall and Sheldrick (2000). 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 the following formula (Jacobson's criterion ‘c’):

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$c=\frac{s_0M_1+s_1M_0}{s_0+s_1}$$

where M_1 is the mean pretreatment BDI score of participants at the CCT, s_1 is the standard deviation of this CCT mean, M_0 is the mean BDI score of the normative sample, and s_0 is the standard deviation for that sample.

Longitudinal comparison of treatment outcome to DeRubeis et al. (2005)— 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).

There are several complicating factors in comparing the DeRubeis et al. (2005) sample and the CCT sample. The treatment protocol in the RCT called for all participants to receive twice weekly therapy sessions for the first four weeks of treatment and to transition to weekly sessions thereafter. Because the more typical course of therapy at CCT is weekly sessions throughout treatment, the frequency of sessions between the two treatment settings is unequal. In addition, in the CCT sample, therapy did not have a fixed endpoint and could be extended (to include more sessions) if a therapist and participant believed it would be useful (e.g., if the participant's improvement was modest, and it was judged that he or she could benefit from more treatment, or conversely, if a participant was making good progress and desired more sessions to consolidate his or her learning). Because of these design constraints, 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 treatment or the other. Therefore, we conducted two primary analyses, each of which controlled for one of these confounds. In one analysis, we use all longitudinal data with a cutoff of 15 weeks. Fifteen weeks was considered the endpoint of treatment in these analyses and all data after this time point was discarded. This choice of endpoint corresponds to one week fewer than the treatment protocol the RCT called for, thus, treatment was not complete in either setting at 15 weeks. By using this as the final week of treatment, we sought to mitigate the potential effects of the planned termination in the RCT. Still, in this analysis, the RCT participants are likely to have received more sessions than the CCT participants, and therefore, this analytic strategy might be expected to bias results in favor of the RCT. In the second analysis, we examined outcomes using longitudinal models wherein the maximum number of sessions is fixed at 20. In this analysis, it did not matter how much chronological time elapsed until the participant received 20 sessions, all data was used up to and including 20 sessions only. Twenty sessions was chosen as a cutoff because it is a frequently used

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

maximum number of sessions of cognitive therapy for depression in RCTs (e.g., Elkin et al., 1989; Hollon et al., 1992; Jarrett et al., 1999). Given that more time would be expected to pass before participants in the CCT could receive 20 sessions, and given the possibility that depressive symptoms might remit on their own for some individuals, this choice of analysis would be expected to bias the results in favor of the CCT.

Results

Characteristics of CCT Sample

Table 1 displays the characteristics of the sample of the CCT participants. Fifty-nine percent of the CCT participants were females and 83% were white. Their mean age was 34.5 years (*SD* = 11.8). Their mean pretreatment BDI score of 26.4 (*SD* = 9.9) falls in the moderate to severe range (Beck, Steer, & Brown, 1996). In the therapy charts were indications that approximately 40% of participants were taking psychiatric medications during the course of treatment. The CCT participants attended an average of 15.9 (*SD* = 16.2) treatment sessions, with a range of 0 (participants who completed an intake and never entered treatment) to 97. The modal number of sessions was 1 (*n* = 16 participants) and the median was 11 (with lower and upper quartiles 4 and 21, respectively).

Symptom Change During Treatment

In order to assess CCT participants at posttreatment, the last recorded BDI score for each participant was used irrespective of the number of sessions the participant attended. A significant reduction in depressive symptoms was observed over the course of treatment (mean posttreatment BDI = 15.9, *SD* = 14.0; t-test of difference from intake scores: $t = 8.94$, $p < .001$).

Clinically Significant Change

Clinically significant change in depressive symptoms as measured on the BDI was assessed. As noted, participants were considered to have demonstrated reliable improvement if their RCI equaled 1.96 or greater. Using this criterion, 61% of the CCT participants experienced reliable improvement over the course of treatment (8% were shown to have 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 15. Thus, if participants had demonstrated reliable change *and* finished treatment with a BDI score below 15, they were considered recovered. Only participants who began treatment with a BDI score of 15 or above (*n* = 175) were included in this analysis, as those with scores lower than 15 did not have the opportunity to meet both criteria. Seventy-nine participants (45% of those who met the first criteria, 36% of total sample) were found to be below the cutoff and therefore considered "recovered."

Comparison of CCT and DeRubeis et al. (2005) Sample

The characteristics of the CCT and RCT samples are compared in Table 1. Participants at the CCT were significantly younger than those in the DeRubeis et al. (2005) study, $\chi^2(1) = 4.94$, $p < .001$, $d = 0.46$. The CCT sample contained a significantly higher proportion of participants with Axis II diagnoses, $\chi^2(1) = 7.14$, $p < .01$, but the RCT sample contained higher rates of diagnosed co-occurring substance abuse $\chi^2(1) = 12.35$, $p < .001$, double depression $\chi^2(1) = 3.85$, $p < .05$, and recurrent depression $\chi^2(1) = 16.80$, $p < .001$. The DeRubeis et al. sample also had a higher mean intake BDI score, $t(455) = 5.28$, $p < .001$. The two samples did not differ in terms of percentage of participants who were female, Caucasian, married, unemployed, or who had Axis I comorbidity. The mean number of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

sessions attended by participants did not differ between the two samples, $t(275) = 1.61$, $p = .11$.

Longitudinal Comparison of DeRubeis et al. (2005) and CCT Data

Comparison at 15 weeks—BDI scores of participants in the DeRubeis et al. (2005) study were compared with BDI scores of participants at the CCT at 15 weeks using a hierarchical linear model (HLM). BDI score (obtained at each session) was the dependent variable and the independent variable of interest was treatment setting (i.e., RCT or the CCT). Since the two groups were shown to have differences in depressive symptoms prior to treatment, grand-mean-centered intake BDI scores were used as a covariate. All interactions were tested. The three-way interaction between treatment setting, intake BDI, and time was not significant, $F(1, 147) = 0.44$, $p = .51$ and was therefore removed from the model. In the resulting model, at 15 weeks there was no statistically significant difference in estimated mean BDI scores for the RCT versus CCT participants (RCT mean = 10.2, $SE = 1.8$; CCT mean = 13.6, $SE = 1.1$; $F(1, 182) = 2.64$, $p = .11$). There was, however, a significant interaction between treatment setting and intake BDI score, $F(1, 223) = 11.26$, $p < .001$. For participants with levels of symptoms at intake that fell one standard deviation below the grand mean, treatment outcomes were roughly equivalent between DeRubeis et al. and CCT (Figure 1). For participants with mean levels of symptoms at intake, those treated within DeRubeis et al. had better treatment outcomes than those treated at CCT. For participants with levels one standard deviation above the mean, this pattern was even more pronounced.

As reported above, participants in the two settings differed with respect to the rates of Axis II comorbidity, substance abuse comorbidity, recurrent depression and double depression. To control for the effects of these variables, they were added as covariates (using unweighted effects coding) to the HLM model described above. The interaction of each covariate with intake BDI, time and both intake BDI and time were also entered. Covariate terms with p-values above 0.50 were removed and the model was re-run. The resulting model indicated that treatment setting predicted week 15 scores at a trend level, $F(1, 190) = 3.08$, $p = .08$, with participants in the RCT exhibiting superior outcomes to those at the CCT. As in the previous analysis, there was a significant interaction between treatment setting and intake BDI score $F(1, 219) = 6.50$, $p = .01$, in the same direction as the one described above.

Comparison at 20 sessions—BDI scores of participants in DeRubeis et al. (2005) were compared with BDI scores of participants in CCT at the 20th session of treatment, using HLM. As before, centered intake BDI scores were used as a covariate and all interactions were tested. The three-way interaction between treatment setting, intake BDI, and time was not significant, $F(1, 109) = 0.04$, $p = .29$, and was removed from the model. In the resulting model, at 20 weeks there was not a significant difference in estimated mean BDI scores (RCT mean = 10.6, $SE = 2.1$; CCT mean = 9.9, $SE = 1.3$; $F(1, 130) = 0.09$, $p = .77$). Also convergent with the data at 15 weeks, there was a significant interaction between treatment setting and intake BDI score, $F(1, 236) = 6.35$, $p = .01$. The direction of the effect was similar to that observed at 15 weeks (Figure 2). Specifically, participants with low levels of intake depressive symptoms were predicted to have better treatment outcomes when treated at CCT than in the RCT. Those with more moderate symptoms at intake reported similar treatment outcomes in the two settings. Those with a greater degree of intake symptom levels displayed better treatment outcomes if treated in the RCT than at CCT.

The model of the BDI scores at 20 sessions was expanded to include the four potentially confounding covariates described for the week 15 analysis. There was no main effect of treatment setting $F(1, 131) = 0.00$, $p = .97$, but again there was an interaction between

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

setting and intake BDI score $F(1, 234) = 6.10$, $p = .01$, in the same direction as the one described above.

Secondary Analysis of the Effects of Medication

As previously noted, antidepressant medication was not offered in the cognitive therapy arm of the RCT, whereas participants at the CCT did have the option of pursuing concurrent pharmacotherapy outside the clinic. Forty percent of the CCT participants were documented to have been receiving concurrent medication. The effect of medication on treatment outcomes for participants at the CCT was examined. At 15 weeks, the mean BDI score for the CCT participants who had received medication was 18.4 (*SD* = 14.2), compared to 15.4 (*SD* = 13.0) for those who had not. A general linear model was applied predicting 15-week BDI score from medication status with intake BDI score as a covariate. Medication status did not predict treatment outcome, $F(1, 191) = 0.69$, $p = .41$, nor did it do so in similar analyses of BDI scores at 20 sessions.

Discussion

Depressed individuals treated with cognitive therapy in a routine clinical care setting were found to show significant improvement in symptoms over the course of treatment, with 61% of participants demonstrating reliable improvement, and of these, 45% (or 36% of the total sample) considered to be recovered at the end of treatment. These outcomes are similar or superior to those that have been reported in other non-research settings. For example, Persons and colleagues found that 57% of private practice treatment completers demonstrated reliable improvement and ended treatment with scores that were within a functional distribution, and Merrill and colleagues reported that 48% of their community sample treated with CT evidenced reliable improvement. Westbrook and Kirk (2005) report that 52% of participants with a likely diagnosis of depression experienced clinically significant change and that 36% of these participants could be considered recovered. However, this study moves beyond previous research reports by using a sample of participants diagnosed via structured diagnostic interviews, as is the norm in RCTs. It provides evidence in regard to one specific disorder, in a clinic specializing in cognitive therapy under routine conditions. We hope that these results will provide a benchmark for future use of the effectiveness of cognitive therapy for depression in routine care.

When a detailed analysis was performed comparing treatment outcomes at the CCT with data obtained from a large RCT for depression (DeRubeis et al., 2005), there was little evidence of the superiority of either setting overall. One virtue of these analyses is that we were able to control for differences between the participant samples. The results of these analyses suggested that overall, symptom levels at the end of treatment did not differ between the two settings. However, our analyses also indicated that RCT participants with higher symptom severity at intake improved more than did their counterparts in the clinic sample. There was a less robust indication that participants with lower symptom severity at intake fared somewhat better in the clinic, relative to the RCT.

There are several limitations of the current study, some of which are unavoidable when an uncontrolled outpatient sample is employed. One limitation is that we examined only one outcome measure and only one form of treatment at one outpatient clinic. It may be that other forms of treatment do not show similar levels of effectiveness in the real world, or that cognitive therapy would perform differently in another setting or under different conditions. Although the CCT appears to be a clinically representative setting, due to its emphasis on cognitive therapy and high levels of therapist training, as well as sophisticated assessment methods, it may not fully represent other outpatient clinics. In addition, treatment outcome was assessed using the BDI only. It is possible that different outcomes may obtain in future

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

studies that use non-self report measures, however, the use of more time-intensive, interviewer-administered measures may not be feasible in applied clinical settings like the CCT. Further, this research lacked a control group against which to compare the CCT outcomes. When outcomes were compared with those obtained in an RCT, it was not possible to equate the frequency of sessions or overall treatment length. Additionally, the DeRubeis et al. (2005) RCT excluded participants with mild depression severity, so that the two samples were not well matched in regard to their severity distributions.

A related limitation is that the participant samples may not have been large enough to detect all differences between the treatment settings. Therapist adherence was not assessed at CCT, whereas the RCT therapists were aware that adherence was monitored via a review of video recordings of their sessions. Although the CCT is a cognitive therapy clinic, it is unknown whether therapists strictly adhere to a cognitive model or whether they incorporate techniques from other therapeutic traditions. Further, the frequency and total number of therapy sessions was uncontrolled. This may have had an effect on the comparisons drawn between the CCT and the RCT. For example, RCT therapists and participants may have worked together more efficiently knowing that they had limited time to accomplish their goals (Rubens, 1983). Additionally, the use of medications was not controlled or standardized in the CCT sample. Although the results of secondary analyses suggested that medications did not play a role in outcomes, these may have exerted some influence on the delivery of the treatment. Such factors are inherent in gathering information from a sample with conditions that are clinically representative, but they must be taken into account when examining the conclusions of the research.

There are several directions that would be interesting for future investigation. First, researchers should attempt to engage larger and more diverse samples in order to maximize power to detect differences, as well as take the additional step of testing for equivalence of treatment outcomes (Jones et al., 1996). Second, it would be helpful to examine a larger range of patient outcomes. One possibility is to examine change in the cognitive-affective and somatic factors of the BDI (Whisman, Perez & Ramel, 2000). The BDI score changes found in this study may be due to a change in one or both of these factors. To date, evidence suggests that cognitive change is often demonstrated in both cognitive therapy and medication treatments (Hollon, 2006), but changes in the two factors could potentially vary across sites or according to adjunctive medication status.

There are other areas that would be of interest to investigate as well. These include patient expectations of treatment, functional status, degree of hopelessness, and patient and therapist alliance. Due to the naturalistic setting in this study, such information was not available, but would likely be a fruitful area for future research. It would also be interesting to investigate the effect of a treatment manual on treatment delivery and outcome. This could include examining the alliance in each setting, and looking for differential relationships between the alliance and outcome in each. For instance, Carroll, Nich and Rounsaville (1997) found that patients treated in an active manualized treatment condition had greater alliance scores than those in a control treatment, but that the alliance was more strongly linked to outcome in the control treatment. The effect of imposing a time limit on treatment could also be further investigated. Therapists and patients may be more likely to enact changes more rapidly when there is a limited amount of time, particularly if this design feature is known in advance and factored into the treatment plan (Reynolds, Stiles, Barkham, Shapiro, Hardy, & Rees, 1996). Finally, this study did not examine therapist and client alliance in the two settings, and a comparison of therapeutic relationships in the two settings would be of great interest.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Implications

These results suggest that, in general, the effectiveness of cognitive therapy for depression when administered under routine clinical conditions can be comparable to that found in RCTs. Results of the comparison of outcomes between the CCT and the DeRubeis et al. (2005) RCT suggest that for participants with low levels of symptoms at the start of treatment, CT in routine care may produce similar or better outcomes than CT as delivered in an RCT. However, for participants with moderate to severe symptoms, outcomes may be better when treatment is delivered within an RCT than when delivered in this outpatient setting. This suggests that clinicians treating participants with moderate to severe symptoms may benefit from making modifications to treatment in order to attain outcomes similar to those evidenced in RCTs. It is possible that gains could be incurred for the more severe participants by structuring treatments to more closely resemble those in RCTs. For instance, increased consultation and supervision, as well as periodic assessments of adherence to therapeutic modality may yield important benefits. Participants, in particular those with high levels of symptom severity, may also evidence enhanced outcomes when given more frequent sessions at the start of treatment. Future research, employing strategies that maximize clinical representativeness, should continue to focus on the impact of these factors on treatment outcomes.

Acknowledgments

We thank Alexandra Sibley and Catherine Schafer for their help with data entry. We also thank Kathleen Carroll for her helpful comments on the manuscript.

Role of Funding Source: This research and manuscript preparation was supported by grants MH47383 (Dr. Beck), K99MH080100 (Dr. Stirman) and MH50129 (R10) (Dr. DeRubeis) and MH55875 (R10) and MH01697 (K02) (Dr. Hollon) from the National Institute of Mental Health, Bethesda, MD and by grant R49/CCR316866 (Dr. Beck) by the Center for Disease Control. The NIMH and CDC had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

References

Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of beck depression inventories-IA and -II in psychiatric outpatients. J Pers Assess 1996;67(3):588–597. [PubMed: 8991972]

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571. [PubMed: 13688369]

Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: A review of meta-analyses. Clin Psychol Rev 2006:17–31. [PubMed: 16199119]

Carroll KM, Nich C, Rounsaville BJ. Contribution of the therapeutic alliance to outcome in active versus control psychotherapies. J Consult Clin Psychol 1997;65:510–514. [PubMed: 9170775]

Carroll KM, Rounsaville BJ. A vision of the next generation of behavioral therapies research in the addictions. Addiction 2007;102(6):850–862. [PubMed: 17523974]

Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol 1998;66(1):7–18. [PubMed: 9489259]

Cohen J. A power primer. Psychol Bull 1992;112:155–159. [PubMed: 19565683]

Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008;76(6):909–922. [PubMed: 19045960]

DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive therapy vs. medications in the treatment of moderate to severe depression. Arch Gen Psychiatry 2005;62(4):409–416. [PubMed: 15809408]

Dozois D, Dobson K, Ahnberg J. A psychometric evaluation of the Beck Depression Inventory-II. Psychol Assess 1998;10:83–89.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Elkin I, Shea MT, Watkins JT, Imber SD. National institute of mental health treatment of depression collaborative research program: General effectiveness of treatments. Arch Gen Psychiatry 1989;46(11):971–982. [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, US: John Wiley & Sons Inc.; 2004.

Gloaguen J, Cottraux M, Cucherat, Blackburn I. A meta-analysis of the effects of cognitive therapy in depressed patients. J Affect Disord 1998;49:59–72. [PubMed: 9574861]

Goldfried MR, Wolfe BE. Toward a more clinically valid approach to therapy research. J Consult Clin Psychol 1998;66(1):143–150. [PubMed: 9489268]

Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, et al. Cognitive therapy and pharmacotherapy for depression: Singly and in combination. Arch Gen Psychiatry 1992;49(10):774–781. [PubMed: 1417429]

Hollon, SD. Cognitive therapy in the treatment and prevention of depression. In: Joiner, TE.; Brown, JS.; Kistner, J., editors. The interpersonal, cognitive, and social nature of depression. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2006. p. 133-151.

Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. J Consult Clin Psychol 1999;67(3):300–307. [PubMed: 10369050]

Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59:12–19. [PubMed: 2002127]

Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC. Treatment of atypical depression with cognitive therapy or phenelzine: A double-blind, placebo-controlled trial. Arch Gen Psychiatry 1999;56(5):431–437. [PubMed: 10232298]

Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36–39. [PubMed: 8664772]

Kendall PC, Sheldrick RC. Normative data for normative comparisons. J Consult Clin Psychol 2000;68:767–773. [PubMed: 11068962]

McFall, RM. Consumer satisfaction as a way of evaluating psychotherapy: Ecological validity and all that versus the good old randomized trials (panel discussion). 6th Annual Convention of the American Association of Applied and Preventative Psychology; San Francisco. 1996 Jul 1.

Merrill KA, Tolbert VE, Wade WA. Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. J Consult Clin Psychol 2003;71(2):404–409. [PubMed: 12699035]

Minami T, Wampold BE, Serlin RC, Hamilton EG, Brown GS, Kircher JC. Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. J Consult Clin Psychol 2008;76(1):116–124. [PubMed: 18229989]

Minami T, Wampold BE, Serlin RC, Kircher JC, Brown GS. Benchmarks for psychotherapy efficacy in adult major depression. J Consult Clin Psychol 2007;75(2):232–243. [PubMed: 17469881]

Persons JB, Bostrom A, Bertagnolli A. Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. Cog Ther Resear 1999;23(5):535–548.

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. [PubMed: 8698952]

Rubens R. An independent share in the work: Some thoughts on time-limited psychotherapy. A response to Moss New Ideas Psychol 1983;1:177–182.

Seligman M. The effectiveness of psychotherapy: The Consumer Reports study. Am Psychol 1995;50:965–974. [PubMed: 8561380]

Shadish W, Matt G, Navarro A, Siegle G, Crits-Christoph P, Hazelrigg M, et al. Evidence that therapy works in clinically representative conditions. J Consult Clin Psychol 1997;65:355–365. [PubMed: 9170759]

Shadish W, Matt G, Navarro A, Phillips G. The effects of psychological therapies under clinically representative conditions: a meta-analysis. Psychol Bull 2000;126:512–529. [PubMed: 10900994]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Strunk D, DeRubeis R. Cognitive therapy for depression: A review of its efficacy. J Cog Psychother 2001;15:289–297.

Wade W, Treat T, Stuart G. Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. J Consult Clin Psychol 1998;66:231–239. [PubMed: 9583326]

Westbrook D, Kirk J. The clinical effectiveness of cognitive behaviour therapy: outcome for a large sample of adults treated in routine practice. Behav Res Ther 2005;43:1243–1261. [PubMed: 16086979]

Westen D, Novotny CM, Thompson-Brenner H. The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. Psychol Bull 2004;130(4):631–663. [PubMed: 15250817]

Whisman M, Perez J, Ramel W. Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. J Clin Psychol 2000;56:545–551. [PubMed: 10775046]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

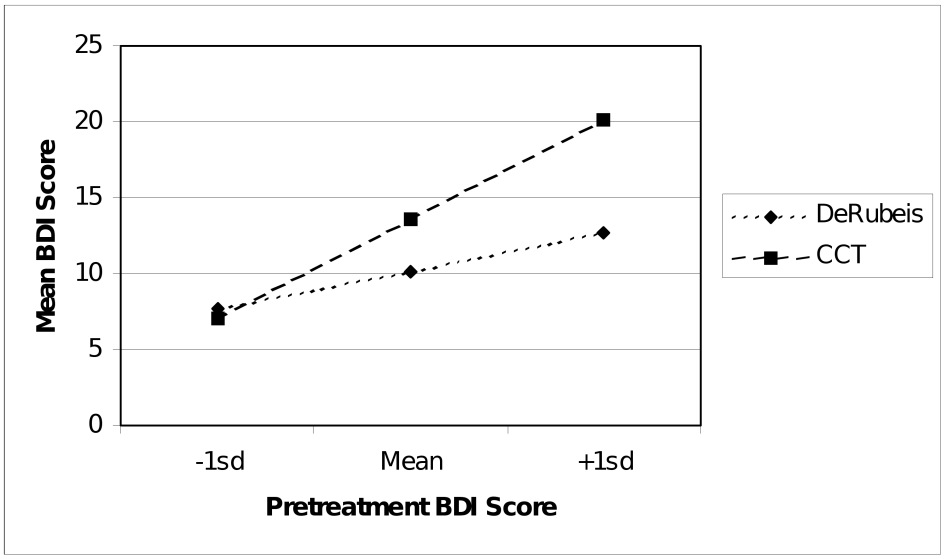

Figure 1.
Intake BDI by treatment setting interaction at 15 weeks.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

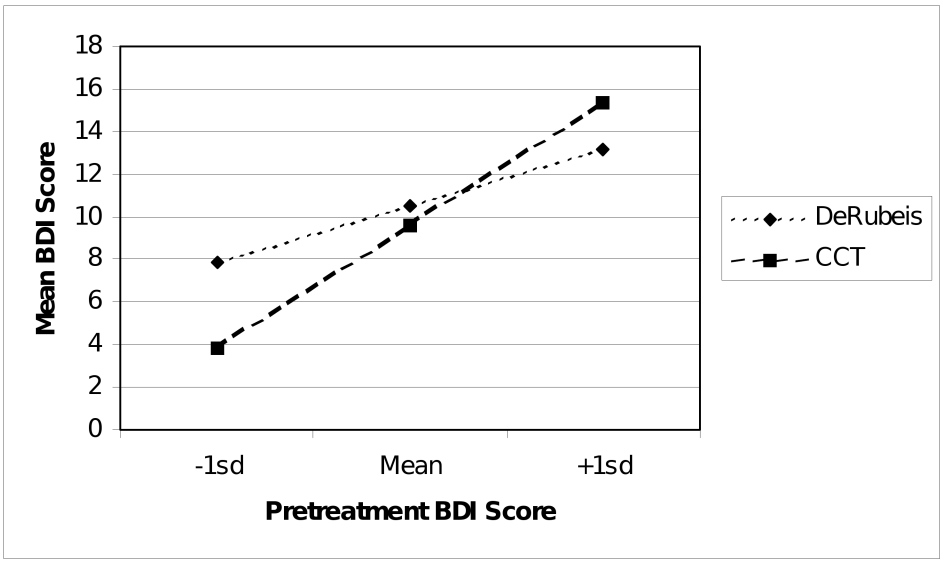

Figure 2.
Intake BDI by treatment setting interaction at 20 sessions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Demographic and Diagnostic Characteristics of Participants Treated at CCT and in the DeRubeis et al. (2005) Study

Factor of Comparison	CCT N=217	DeRubeis et al. N=240	t / χ^2	p
Mean Pretreatment BDI Score (SD)	26.4 (9.9)	30.9 (8.6)	5.28	<.0001
Mean Age (SD)	34.5 (11.8)	40 (12)	4.94	<.0001
Female, %	59	59	0.01	.97
Minority, %	17	16	0.01	.90
Unmarried, %	70	67	0.33	.56
Unemployed, %	18	18	0.00	.99
Axis I Comorbidity	70	72	0.14	.71
Axis II Comorbidity, %	61	48	7.14	<.01
Substance Disorders, %	20	35	12.35	<.001
Double Depression, %	17	25	3.85	.049
Recurrent Depression, %	75	90	16.80	<.0001