Long-Term Follow-Up of a Randomized Controlled Trial Comparing Acceptance and Commitment Therapy and Standard Cognitive Behavior Therapy for Anxiety and Depression

Evan M. Forman  
Jena A. Shaw  
Elizabeth M. Goetter  
James D. Herbert  
Jennie A. Park  
Drexel University  
Erica K. Yuen  
Medical University of South Carolina

The present study represents one of the first comparisons of the long-term effectiveness of traditional cognitive behavior therapy (i.e., Beckian cognitive therapy; CT) and acceptance and commitment therapy (ACT). One hundred thirty-two anxious or depressed outpatients were randomly assigned to receive either CT or ACT, and were assessed at posttreatment \( n = 90 \) and at 1.5-year \( n = 91 \) follow-up. As previously reported, the two treatments were equivalently effective at posttreatment according to measures of depression, anxiety, overall (social/occupational/symptom-related) functioning, and quality of life. However, current results suggest that treatment gains were better maintained at follow-up in the CT condition. Clinical significance analyses revealed that, at follow-up, one-third more CT patients were in the clinically normative range in terms of depressive symptoms and more than twice as many CT patients were in the normative range in terms of functioning levels. The possible long-term advantage of CT relative to ACT in this population is discussed.

Keywords: acceptance-based behavior therapy; psychotherapy outcome; depression; anxiety; long-term follow-up

The term cognitive behavioral therapy (CBT) reflects a broad collection of evidence-based approaches that have become the most widely utilized and researched of all psychotherapeutic methods (Norcross, Hedges, & Castle, 2002), with Beckian cognitive therapy (CT; Beck, 1991) representing the most widely used and empirically supported form of CBT (Butler, Chapman, Forman, & Beck, 2006; Hofmann & Smits, 2008). A newer subcategory of CBT, sometimes referred to as acceptance-based behavior therapies, has risen to prominence in recent years. Examples include mindfulness-based cognitive therapy (MBCT; Z. V. Segal, Williams, & Teasdale, 2002), mindfulness-based stress reduction (Kabat-Zinn, 1990), acceptance-based behavior therapy for generalized anxiety disorder (Roemer & Orsillo, 2005), dialectical behavior therapy (DBT; Linehan, 1993), and acceptance and commitment therapy (ACT; Hayes,
Strosahl, & Wilson, 1999), among others. Of these, ACT has received the most attention in terms of empirical study (Hayes, Levin, Plumb, Boulanger, & Pistorello, in press; Hayes, Luoma, Bond, Masuda, & Lillis, 2006) and scientific debate (Arch & Craske, 2008; Corrigan, 2001, 2002; Gaudiano, 2009a, 2009b; Hayes, 2002, 2008; Hayes et al., in press; Herbert & Forman, in press, 2011; Hofmann & Asmundson, 2008; Hofmann & Asmundson, in press; Öst, 2008, 2009). At the level of technology, there are some important differences in how ACT and CT treat psychopathology (Forman & Herbert, 2009). CT makes use of cognitive disputation and other forms of reappraisal (including behavioral experiments) designed to correct systematic biases in information processing, with the goal of reducing symptom intensity (Beck, Rush, Shaw, & Emery, 1979). The goal of ACT is not symptom reduction per se, but helping patients to “accept” difficult internal experiences (thoughts, images, emotions, sensations) in the service of engaging in values-consistent behavior change.

ACT has demonstrated preliminary effectiveness across a range of problem behaviors, including mood (Zettle & Hayes, 1986) and anxiety (Block, 2003; L. A. Brown et al., 2011; Dalrymple & Herbert, 2007; Roemer, Salters-Pedneault, & Orsillo, 2006; Twohig, Hayes, & Masuda, 2006) disorders, psychosis (Bach & Hayes, 2002; Gaudiano & Herbert, 2006), polysubstance abuse (Hayes, Wilson, et al., 2004), and smoking (R. A. Brown, Lejuez, Kahler, Strong, & Zvolensky, 2005; Hernández-López, Luciano Soriano, Bricker, Roales-Nieto, & Montesinos Marin, 2009; Hernandez Lopez, Roales Nieto, Luciano Soriano, & Montesinos Marin, 2005), among others. A meta-analysis (Hayes, et al., 2006) reported ACT to be superior to active treatments, including standard CBT. However, Öst (2008) has criticized the rigor of the trials on which the meta-analysis relied, and a subsequent meta-analysis (Powers, Zum Vörde Sive Vörding, & Emmelkamp, 2009) concluded that ACT was equally effective as established treatments (but see also Gaudiano, 2009a). The RCT upon which the current study was based similarly detected no differences in efficacy between ACT and CT at posttreatment in the treatment of depression and anxiety (Forman, Herbert, Moitra, Yeomans, & Geller, 2007).

It is important to examine longer-term effectiveness of psychotherapies, as immediate effects can fade over time, patients might take time to master skills learned in treatment, and treatments that are equivalent at one time point can diverge later (Gifford et al., 2004; Lappalainen et al., 2007). Generally speaking, standard CBT has demonstrated longer-term efficacy (Butler et al., 2006; Gloaguen, Cottraux, Cucherat, & Blackburn, 1998; Shapiro et al., 1994), though several recent reviews suggest effects weaken considerably after 1 year (e.g., Durham, Higgins, Chambers, Swan, & Dow, 2011). Open trials and trials comparing ACT to nonactive treatments or treatment-as-usual support ACT’s lasting benefits for anxiety (Dalrymple & Herbert, 2007; Osman, Wilson, Storaasli, & McNeill, 2006; Twohig, 2008; Zettle, 2003), depression (Blackledge & Hayes, 2006), trichotillomania (Woods, Wetterneck, & Flessner, 2006), psychosis (Bach & Hayes, 2002; Gaudiano & Herbert, 2006), substance dependence (Hayes, Wilson, et al., 2004; Stotts, Masuda, & Wilson, 2009; Twohig, Shoenberger, & Hayes, 2007), smoking (Gifford et al., 2004), obesity (Forman, Butryn, Hoffman, & Shoenberger, & Hayes, 2007; Lillis, Hayes, Bunting, & Masuda, 2009; Tapper et al., 2009), and chronic pain (Dahl, Wilson, & Nilsson, 2004; McCracken, MacKichan, & Eccleston, 2007; Vowles & McCracken, 2008). However, the follow-up periods of these studies tended to be relatively short (i.e., 1 to 3 months), limiting the extent to which conclusions can be drawn about longer-term effects. Moreover, few of these trials compared ACT to another well-established, active intervention, and none compared ACT to traditional CBT.

The current study compares long-term (18-month follow-up) outcomes from a group of outpatients randomly assigned to receive either CT or ACT. The study is an extension of an earlier report of posttreatment outcomes (Forman, Hebert, et al., 2007). As mentioned, the effects of standard CBT may attenuate in the longer-term (Durham et al., 2011), and some proponents have hypothesized that ACT might hold certain advantages over standard (i.e., Beckian) CBT because ACT is argued to be more tightly linked to basic research on its mechanisms and underlying theory (Hayes, 2008; Hayes et al., in press). ACT proponents have also questioned the putative mechanisms of change of CT (i.e., modification of the content of dysfunctional cognitions; Hayes, 2008; Hayes, Villatte, Levin, & Hildebrandt, 2011). On the other hand, standard CBT has proven long-term efficacy (Butler et al., 2006; Gloaguen et al., 1998; Shapiro et al., 1994). Perhaps an overriding consideration is the accumulating evidence suggesting that it is the behavioral elements of treatment that represent the mechanisms of action, and that other components are superfluous (Dimidjian et al., 2006; Longmore & Worrell, 2007). Thus, no specific hypotheses were made regarding differential long-term effectiveness of the two treatments. Given previous findings that mindfulness and acceptance variables moderate the impact of treatment, we tentatively hypothesized such a moderation effect at follow-up.
Method
Participants were 132 nonbaccalaureate health science students presenting for treatment at a student counseling center (Figure 1). Inclusion criteria were purposefully broad to maximize external validity. Individuals were eligible if they were seeking individual psychotherapy (as opposed to, for example, couples therapy, or study skills training), reported at least moderate levels of anxiety and/or depression (i.e., score of 9 or above on the Beck Depression Inventory [BDI-II; Beck, Steer, & Brown, 1996] or Beck Anxiety Inventory [BAI; Beck, Epstein, Brown, & Steer, 1988]),1 were fluent in English, and were not psychotic.

Participants ranged in age from 18 to 52 (M = 26.7, SD = 6.4). The majority were female (79.5%) and white (70.7%); half (50.7%) were single, and a small number were taking psychiatric medication (16.7%).2 Diagnoses were obtained by a structured interview, the Mini-International Neuropsychiatric Interview (MINI), which is a widely used diagnostic interview with shorter administration times, but comparable reliability and validity (Lecrubier et al., 1997; Sheehan et al., 1998), to the Structured Clinical Interview for DSM Disorders (SCID; First, Williams, Spitzer, & Gibbon, 2007). The large majority of patients met criteria for anxiety disorders (49.2%; 24.2% generalized anxiety disorder, 12.1% anxiety disorder NOS, 5.3% panic disorder, 5.3% social anxiety disorder, 3.1% OCD, 2.3% specific phobia, 1.5% PTSD) and/or depressive disorders (37.1%; 22% recurrent MDD, 9.8% depressive disorder NOS, 7.6% single episode MDD, 3.0% dysthymia). Other diagnoses included adjustment disorders (6.8%) and eating disorders (3.8%).

1 We elected to use a cutoff score of ≥ 9 on the BDI-II and BAI to indicate elevated levels of depression and anxiety, respectively, based on data examining normative data for adult samples on the BDI-II (D. L. Segal, Coolidge, Cahill, & O’Riley, 2008; Whisman, Perez, & Ramel, 2000) and BAI (Gillis, Haaga, & Ford, 1995).

2 Seven participants who reported taking psychiatric medications at baseline reported that they no longer were taking medications at follow-up. As a check, the analyses described below were repeated without these 7 participants and results were equivalent.
diagnostic category (depressive disorder, anxiety disorder, both depressive and anxiety disorders or neither depressive nor anxiety disorder) was established for analytical purposes. Diagnostic assessors had received 4 hours of training on the MINI, along with a minimum of 3 hours of supervised practice. Additionally, diagnostic interviews were recorded and were reviewed by doctoral-level supervisors.

**PROCEDURE**

After providing consent to participate in the study, patients were assigned to either the ACT ($n=69$) or CT ($n=63$) condition via stratified block randomization determined by total score on the Outcome Questionnaire (OQ; Lambert, Hansen, et al., 1996). As the course of therapy was intended to be naturalistic, there was not a standardized number of sessions for either condition; termination was based on mutual consensus between the therapist and patient. However, the average number of sessions across conditions was similar for ACT (18.10, $SD=15.91$) and CT (16.37, $SD=13.64$; $t(130)$, $p=.50$). Both interventions included psychoeducation, coping skills development, and behavioral (especially behavioral exposure) exercises that were presented within the theoretical framework of the assigned intervention. CT incorporated discussion of automatic thoughts, core beliefs, and schemas, identification of cognitive distortions, cognitive disputation, and cognitive restructuring, whereas ACT emphasized experiential acceptance, mindfulness training, clarification of personal values, and willingness to experience internal distress for the sake of living consistently with one’s values.

Data were collected at baseline, posttreatment, and at approximately 18-month $^3$ follow-up. Individuals were paid $20.00 for completing the follow-up procedures, but not for any of the other assessments.

**Therapist Allegiance, Fidelity, and Competence**

Therapists ($n=28$) were doctoral-level graduate students who received specialized training (a 30-hour workshop) and weekly supervision in both ACT and CT. The training and supervision sessions were led by licensed clinical psychologists with several years’ combined experience in ACT and CT. Therapists administered treatments to both ACT and CT patients. As might be expected given the dual training model, no significant differences in therapist allegiance were detected in terms of which treatment therapists judged to be more effective (ACT=53%; $\chi^2=1.29$, $p=.26$) or preferred (CT=61%; $\chi^2=.14$, $p=.71$). Session-by-session treatment manuals were not utilized; instead, therapists followed general ACT or CT treatment outlines that were specific to diagnosis. All sessions were audio recorded and a random selection of 3 sessions per study participant were rated using the Drexel University ACT and CT Adherence and Rating Scale (McGrath, Forman, & Herbert, 2009), revealing that adherence, contamination, and competence were statistically equivalent between groups. Specifically, an average of 36.2% (ACT) and 33.3% (CT) of time was spent on treatment-specific components, with the majority of the remaining time spent on common treatment components, e.g., behavioral interventions and nonspecific, active listening/supportive interventions. Contamination rates were low: 4.3% (ACT) and 5.7% (CT) of time was spent on aspects associated with the nonassigned treatment condition. Competence of therapist was rated as “good,” “very good,” or “excellent” in 91% of ACT sessions and 93% of CT sessions.

**MEASURES**

**Beck Depression Inventory – 2nd Edition (BDI-II; Beck et al., 1996)**

The BDI-II is an extensively used 21-item assessment tool, designed to assess the severity of current depressive symptoms. The BDI-II has been shown to have good reliability and strong content, concurrent, and discriminant validity in both clinical and nonclinical samples (Beck, Steer, & Garbin, 1988).

**Beck Anxiety Inventory (BAI; Beck, Epstein, et al., 1988)**

The BAI assesses the severity of anxiety related symptoms. The BAI is a 21-item self-report measure with high internal consistency (alpha=.92) and good reliability and validity in clinical populations (Beck et al.).

**Outcome Questionnaire (OQ; Lambert, Hansen, et al., 1996)**

The OQ is designed to measure patient functioning in subjective distress, interpersonal relationships, and social role performance. The 45 items yield a total score from 0 to 180, with higher scores indicating poorer functioning. The OQ has demonstrated excellent internal consistency and appropriate content and concurrent validity (Lambert, Burlingame, et al., 1996).

**Quality of Life Index (QOLI; Frisch et al., 2005)**

The QOLI is a measure of life satisfaction rooted in the view that overall life quality is the sum of satisfaction in a variety of life domains. Clients are asked to rate the importance of (0=not important

---

$^3$ Follow-up assessments varied from 14 to 20 months (mode=18 months). The nature of the sample was such that most participants had left the region, and locating them proved to be slow and arduous, producing variability in follow-up times. Due to this variability, time to follow-up was entered as a covariate in all analyses.
to 2 = extremely important and satisfaction with (-3 = very dissatisfied to 3 = very satisfied) a variety of life domains. The index takes the sum of the products of each domain. Test–retest coefficients and internal consistency coefficients are very good (Frisch et al.).

Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004)
The KIMS is a 39-item measure of four components of mindfulness: observing, describing, acting with awareness, and accepting without judgment (Baer et al., 2004). Items are rated on a 5-point Likert scale ranging from 1 (never or very rarely true) to 5 (almost always or always true). The measure was found to have high internal consistency, adequate to good test-retest reliability, and validation analyses providing support for the relationship between mindfulness and mental health (Baer et al.).

Acceptance and Action Questionnaire (AAQ; Hayes, Strosahl, et al., 2004)
The AAQ is a 9-item measure of the extent to which an individual demonstrates an accepting attitude towards negative feelings and experiences and the ability to take action even when feeling dysphoric or uncertain. Items are rated on a 1 (never true) to 7 (always true) scale, with higher scores indicating greater levels of experiential avoidance. The AAQ has demonstrated very good internal consistency, and has adequate criterion-related, predictive, and convergent validities (Hayes et al.).

Global Assessment of Functioning (GAF) and Clinical Global Impression (CGI)
At baseline and posttest, but not at follow-up, clinicians rated GAF (Spitzer, Gibbon, Williams, & Endicott, 1996), which ranges from 1 (persistent danger of hurting oneself or others) to 100 (superior functioning), and also the CGI (Guy, 1976), which ranges from 1 (normal) to 7 (extreme illness).

Results

Participant enrollment
Key variables, including primary diagnostic category, \( \chi^2(1) = .69, p = .88 \), and days to follow-up (\( t = -.20, p = .84 \)), were equivalent by assignment to condition. In addition, there were no significant differences between those who provided follow-up data (\( n = 91 \))

Table 1
Baseline Differences in Demographic Variables

<table>
<thead>
<tr>
<th></th>
<th>Competed follow-up (%)</th>
<th>No follow-up (%)</th>
<th>Comparisons using logistic regression (Wald's ( \chi^2 ) (p))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT</td>
<td>CT</td>
<td>ACT</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>82.6</td>
<td>80.0</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Married</td>
<td>47.8</td>
<td>55.6</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>White</td>
<td>76.1</td>
<td>68.9</td>
</tr>
<tr>
<td>Comorbidity</td>
<td>Present</td>
<td>30.4</td>
<td>40.0</td>
</tr>
</tbody>
</table>

Note. CT = Cognitive Therapy; ACT = Acceptance and Commitment Therapy.

Table 2
Baseline Differences by Follow-up Status

<table>
<thead>
<tr>
<th></th>
<th>Completed follow-up (( M (SD) ))</th>
<th>No follow-up (( M (SD) ))</th>
<th>Comparisons using ANOVA (( F ) (p))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT</td>
<td>CT</td>
<td>ACT</td>
</tr>
<tr>
<td>BDI</td>
<td>19.2 (10.4)</td>
<td>19.2 (10.2)</td>
<td>16.5 (7.7)</td>
</tr>
<tr>
<td>BAI</td>
<td>15.0 (10.4)</td>
<td>15.5 (11.2)</td>
<td>14.3 (13.6)</td>
</tr>
<tr>
<td>OQ</td>
<td>81.4 (16.7)</td>
<td>82.3 (16.8)</td>
<td>75.1 (15.3)</td>
</tr>
<tr>
<td>OQOLI</td>
<td>1.04 (2.49)</td>
<td>0.51 (2.08)</td>
<td>0.80 (1.71)</td>
</tr>
<tr>
<td>AAQ</td>
<td>53.0 (5.6)</td>
<td>53.5 (6.8)</td>
<td>53.0 (7.9)</td>
</tr>
<tr>
<td>KIMS-Observe</td>
<td>25.0 (8.5)</td>
<td>24.9 (9.7)</td>
<td>22.4 (7.6)</td>
</tr>
<tr>
<td>KIMS-Describe</td>
<td>18.6 (6.6)</td>
<td>20.1 (7.0)</td>
<td>19.4 (5.9)</td>
</tr>
<tr>
<td>KIMS-Act with Awareness</td>
<td>17.6 (6.3)</td>
<td>17.6 (6.6)</td>
<td>17.8 (4.1)</td>
</tr>
<tr>
<td>KIMS- Acceptance</td>
<td>16.8 (7.4)</td>
<td>18.0 (7.8)</td>
<td>18.5 (7.0)</td>
</tr>
<tr>
<td>CGI</td>
<td>3.4 (1.2)</td>
<td>3.6 (1.1)</td>
<td>3.1 (1.5)</td>
</tr>
<tr>
<td>GAF</td>
<td>63.7 (9.7)</td>
<td>61.8 (10.6)</td>
<td>66.8 (13.5)</td>
</tr>
</tbody>
</table>

Note. CT = Cognitive Therapy; ACT = Acceptance and Commitment Therapy; BDI = Beck Depression Inventory; BAI = Beck Anxiety Inventory; OQ = Outcome Questionnaire; AAQ = Acceptance and Action Questionnaire; KIMS = Kentucky Inventory of Mindfulness Skills; CGI = Clinical Global Impression; GAF = Global Assessment of Functioning.
and those who did not on key demographic or outcome variables, but those diagnosed with an anxiety disorder were more likely to have completed the follow-up assessment, $\chi^2(3)=8.11$, $p=.04$. A roughly equal percentage of ACT ($n=67\%$) and CT ($n=71\%$) participants completed follow-up, $\chi^2(1)=0.35$, $p=.56$. To probe yet further, a series of analyses were performed in order to determine if the follow-up status (i.e., whether or not a participant was available at follow-up), the treatment condition, or the interaction of follow-up status and condition was related to key demographic, comorbidity, and outcome variables. Logistic regression and ANOVAs indicated that neither follow-up status nor treatment condition nor their interaction was related to baseline characteristics of the sample (Tables 1 and 2).

**Analytic Strategy**

Because we have previously reported posttreatment outcome results on an overlapping sample (Forman et al., 2007), we focus here on follow-up analyses. We conducted long-term follow-up analyses on two samples: (a) an intention-to-treat (ITT) sample ($n=132$) with missing follow-up scores imputed using Expectation Maximization (EM); a relatively sophisticated method of estimating missing values based on iteratively estimating missing data and estimating relationships between observed relationships in the full dataset including, in this case, demographic variables and baseline, posttreatment and observed follow-up values; Dempster, Laird & Rubin, 1977; Schafer & Graham, 2002); and (b) a sample ($n=91$) of those who had presented for follow-up and who were treatment completers (i.e., all those who presented for follow-up were regarded as completers, which we had defined a priori as having received $\geq 5$ sessions).

**Long-term outcomes across treatments**

Significant and large main effects of time on depression, $F(1, 127)=25.86$, $p<.001$, $\eta^2_p=.16$, $f=.44$; anxiety, $F(1, 127)=16.21$, $p<.001$, $\eta^2_p=.11$, $f=.35$; general functioning, $F(1, 127)=19.55$, $p<.001$, $\eta^2_p=.13$, $f=.39$; and quality of life, $F(1, 124)=4.36$, $p=.02$, $\eta^2_p=.03$, $f=.18$ were obtained from pretreatment to follow-up (and from pre- to posttreatment) in the ITT analyses. The complete analysis revealed similarly sized effects of time on depression, $F(1, 70)=25.56$, $p<.001$, $\eta^2_p=.27$, $f=.61$; anxiety, $F(1, 70)=19.43$, $p<.001$, $\eta^2_p=.22$, $f=.53$; general functioning, $F(1, 70)=19.46$, $p<.001$, $\eta^2_p=.22$, $f=.53$; and quality of life, $F(1, 82)=2.41$, $p=.10$, $\eta^2_p=.03$, $f=.18$. As an ad hoc analysis, diagnostic category (anxiety disorder, depressive disorder, both anxiety and depressive disorders) was entered as a moderator. Results revealed that those diagnosed with an anxiety disorder alone start at higher levels and improve faster on the BAI, $F(1, 68)=3.01$, $p=.01$, $\eta^2_p=.13$, and those diagnosed with comorbid depression and anxiety disorders improve more slowly on the BDI, $F(1, 68)=2.04$, $p=.06$, $\eta^2_p=.09$.

**Long-term outcomes by treatment group**

As in our previous analysis, no differences emerged by group at posttreatment. However, at follow-up, a Group×Time interaction effect emerged. In the full ITT sample, the CT group exhibited greater maintenance of treatment effects for depression, $F(1, 128)=5.36$, $p=.02$, $\eta^2_p=.04$, $f=.21$; general functioning, $F(1, 128)=5.08$, $p=.03$, $\eta^2_p=.04$, $f=.22$; quality of life (weak trend; $F(1, 124)=2.78$, $p=.08$, $\eta^2_p=.02$, $f=.14$); though not for anxiety, $F(1, 128)=.28$, $p=.63$, $\eta^2_p<.01$, $f=.07$. Results for the completers sample followed a similar pattern: general functioning, $F(1, 68)=2.40$, $p=.15$, $\eta^2_p=.03$, $f=.19$; depression, $F(1, 68)=2.73$, $p=.07$, $\eta^2_p=.04$, $f=.21$; quality of life, $F(1, 82)=1.70$, $p=.19$, $\eta^2_p=.02$, $f=.13$; and anxiety, $F(1, 68)=.68$, $p=.51$, $\eta^2_p=.01$, $f=.10$. Thus, analyses revealed a small- to moderate-sized interaction effect on depression, functioning, and quality of life and a near-zero effect on

---

4 We also considered utilizing Hierarchical Linear Modeling (HLM) because of its ability to model data at the individual and group (i.e., nested) levels and its ability to handle missing data. However, the former was not a focus of the current study and the latter could be even more conservatively handled through the Expectation-Maximization (EM) algorithm (Hedeker & Gibbons, 1997). Of note, we obtained similar results with missing data included and excluded, suggesting that missing data are not having a substantial effect on outcome. Additionally, some experts in HLM caution against its use in naturalistic studies. HLM can lead to a misinterpretation of the rate of progress being made at different points in treatment because it treats participants who end treatment “prematurely” as though they are going forward with therapy after already having made gains when in fact that is not the case (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). Finally, in an attempt to be maximally conservative, we re-ran ITT analyses making different assumptions about missing data, and obtained equivalent results (see footnote 5).

5 Despite the reassuring analyses indicating that follow-up status did not impact outcomes, it is possible that the trajectories of those who presented for follow-up assessments differed from those who did not. To further increase confidence in our findings in light of missing data, we repeated the long-term follow-up analyses using two additional samples derived from alternate assumptions: an ITT sample with baseline values carried forward (which makes the conservative assumption that those who did not present for follow-up regressed to baseline values) and an ITT sample with the last assessment observation carried forward (which makes the assumption that no change occurred between posttreatment and follow-up). Results from these two samples were equivalent to those obtained from the EM and completer samples.
anxiety. Observed values at baseline, posttreatment, and follow-up are listed in Table 3. No support was obtained for the hypothesis that either diagnostic status or mindfulness or acceptance measures would further qualify the Time×Treatment effect (\(p_s \geq .16, \eta_{p}^2 \leq .02\)), with one exception; KIMS-Describe moderated the Time×Treatment effect on the QOLI, \(F(2, 119) = 13.89, p = .003, \eta_{p}^2 = .05\). Completer-only analyses yielded equivalent findings.

**Clinical significance**

Clinical significance was analyzed using the recommendations of Jacobson and Truax (1991). We determined the proportion of participants who had “recovered,” i.e., were in the “clinical” range at baseline and “normative” range at follow-up (with cutoff determined by means and standard deviations from normative and clinical samples for each measure; see Jacobson & Truax, 1991) by an amount that exceeded the reliable change index for the measure (determined by reliability and standard deviation of each measure; see Jacobson & Truax). For the BDI, 81.8% of CT patients versus 60.7% of ACT patients reliably recovered. For the BAI, these numbers were 72.7% (CT) and 56.0% (ACT), for OQ they were 46.4% (CT) and 22.6% (ACT), and for QOLI they were 37.8% CT and 22.9% ACT (Figure 2).

**Discussion**

ACT has demonstrated effectiveness across a wide range of problem behaviors (Hayes, et al., 2006; Powers, et al., 2009); however, ACT intervention studies often lack long-term follow-up analyses and active comparisons. The current study sought to address the question of ACT’s long-term effectiveness compared to CT.

The results reveal that the two treatments are equally effective in the short term: both were successful in maintaining improvements in depression, anxiety, and general functioning. Yet, statistical comparisons of long-term outcomes suggest that CT has a slight advantage over ACT in the long-term maintenance of gains, at least with respect to depressive symptoms and general functioning. By follow-up, quality of life appeared to increase at a steeper rate in those receiving CT (versus ACT), though this result is tempered by the fact that ending points were equivalent. In contrast to some other reports, baseline levels of acceptance and mindfulness did not qualify these effects.

Mean differences in symptom level change were modest (e.g., 3 BDI points); however, clinical significance analyses revealed that roughly one-third more CT than ACT patients remained in the recovered range at long-term follow-up, with similar differences for quality of life. Furthermore, in the case of the OQ—which is a good all-around indicator of interpersonal and occupational functioning as well as symptom distress—more than twice as many CT as ACT patients were in the normative range at follow-up.

Several possible explanations can be offered for these findings. First, it is possible that CT is a more intuitive and simpler intervention, thus facilitating the ability of patients to apply the learned strategies and skills independently once treatment has terminated. As we have argued elsewhere (Forman & Herbert, 2009), CT is highly compatible with folk psychology.
whereas ACT tends to be more counterintuitive. In a sense, ACT requires the individual to apply concepts that often run counter to prevailing social norms and customs, and as the time away from the therapist’s support and guidance increases, the more difficult it may be for some individuals to implement ACT principles in the face of these prevailing contingencies. Another possibility is that study therapists were better able to administer CT (although the comparable competence and adherence data do not support this hypothesis), perhaps because they were relatively inexperienced (though well-trained). Finally, most of the patients in this study were relatively high functioning and had only modest symptom levels. Some authors have speculated that acceptance-based psychotherapies are especially well suited to more severe, treatment-resistant patients (Kenny & Williams, 2007; Twohig, 2009). No matter the explanation, it challenges the notion that ACT contains more active ingredients than does Beckian CT or that its principles have strong empirical backing.

The present study is the first to our knowledge to directly compare the long-term efficacy of ACT to another gold-standard CBT. Strengths of the study include true random assignment, the long-term follow-up assessment, and therapists whose adherence, competence, and absence of treatment allegiance have been ensured. The study also emphasized external validity using minimal exclusion criteria, and a naturalistic course of treatment based on general guidelines instead of highly prescriptive manuals. In addition, results were verified across multiple imputed datasets with varying assumptions. These strengths are tempered by complementary limitations including diagnostic variability, intervention variability due to minimalist treatment manualization, relatively novice therapists, the lack of a no-treatment control and the lack of formal mediational analyses. The fact that most measures were self-report represents an additional weakness in that data can be limited by self-knowledge and/or response bias (and can reflect experimenter demand). The absence of data on subsequent treatment during the follow-up period should also be noted. In addition, retention was modest (though comparable to similar trials), likely because of the high mobility of the sample. The convergence of findings across various IIT and completer samples partially attenuates this concern.

The results of this study raise interesting questions about the long-term maintenance of gains in a newer, acceptance-based model of CBT relative to standard CT. These findings merit further study, and replication is needed before definitive conclusions can be drawn. Such research would be augmented by an even longer follow-up period and blind clinical assessments, as well as a more clinically impaired sample treated by expert therapists. Studies of moderating variables are especially needed to assess the potential advantages of matching patient characteristics with the intervention most likely to yield long-term benefits.

References


**RECEIVED:** July 9, 2011  
**ACCEPTED:** April 16, 2012  
Available online 24 April 2012