Cognitive Behaviour Therapy
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/sbeh20

Effectiveness of Cognitive Behavior Therapy for Severe Mood Disorders in an Acute Psychiatric Naturalistic Setting: A Benchmarking Study

Thröstur Björgvinsson[a], Sarah J. Kertz[a], Joseph S. Bigda-Peyton[a], David H. Rosmarin[a], Idan M. Aderka[a] & Edmund C. Neuhaus[a]

[a] Department of Psychiatry, McLean Hospital/Harvard Medical School, Belmont, MA, USA

Published online: 31 Mar 2014.

To cite this article: Thröstur Björgvinsson, Sarah J. Kertz, Joseph S. Bigda-Peyton, David H. Rosmarin, Idan M. Aderka & Edmund C. Neuhaus (2014): Effectiveness of Cognitive Behavior Therapy for Severe Mood Disorders in an Acute Psychiatric Naturalistic Setting: A Benchmarking Study, Cognitive Behaviour Therapy, DOI: 10.1080/16506073.2014.901988

To link to this article: http://dx.doi.org/10.1080/16506073.2014.901988

Cognitive Behaviour Therapy, 2014
http://dx.doi.org/10.1080/16506073.2014.901988

Routledge
Taylor & Francis Group

Effectiveness of Cognitive Behavior Therapy for Severe Mood Disorders in an Acute Psychiatric Naturalistic Setting: A Benchmarking Study

Thröstur Björgvinsson, Sarah J. Kertz, Joseph S. Bigda-Peyton, David H. Rosmarin, Idan M. Aderka and Edmund C. Neuhaus

Department of Psychiatry, McLean Hospital/Harvard Medical School, Belmont, MA, USA

Abstract. The current study examined the effectiveness of brief cognitive behavior therapy (CBT) for severe mood disorders in an acute naturalistic setting. The sample included 951 individuals with either major depressive disorder ($n = 857$) or bipolar disorder with depressed mood ($n = 94$). Participants completed a battery of self-report measures assessing depression, overall well-being, and a range of secondary outcomes both before and after treatment. We found significant reductions in depressive symptoms, worry, self-harm, emotional lability, and substance abuse, as well as significant improvements in well-being and interpersonal relationships, post-treatment. Comparable to outpatient studies, 30% of the sample evidenced recovery from depression. Comparison of findings to benchmark studies indicated that, although the current sample started treatment with severe depressive symptoms and were in treatment for average of only 10 days, the overall magnitude of symptom improvement was similar to that of randomized controlled trials. Limitations of the study include a lack of control group, a limitation of most naturalistic studies. These findings indicate that interventions developed in controlled research settings on the efficacy of CBT can be transported to naturalistic, "real world" settings, and that brief CBT delivered in a partial hospital program is effective for many patients with severe depressive symptoms. *Key words: cognitive behavioral therapy; depression; benchmarking study; treatment outcome; naturalistic setting.*

Received 30 October 2013; Accepted 4 March 2014

Correspondence address: Thröstur Björgvinsson, Behavioral Health Partial Program, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA. Tel: +1617 855 4180. Email: tbjorgvinsson@mclean.harvard.edu

Mood disorders are among the most common psychiatric disorders and are associated with significant functional impairment (Kessler, Chiu, Demler, & Walters, 2005) and tremendous healthcare costs, totaling $83.1 billion in the USA in 2000 (Donohue & Pincus, 2007). The efficacy of cognitive behavior therapy (CBT) for symptoms of depression is well-established; however, there are far fewer studies investigating the effectiveness of CBT in naturalistic, "real world" conditions. Further, while some literature points to the efficacy of CBT for severe mood disorders (e.g., DeRubeis et al., 2005), nearly all effectiveness studies of CBT have used a traditional, once-weekly, individual outpatient treatment model, despite the prevalence of other treatment delivery models. Thus, the current study was designed to examine the effectiveness of CBT delivered in an increasingly common but understudied treatment setting, the partial hospital. On the continuum of clinical care, partial hospital programs are located between 24-h inpatient care on one end, and more traditional 1-h per week outpatient treatment on the other. Given that partial hospital programs are increasingly popular treatment options, especially for patients experiencing acute worsening of symptoms and those with severe and chronic symptoms (Kiser, Lefkovitz, Kennedy, & Knight, 2010), additional study of such programs is warranted.

While randomized controlled trials (RCTs) of CBT for depression are conducted under conditions with high internal validity and are ideal for examining efficacy, effectiveness

studies conducted under "real world" conditions emphasizing external validity are necessary to establish generalizability. In contrast to RCTs, there is a paucity of research examining CBT for mood disorders in naturalistic settings. The limited literature available is promising and suggests that CBT can be successfully transported from the laboratory to clinical practice. For example, one study found that 45 individuals with depression receiving CBT in a private clinical practice evidenced significant pre- to post-treatment improvements that were comparable to those found in research settings (Persons, Bostrom, & Bertagbolli, 1999). Another study indicated that diagnostically mixed clients receiving CBT under routine outpatient treatment evidenced significant improvements in depression from pre- to post-treatment and an overall recovery rate of 30% (Westbrook & Kirk, 2005). However, research on the implementation of CBT within partial hospitals, involving both group and individual treatments with acute psychiatric patients, remains in its infancy.

Therefore, the current study sought to extend findings on CBT outcomes by focusing on individuals treated in an understudied but increasingly common partial hospital program setting. This intermediate form of treatment is both cost efficient, as it is typically less expensive than inpatient or residential treatment, and appealing, as patients with severe symptoms are able to engage in treatment with minimal disruption to daily life (e.g., while living at home; Kiser et al., 2010). CBT delivered in a partial hospital setting has shown promising preliminary results in a pilot study (Neuhaus, Christopher, Jacob, Guillaumot, & Burns, 2007), but additional research is needed.

The goal of the current study was to examine the effectiveness of CBT for individuals with current depression treated in a brief, 2-week partial hospital program. To address the limitations of previous work, the study included a large sample of individuals who were formally diagnosed and addressed a broad battery of measures including depressive symptoms, worry, psychological well-being, and general psychopathology, to assess treatment effects. Literature-wide benchmarks were also created to facilitate comparison of the magnitude of the outcomes observed in this sample with those published in clinical trials, which allowed for the interpretations of the findings in the context of the broader CBT literature.

Methods

Participants and procedure

Participants were patients seeking treatment at the Behavioral Health Partial Hospital Program (BHP), a partial hospital program in New England. Approval for the study was granted by the hospital's institutional review board, and all participants provided written informed consent to the study. At pre-treatment, patients completed a semi-structured diagnostic interview, a demographics survey, and a battery of self-report measures. The battery of self-report measures was repeated at post-treatment. Participants were informed that they were free to stop the study at any point without any consequence.

Study participants were patients with a current diagnosis of major depressive disorder (MDD) or bipolar disorder with depressed mood (Bipolar Disorder-Depressed), as assessed by the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) or psychiatrist clinical interview. A total of 1727 patients presented for treatment during the study period (July 2010–April 2013). Diagnostic data from the MINI were missing for 288 patients, and thus, diagnoses from the program psychiatrist were used. A total of 30 patients had neither MINI nor psychiatrist diagnoses available and were therefore excluded from further analyses, resulting in $n = 1697$ patients. Of these patients, 992 met criteria for MDD or Bipolar Disorder-Depressed. A small number ($n = 41$) failed to complete the pre-treatment assessment and were excluded. Thus, the final sample included $n = 951$ individuals with MDD ($n = 857$) or Bipolar Disorder-Depressed ($n = 94$). Slightly more than half (57%) were referred from the community (e.g., outpatient referrals) for higher level of care and 43% were referred from inpatient units as a step down to lower level of care.

Bonferroni corrected t-tests indicated that treatment completers and non-completers did not differ at pre-treatment on primary outcome variables of CESD-10 depression, $t(949) = .62$, $p > .05$, Cohen's $d = .88$, or SOS

Table 1. *Demographic and diagnostic characteristics*

	Total ($n = 951$)
Age (*M*, SD)	36.27 (14.44)
Gender (Female)	60% (570)
Race	
Non-hispanic white	87% (825)
Asian	4% (40)
Latino/a	3% (26)
Other	4% (41)
Choose not to respond	2% (19)
Currently unemployed	53% (501)
Currently a student	28% (263)
Currently married	27% (253)
Generalized anxiety disorder – current[a]	33% (311)
Panic disorder – current	12% (113)
Social anxiety disorder – current	20% (188)
Obsessive-compulsive disorder – current	8% (77)
Post-traumatic stress disorder – current	10% (94)
Alcohol abuse – current	6% (55)
Alcohol dependence – current	9% (85)

[a] *Cells below may not add to the total number of participants as comorbid diagnoses are represented.*

well-being, $t(917) = -1.77$, $p > .05$, Cohen's $d = .48$. Thus, their symptom severity at pretreatment did not appear to influence the likelihood of successful treatment completion. *T*-tests also indicated that MDD and Bipolar-Depressed patients did not differ in pretreatment CESD-10 depression, $t(949) = -.81$, $p > .05$, Cohen's $d = .85$, or SOS well-being, $t(917) = .88$, $p > .05$, Cohen's $d = .61$; thus, these two samples were combined for subsequent analyses. Demographic and diagnostic data for the final sample are presented in Table 1. All patients were taking medications prior to their treatment at BHP and some received medication adjustments during the course of treatment. Unfortunately, these data are not available for inclusion in analyses.

Treatment setting

The BHP offers CBT to individuals with acute symptoms across diverse diagnostic categories (principally mood, anxiety, personality, and psychotic disorders). Individual treatment plans were constructed by clinical team managers who conducted initial intake assessments and managed the overall treatment delivery. Treatment consisted primarily of group CBT provided by psychiatrists, psychologists, social workers, occupational therapists, postdoctoral and graduate level psychology trainees, and mental health counselors. Patients attended five 50-min groups each day, 5 days per week (Monday–Friday). Of these, one group per day focused on behavioral activation, based on a protocol adapted from Martell and colleagues (2010). A second group focused on identifying and challenging negative automatic thoughts guided by a protocol adapted from Beck and colleagues (1979). The remaining group content included modules on psychoeduction, self-monitoring, mindfulness, interpersonal skills, and skills acquisition, adapted from empirically supported CBT manuals (e.g., Beck, Emergy, & Greenberg, 1985; Linehan, 1993). To maintain treatment fidelity, groups utilized treatment protocols designed for the program, derived from established treatment manuals. Group therapy was facilitated by handouts and worksheets provided during the group. Groups were periodically observed by postdoctoral fellows and staff psychologists to ensure adherence to the treatment manuals. These observers had extensive experience in the treatment provided (between 2 and 15 years of experience). In addition to group therapy, patients also received two to three weekly individual CBT sessions from graduate-level psychologists to review and enhance material learned in groups. The average duration of treatment in the sample for this study was 10.36 ($SD = 4.61$) days, including non-treatment weekend days.

Measures

MINI (Sheehan et al., 1998). The MINI is a structured interview assessing for DSM-IV Axis I symptoms (e.g., mood, anxiety, substance abuse, and psychosis). The MINI was administered by doctoral practicum students and interns in clinical psychology who received weekly supervision by a postdoctoral psychology fellow. Training included reviewing administration manuals and completing mock interviews. All clinicians were required to pass a final training interview with their supervisor. The MINI has strong reliability and validity in relation to the

Structured Clinical Interview for DSM-IV (SCID-IV), with inter-rater reliabilities ranging from κs of .89–1.0 (Sheehan et al., 1998). Previous data from a subset of this sample indicated that inter-rater reliability between the MINI and the program psychiatrists was $\kappa = .69$ for MDD and $\kappa = .75$ for Bipolar Disorder-Depressed (Kertz, Bigda-Peyton, Rosmarin, & Björgvinsson, 2012).

Center for the Epidemiological Studies of Depression-10 (CESD-10; Andersen, Malmgren, Carter, & Patrick, 1994). The CESD-10 is a widely used, brief instrument for measuring symptoms of depression. Items are rated from 0 = rarely or none of the time (less than 1 day) to 3 = most or all of the time (5–7 days). The CESD-10 has been shown to be reliable and valid and had high internal consistency in this study ($\alpha = .85$).

Penn State Worry Questionnaire-Abbreviated (PSWQ-A; Hopko et al., 2003b). The PSWQ-A is a validated, single factor, 8-item measure designed to assess worry severity. Derived from the original 16-item instrument (Meyer, Miller, Metzger, & Borkovec, 1990), items are rated on a 5-point Likert type rating scale ranging from 1 (not at all typical of me) to 5 (very typical of me). Total scores range from 8 to 40, with higher scores indicating higher levels of worry. Reliability in the present study was very high ($\alpha = .94$).

Schwartz Outcome Scale (SOS; Blais et al., 1999). The SOS is a well-validated and reliable, single factor, 10-item measure designed to examine a broad domain of psychological health in a variety of settings (Young, Waehler, Laux, McDaniel, & Hilsenroth, 2003). Participants rate items on a 7-point Likert scale from 0 (never) to 6 (all or nearly all of the time). Total scores, with higher scores indicating better psychological health, range from 0 to 60. Internal consistency of the SOS was high in the present study ($\alpha = .90$).

Behavior and Symptom Identification Scale (BASIS-24; Eisen, Normand, Belanger, Spiro, & Esch, 2004). The BASIS-24 is a 24-item measure that demonstrated good psychometric properties across inpatient, outpatient, residential, and partial hospital settings as a broad assessment of psychopathology and associated distress. The BASIS-24 consists of six subscales: (1) Depression/Functioning, (2) Interpersonal Problems, (3) Self-Harm, (4) Emotional Lability, (5) Psychosis, and (6) Substance Abuse/Dependence. Respondents rate items on a 5-point Likert type rating scale from 0 (none of the time) to 4 (all of the time) and higher scores indicating worse functioning. Subscales range from 0–8 (self-harm) to 0–24 (depression/functioning). Subscale reliability in the current study was high ($\alpha >$ .80) for Depressive/Functioning ($\alpha = .85$) and Self-Harm ($\alpha = .83$), adequate for Substance Abuse/Dependence ($\alpha = .77$), Interpersonal Problems ($\alpha = .73$), and Emotional Lability ($\alpha = .71$), and poor for Psychosis ($\alpha = .61$). Thus, the Psychosis scale was not used in subsequent analyses.

Data analytic plan

Approximately 26% of the data were missing due to treatment dropout. Multiple imputation (Rubin, 1987) was used to account for missing data. Multiple imputation is a procedure that replaces missing values with a set of plausible values, which represents uncertainty about the correct value to impute (see Rubin, 1987 for more details). The Markov Chain Monte Carlo method in PRELIS 9.1 was used to generate five imputed data-sets, based on indications that only 3–10 data-sets are necessary with the current proportion of missing data (Rubin, 1987). A complete set of analyses was then conducted on each of the five imputed data-sets using the Statistical Package for the Social Sciences (SPSS) version 19.0. Estimates from the five databases were pooled using Rubin's (1987) guidelines.

Pre-post differences. Changes in symptom scores during the course of treatment were examined using repeated-measures analyses of variance (ANOVA), with time (pre-treatment vs. post-treatment) as the independent variable. The CESD-10 was the primary depression outcome measure, but changes in worry, overall well-being, and BASIS subscale scores were also examined to assess for a range of symptomatic and functional improvement.

Clinical significance. Clinical significance was assessed by examining changes in the primary outcome variable, CESD-10 depression, and SOS overall well-being. Clinical significance was estimated by calculating reliable change index scores and examining whether or not a patient shifted across a normative cut-off point during treatment. We used the method outlined by Jacobson and Truax (1991) for

calculating cut-off point b to establish the threshold for a functional range (2 SDs above the mean of the functional population), so that scores below the cut-off point indicate clinically significant change. We used previously published community norms ($M = 4.76$ $SD = 4.05$; Ybema & van den Bos, 2010) to establish a normative cut-off point of 12.86 for the functional range on the CESD-10. For the SOS, previous research has suggested that a change of 8.5 points indicates reliable change based on Jacobson and Truax criteria (1991) and a score of 41 functions well as a normative cut-off score for patients and non-patients (Blais, Kehl-Fie, & Blais, 2008). The two indices (reliable change index and normative cut-off point) were used to create a classification table of patients as recovered, improved but not recovered, unchanged, or reliably deteriorated in terms of depression (Jacobson & Truax, 1991).

Benchmarking procedure. The search for comparison studies was modeled from previously established procedures (Weersing & Weisz, 2002). Twenty-two studies were located and data from CBT conditions were coded (see Table 2). Depression scores were z-transformed and then aggregated across studies to form a mean CBT benchmark score (Hedges & Olkin, 1985). We first searched for clinical trials of CBT for mood disorders in adults using PsycINFO, PsychLit, and Medline as well as reference lists in identified studies, review articles, book chapters, and meta-analyses (4). We identified 22 clinical trials that (1) treated depressed adults in mental health clinics; (2) included random assignment to one or more psychosocial treatment conditions; (3) were published in an English, peer-reviewed journal; and (4) reported pretreatment means of primary outcomes. Following the procedure outlined by Weersing and Weisz (2002), we used studies conducted within the last 10 years, and preference was given to more recent studies. The primary dimensional depression measure for each study was identified (see Table 2) and norms for nonclinical, community samples of adults were obtained. We then used the mean and SD of the normal sample to compute normative z scores for the CBT treatment group (znt) at pre- and post-treatment. These computations took the general form $znt = (xt - \eta)/(\bar{o})$, where xt was the CBT group mean, η was the normal population mean, and $\bar{o}$ was the normal population standard deviation for the depression measure. By using normative data to construct our z scores, we were able to place all studies on the same metric for comparison, giving the metric clinically significant meaning. Finally, we aggregated scores at the pre-treatment and post-treatment assessments with an unweighted mean, allowing direct computation of standard errors and confidence intervals for CBT benchmark means at pre- and post-treatment. In addition, normative z-scores calculated for CBT and control groups were based on studies with more than 10 participants per cell. This is a common cut point for use of sample-weighted aggregates in meta-analysis (Hedges & Olkin, 1985), suggesting that use of the unweighted mean would not unduly influence results (see Weersing & Weisz, 2002, for additional discussion).

Results

Comparison of pre- and post-treatment scores

An ANOVA was first estimated for CESD-10 scores, with time (pre-treatment, post-treatment) as a within-subjects factor. The effect for time was significant, $F(1, 950) = 1348.24$, $p < .001$, and the effect size was large, Cohen's $d = 1.13$, and participants had lower scores at post-treatment compared to pre-treatment. Pre- and post-treatment means, F statistics, and effect sizes are presented in Table 3. Results for SOS well-being were also significant, $F(1, 950) = 1058.82$, $p < .001$, with a large effect size, Cohen's $d = -.91$. Pre- and post-treatment scores were also examined across a range of outcomes using a series of Bonferroni-corrected ANOVAs. Results were significant for all secondary outcome measures, all $p < .001$. Changes in worry, interpersonal problems, self-harm, and emotional lability were associated with medium effect sizes while the effect size for substance abuse was small.

Clinical significance

Clinically significant change. Using the method outlined above, overall 32% ($n = 304$) of the

Table 2. *Comparison of BHP sample to randomized control trial benchmark studies*

			Demographic characteristics		Clinical/Treatment characteristics		
Study	Percent attrition	Primary measure	Mean age	Percent male	Depression diagnoses	Other diagnoses	Sessions
Bagby et al. (2008)	29.9	HDRS-17	41.9	37.5	100% MDD	Axis II personality traits	18
Bodenmann et al. (2008)	3.3	BDI	45.3	41.7	100% MDD Non-Depressed Partner	–	20
David et al. (2008)	11.1	BDI	37.0	22.2	100% MDD	–	14
DeRubeis et al. (2005)	15.5	HDRS-17	40.0	31.0	90% Recurrent MDD	72% Axis 1 Comorbidity 48% Axis 2 Comorbidity	8
Dimidjian et al. (2006), Dobson et al. (2008)	17.8	BDI	39.9	34.0	57.7% Recurrent MDD 34.4% Chronic MDD	28.2% Any Axis I comorbidity	24
Dozois et al. (2009)	16.6	BDI-II	46.5	26.0	100% MDD	44% Anxiety Disorderbr15% PTSD	15
Faramarzi et al. (2008)	28.0	BDI	28.8	0	100% MDD	36% Infertility diagnosis	10
Freedland et al. (2009)	11.5	BDI	60.6	50.3	66% Major Depression 34% Minor Depression	Recent Coronary Surgery	8
Hopko et al. (2003a)	0	BDI	30.5	64.0	100% MDD	44% Substance Abuse Disorder 40% Anxiety Disorder	7
Klein et al. (2004)	27.2	HRSD-24	44.1	33.9	59.2% Chronic MDD 40.8% MDD w/ DD	25.7% Hx of Sub Ab Disorder 26.8% Hx of Anxiety Disorder	18
Kocsis et al. (2009)	21.7	HDRS-17	44.6	47.0	100% MDD	–	13
Kuyken et al. (2008)	6.5	BDI-II	49.2	53.0	67.5% MDD Remission 32.5% MDD Partial Remission	34% Previous SI Attempt	12
Laidlaw et al. (2008)	9.0	BDI	74.0	27.5	100% MDD	–	8
Manber et al. (2008)	6.6	HDRS-17	48.6	39.0	100% MDD	100% Insomnia-based Sleep Disorder	7
Miranda et al. (2003, 2006)	59.5	HDRS-17	29.3	0	100% MDD	PTSD (*high rates*)	6
Perlis et al. (2002)	36.6	HRSD-24	39.9	45.5	100% MDD	Inclusion = 3 or more MD episodes	19
Rohan et al. (2007)	10.0	BDI-II	45.0	10.0	100% MDD	100% Seasonal Affective	16
Sava et al. (2009)	11.0	BDI	37.0	32.4	100% MDD	–	20
Serfaty et al. (2009)	18.3	BDI-II	74.1	20.6	88% Major Depression 12% Minor Depression	–	12
Smit et al. (2006)	25.0	BDI	42.6	36.8	32% Severe MDD 30% Moderate MDD	15% Social Anxiety Disorder 11% Panic Disorder	11

Strauman et al. (2006)	13.3	BDI	38.6	26.6	38.1% MDD	24.6% Anxiety Disorder	12
					20.6% Dysthymic Disorder	66.7% Axis II Disorder	
Wright et al. (2005)	11.1	BDI	40.2	24.5	100% MDD	–	9
BHP	25.8	CES-D	36.27	40	67.4% MDD	35% Generalized Anxiety D	Varied 8.4 Days
					17.4% Mood Disorder NOS	23% Social Anxiety Disorder	
					15.3% Bipolar Disorder-Depressed	13.4% Panic Disorder	

Note. *Dashes indicate that data were not reported. HRSD-24, Hamilton Rating Scale for Depression – 24-item; SCL-D, Hopkins Symptom Checklist; QIDS-SR, Quick Inventory of Depressive Symptomatology; HDRS-17, Hamilton Rating Scale for Depression – 17-item; BDI, Beck Depression Inventory; BDI-II, Beck Depression Inventory – Revised; BHP, Behavioral Health Partial Hospital Program; MDD, Major Depressive Disorder; Bipolar Disorder-Depressed, Bipolar Disorder – Currently Depressed; Hx, History; Sub Ab, Abuse Disorder.*

total sample achieved reliable change in symptoms of depression, 68% ($n = 644$) did not, and 0.3% ($n = 3$) reliably deteriorated. Next, we examined the proportion of patients who crossed the threshold into the established functional range (within 2 SDs of a non-clinical mean) at post-treatment. Of those who were above the cut-off score for depression at pre-treatment ($n = 740$), 52% ($n = 383$) improved and crossed into the functional range at post-treatment. Of those who were below the cut-off score at pre-treatment, 90% ($n = 190$) remained below cut-off score for depression at post-treatment and therefore did not worsen. Results were then classified using both reliable change and return to functional range (Jacobson, Roberts, Berns, & McGlinchey, 1999). Of those above the cut-off score at pre-treatment ($n = 740$), 32% ($n = 237$) were recovered (crossed the threshold and the change was reliable), 8% ($n = 58$) were improved but not recovered (experienced reliable change but were still above the cut-off score), 20% ($n = 146$) were unchanged (below the threshold but change was not reliable), and 40% ($n = 299$) were above the cut-off score and change was not reliable.

In terms of well-being, 95% ($n = 905$) of patients were below the cut-off score at pre-treatment, indicating poor well-being. Of these, 15% ($n = 139$) experienced both reliable change and crossed the threshold for normative well-being and could be considered recovered. An additional 2% ($n = 18$) improved and were above the cut-off score at pre-treatment but change was not reliable; 40% ($n = 366$) experienced reliable improvement but were still below the normative cut-off score at post-treatment. Finally, 1% ($n = 11$) saw reliable deterioration in well-being and remained below the cut-off score at post-treatment.

Benchmarking strategy. Each patient's pre-treatment and post-treatment depression scores were transformed to normative *z*-scores. Results suggested that the current sample was more symptomatic at pre-treatment compared to RCT samples (*z*-score of 3.37 compared to 2.68). At post-treatment, the mean *z*-score of the current sample was 1.65, which was higher than that of .86 for the benchmark CBT outpatient conditions. This indicates that at the conclusion of partial

Table 3. *Overall means, SDs, F statistics, and within-group effect sizes across outcome measures*

	Pre treatment *M* (SD)	Post-treatment *M* (SD)	*F* statistic	Overall Cohen's *d*	95% CI on Cohen's *d*
CESD-10	18.39 (6.38)	11.47 (5.83)	1348.24	1.13	1.14, 1.15
SOS	21.32 (11.19)	31.69 (11.68)	1058.82	−0.91	−0.93, −0.88
PSWQ-A	29.70 (8.40)	25.86 (8.01)	403.45	0.47	0.45, .49
BASIS-24-DF	2.50 (0.82)	1.61 (0.69)	1305.00	1.18	1.15, 1.20
BASIS-24-IP	1.45 (0.77)	1.11 (0.66)	216.72	0.47	0.46, 0.47
BASIS-24-SH	0.87 (0.96)	0.41 (0.64)	353.53	0.56	0.54, 0.59
BASIS-24-EL	1.72 (0.81)	1.32 (0.8)	231.66	0.43	0.42, 0.45
BASIS-24-SA	0.45 (0.69)	0.33 (0.52)	60.78	0.20	0.17, 0.23

Note. *CESD-10, Center for the Epidemiological Studies of Depression-Short Form; PSWQ-A, Penn State Worry Questionnaire-Short Form; SOS, Schwartz Outcome Scale; BASIS-24, Behavior and Symptom Identification Scale, DF, Depressive Functioning subscale, IP, Interpersonal Problem subscale, SH, Self-Harm subscale, EL, Emotional Lability, SA, Abuse/Dependence; CESD-10 scores doubled to facilitate comparison with other benchmark studies.*

hospital treatment, patients are more symptomatic than when concluding outpatient treatment, which is to be expected. However, the change in *z*-scores was very similar, with an average decrease of 1.72 in the current sample and 1.83 for the benchmark CBT outpatient conditions. Thus, the current sample reported higher depression severity at both pre- and post-treatment compared to those in the benchmark CBT outpatient conditions, but evidenced comparable level of absolute change.

Discussion

Overall, findings indicated that individuals with severe depression who completed partial hospital treatment experienced significant reductions from pre- to post-treatment on a range of outcomes, including depression, worry, self-harm, emotional lability, and substance abuse, as well as significant improvements in well-being and interpersonal problems. Effect sizes for depression symptom reduction were large and comparable to those reported in other outpatient naturalistic settings (Westbrook & Kirk, 2005) and the benchmark RCTs. Results also indicated that 32% of patients achieved reliable change, and approximately one-third of those who experienced clinical levels of depression severity on the CESD-10 at pre-treatment were considered recovered at post-treatment. This recovery rate is comparable to a naturalistic study of individual outpatient treatment, provided once a week, in which 30% of the sample recovered (Westbrook & Kirk, 2005). In addition, half of the patients in the current sample who started treatment in the clinical range of the CESD-10 were within the normal range at post-treatment. Thus, CBT provided in a psychiatrically acute partial hospital program in this sample resulted in effect sizes for depression that are comparable to those seen in RCTs but slightly lower rates of reliable change and recovery.

The finding that CBT was effective overall in the current sample is encouraging in light of the severity of the patients' symptoms. At pre-treatment, patients in this sample were more symptomatic than most patient samples represented in benchmark comparison RCTs. For example, although the average level of depression in our sample was higher than the average pre-treatment scores of individuals in benchmark comparison RCTs, the magnitude of change was similar. Results provide additional evidence that CBT is effective for individuals with very severe symptoms who may be excluded from traditional RCTs; further, this is the first known study to suggest that a brief, partial hospitalization program is effective in reducing symptoms of depression, improving overall well-being, and a range of other outcomes, including worry, interpersonal problems, self-harm, emotional lability, and substance abuse.

Despite these promising results, not all patients in the sample improved. It should be noted that 20% of patients were unchanged over the course of treatment and 40% reported CESD-10 scores above the clinical

cut-off score at post-treatment. Further, a small number of patients (<1%) showed reliable deterioration, a rate consistent with other studies (McEvoy & Nathan, 2007; Persons et al., 1999; Westbrook & Kirk, 2005). It may be that the clinical significance criterion is not the most accurate estimate of successful treatment at a partial hospital level of care. That is, partial hospital treatment is designed to ameliorate symptoms to the point that outpatient treatment is appropriate. The objective is not to return patients to the normal population, as is the goal of outpatient therapy; thus, this standard may be too stringent for the evaluation of a partial hospital program. Future research investigating the effectiveness of higher levels of care (e.g., inpatient, intensive outpatient programs, etc.) may also benefit from modifications to traditional estimates of "successful" outcomes consistent with treatment goals. In addition, research investigating which patients are most likely to benefit from partial hospitalization, as opposed to those who might improve with inpatient hospitalizations, is needed to both enhance patient outcomes and reduce health care costs.

Data in the current study have many of the limitations common in effectiveness research (for a full discussion, see Westbrook & Kirk, 2005). The study had no control group and we cannot rule out the possibility that improvement in symptoms may potentially be a result of change over time and regression to the mean (regardless of treatment). However, findings regarding the chronic nature of depression (Wiersma et al., 2011), and the similarities between our results and those reported in benchmark studies, reduce the probability that change occurred due to the passage of time alone. In addition, no data were available on patients' current medications and these effects were therefore not accounted for in the current analyses. Attrition levels in this sample (25%) were also higher than the average benchmark study (17.7%), but comparable to other naturalistic studies of CBT for depression (e.g., 22.5% attrition rate reported by Westbrook & Kirk, 2005). Although treatment completers and non-completers did not differ in pre-treatment CESD-10 depression, acute worsening of symptoms may be more likely in this sample than in outpatient samples, putting participants at higher risk for inpatient admission (and thus for dropout from the present study). In addition, although we used multiple imputation to address missing data, we were unable obtain data from non-completers. Outcome studies that could track patients who fail to complete treatment would be tremendously beneficial. Thus, findings from this study should be interpreted with these limitations in mind.

Conclusions

The current study extends previous research on the effectiveness of CBT by examining a large sample of partial hospital patients with severe depression treated with brief, manualized CBT. Our findings indicate that treatment was associated with reduced depressive symptoms and improved well-being. Findings also provide additional support for the effectiveness of CBT across the continuum of care and in diverse treatment provision settings.

References

Andersen, E.M., Malmgren, J.A., Carter, W.B., & Patrick, D.L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (center for epidemiologic studies depression scale). *American Journal of Preventive Medicine*, *10*, 77–84.

Bagby, R., Quilty, L.C., Segal, Z.V., McBride, C., Kennedy, S., & Costa, P. (2008). Personality and differential treatment response in major depression: A randomized controlled trial comparing cognitive-behavioural therapy and pharmacotherapy. *The Canadian Journal of Psychiatry*, *53*, 361–370.

Beck, A.T., Emergy, G., & Greenberg, R.L. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York, NY: Basic Books.

Beck, A.T., Rush, J.A., Sahw, B.F., & Emery, G. (1979). *Cognitive therapy for depression*. New York, NY: Guilford Press.

Blais, M.A., Lenderking, W.R., Baer, L., deLorell, A., Peets, K., Leahy, L., & Burns, C. (1999). Development and initial validation of a brief mental health outcome measure. *Journal of Personality Assessment*, *73*, 359–373. doi:10.1207/S15327752JPA7303_5

Blais, M.R., Kehl-Fie, K., & Blais, M.A. (2008). The Schwartz Outcome Scale©: A brief measure of psychological well-being. *Patient Reported Outcomes Newsletter*, *39*, 19–21.

Bodenmann, G., Plancherel, B., Beach, S.H., Widmer, K., Gabriel, B., Meuwly, N., & Schramm, E. (2008). Effects of coping-oriented couples therapy on depression: A

randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *76*, 944–954. doi:10.1037/a0013467

David, D., Szentagotai, A., Lupu, V., & Cosman, D. (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: A randomized clinical trial, posttreatment outcomes, and six-month follow-up. *Journal of Clinical Psychology*, *64*, 728–746. doi:10.1002/jclp.20487

DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., Shelton, R.C., Young, P.R., Salomon, R.M., & Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, *62*, 409–416. doi:10.1001/archpsyc.62.4.409

Dimidjian, S., Hollon, S.D., Dobson, K.S., Schmaling, K.B., Kohlenberg, R.J., Addis, M. E., & Jacobson, N.S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, *74*, 658–670. doi:10.1037/0022-006X.74.4.658

Dobson, K.S., Hollon, S.D., Dimidjian, S., Schmaling, K.B., Kohlenberg, R.J., Gallop, R. J., & Jacobson, N.S. (2008). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology*, *76*, 468–477. doi:10.1037/0022-006X.76.3.468

Donohue, J.M., & Pincus, H. (2007). Reducing the societal burden of depression: A review of economic costs, quality of care and effects of treatment. *Pharmacoeconomics*, *25*, 7–24 doi:10.2165/00019053-200725010-00003

Dozois, D.A., Bieling, P.J., Patelis-Siotis, I., Hoar, L., Chudzik, S., McCabe, K., & Westra, H.A. (2009). Changes in self-schema structure in cognitive therapy for major depressive disorder: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *77*, 1078–1088. doi:10.1037/a0016886

Eisen, S.V., Normand, S.L.T., Belanger, A.J., Spiro, A., & Esch, D. (2004). The revised behavior and symptom identification scale BASIS-R. *Medical Care*, *42*, 1230–1241. doi:10.1097/00005650-200412000-00010

Faramarzi, M., Alipor, A., Esmaelzadeh, S., Kheirkhah, F., Poladi, K., & Pash, H. (2008). Treatment of depression and anxiety in infertile women: Cognitive behavioral therapy versus fluoxetine. *Journal of Affective Disorders*, *108*, 159–164. doi:10.1016/j.jad.2007.09.002

Freedland, K.E., Skala, J.A., Carney, R.M., Rubin, E.H., Lustman, P.J., Dávila-Román, V.G., & Hogue, C.W. (2009). Treatment of depression after coronary artery bypass surgery: A randomized controlled trial. *Archives of General Psychiatry*, *66*, 387–396. doi:10.1001/archgenpsychiatry.2009.7

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic.

Hopko, D.R., Lejuez, C.W., LePage, J.P., Hopko, S.D., & McNeil, D.W. (2003a). A brief behavioral activation treatment for depression: A randomized pilot trial within an inpatient psychiatric hospital. *Behavior Modification*, *27*, 458–469. doi:10.1177/0145445503255489

Hopko, D., Stanley, M., Reas, D., Wetherell, J., Beck, J., Novy, D., & Averill, P.M. (2003b). Assessing worry in older adults: Confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment*, *15*, 173–183. doi:10.1037/1040-3590.15.2.173

Jacobson, N.S., Roberts, L.J., Berns, S.B., & McGlinchey, J.B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, *67*, 300–307. doi:10.1037/0022-006X.67.3.300

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. doi:10.1037/0022-006X.59.1.12

Kertz, S.J., Bigda-Peyton, J.S., Rosmarin, D.H., & Björgvinsson, T. (2012). The importance of worry across diagnostic presentations: Prevalence, severity, and associated symptoms in a partial hospital setting. *Journal of Anxiety Disorders*, *26*, 126–133.

Kessler, R.C., Chiu, W.T., Demler, O., & Walters, E.E. (2005). Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry*, *62*, 617–627. doi:10.1176/appi.ajp.163.6.1074

Kiser, L.J., Lefkovitz, P.M., Kennedy, L.L., & Knight, M.A. (2010). *Continuum of ambulatory behavioral health services*. Association of ambulatory behavioral health services. Retrieved from http://www.aabh.org/content/continuum-ambulatory-behavioral-health-services

Klein, D.N., Santiago, N.J., Vivian, D., Blalock, J. A., Kocsis, J.H., Markowitz, J.C., & Keller, M. B. (2004). Cognitive-behavioral analysis system of psychotherapy as a maintenance treatment for chronic depression. *Journal of Consulting and Clinical Psychology*, *72*, 681–688. doi:10.1037/0022-006X.72.4.681

Kocsis, J.H., Gelenberg, A.J., Rothbaum, B.O., Klein, D.N., Trivedi, M.H., Manber, R., & Thase, M.E. (2009). Cognitive behavioral analysis system of psychotherapy and brief supportive psychotherapy for augmentation of antidepressant nonresponse in chronic depression: The REVAMP trial. *Archives of General Psychiatry*, *66*, 1178–1188. doi:10.1001/archgenpsychiatry.2009.144

Kuyken, W., Byford, S., Taylor, R.S., Watkins, E., Holden, E., White, K., & Teasdale, J.D. (2008). Mindfulness-based cognitive therapy to prevent relapse in recurrent depression. *Journal of*

Consulting and Clinical Psychology, *76*, 966–978. doi:10.1037/a0013786

Laidlaw, K., Davidson, K., Toner, H., Jackson, G., Clark, S., Law, J., & Cross, S. (2008). A randomised controlled trial of cognitive behaviour therapy vs treatment as usual in the treatment of mild to moderate late life depression. *International Journal of Geriatric Psychiatry*, *23*, 843–850. doi:10.1002/gps.1993

Linehan, M.M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: Guilford Press.

Manber, R., Edinger, J.D., Gress, J.L., Pedro-Salcedo, M., Kuo, T.F., & Kalista, T. (2008). Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disorder and insomnia. *Sleep: Journal of Sleep and Sleep Disorders Research*, *31*, 489–495.

Martell, C.R., Dimidjian, S., & Herman-Dunn, R. (2010). *Behavioral activation for depression: A clinician's guide*. New York, NY: Guilford Press.

McEvoy, P.M., & Nathan, P. (2007). Effectiveness of cognitive behavior therapy for diagnostically heterogeneous groups: A benchmarking study. *Journal of Consulting and Clinical Psychology*, *75*, 344–350. doi:10.1037/0022-006X.75.2.344

Meyer, T.J., Miller, M.L., Metzger, R.L., & Borkovec, T.D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *28*, 487–495. doi:10.1016/0005-7967(90)90135-6

Miranda, J., Chung, J.Y., Green, B.L., Krupnick, J., Siddique, J., Revicki, D.A., & Belin, T. (2003). Treating depression in predominantly Low-income young minority women: A randomized controlled trial. *Journal of the American Medical Association*, *290*, 57–65. doi:10.1001/jama.290.1.57

Miranda, J., Green, B.L., Krupnick, J.L., Chung, J., Siddique, J., Belin, T., & Revicki, D. (2006). One-year outcomes of a randomized clinical trial treating depression in low-income minority women. *Journal of Consulting and Clinical Psychology*, *74*, 99–111. doi:10.1037/0022-006X.74.1.99

Neuhaus, E.C., Christopher, M., Jacob, K., Guillaumot, J., & Burns, J.P. (2007). Short-term cognitive behavioral partial hospital treatment: A pilot study. *Journal of Psychiatric Practice*, *13*, 298–307. doi:10.1097/01.pra.0000290668.10107.f3

Perlis, R.H., Nierenberg, A.A., Alpert, J.E., Pava, J., Matthews, J.D., Buchin, J., & Fava, M. (2002). The effects of adding cognitive therapy to fluoxetine dose increase on risk of relapse and residual depressive symptoms in continuation treatment of major depressive disorder. *Journal of Clinical Psychopharmacology*, *22*, 474–480. doi:10.1097/00004714-200210000-00006

Persons, J.B., Bostrom, A., & Bertagbolli, A. (1999). Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cognitive Therapy and Research*, *23*, 535–548. doi:10.1023/A:1018724505659

Rohan, K.J., Roecklein, K.A., Tierney-Lindsey, K., Johnson, L.G., Lippy, R.D., Lacy, T.J., & Barton, F.B. (2007). A randomized controlled trial of cognitive-behavioral therapy, light therapy, and their combination for seasonal affective disorder. *Journal of Consulting and Clinical Psychology*, *75*, 489–500. doi:10.1037/0022-006X.75.3.489

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: J. Wiley & Sons.

Sava, F.A., Yates, B.T., Lupu, V., Szentagotai, A., & David, D. (2009). Cost-effectiveness and cost-utility of cognitive therapy, rational emotive behavioral therapy, and fluoxetine (Prozac) in treating clinical depression: A randomized clinical trial. *Journal of Clinical Psychology*, *65*, 36–52. doi:10.1002/jclp.20550

Serfaty, M., Haworth, D., Blanchard, M., Buszewicz, M., Murad, S., & King, M. (2009). Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: A randomized controlled trial. *Archives of General Psychiatry*, *66*, 1332–1340. doi:10.1001/archgenpsychiatry.2009.165

Sheehan, D.V., Lecrubier, Y., Sheehan, H.K., Amorim, P., Janavs, J., Weiller, E., & Dunbar, G.C. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(suppl 20), 22–33.

Smit, A.A., Kluiter, H.H., Conradi, H.J., van der Meer, K.K., Tiemens, B.G., Jenner, J.A., & Ormel, J.J. (2006). Short-term effects of enhanced treatment for depression in primary care: Results from a randomized controlled trial. *Psychological Medicine*, *36*, 15–26. doi:10.1017/S0033291705006318

Strauman, T.J., Vieth, A.Z., Merrill, K.A., Kolden, G.G., Woods, T.E., Klein, M.H., & Kwapil, L. (2006). Self-system therapy as an intervention for self-regulatory dysfunction in depression: A randomized comparison with cognitive therapy. *Journal of Consulting and Clinical Psychology*, *74*, 367–376. doi:10.1037/0022-006X.74.2.367

Weersing, V., & Weisz, J.R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology*, *70*, 299–310. doi:10.1037/0022-006X.70.2.299

Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy*, *43*, 1243–1261. doi:10.1016/j.brat.2004.09.006

Wiersma, J.E., Titov, N., Van Schaik, D.J.F., Van Oppen, P., Beekman, A.T.F., Cuijpers, P., & Andrews, G.A. (2011). Treating chronic symptoms of depression in the virtual clinic: Findings on chronicity of depression in patients treated with internet-based computerized cognitive

behaviour therapy for depression. *Psychotherapy and Psychosomatics*, *80*, 313–315. doi:10.1159/000323167

Wright, J.H., Wright, A.S., Albano, A., Basco, M. R., Goldsmith, L., Raffield, T., & Otto, M.W. (2005). Computer-assisted cognitive therapy for depression: Maintaining efficacy while reducing therapist time. *The American Journal of Psychiatry*, *162*, 1158–1164. doi:10.1176/appi.ajp.162.6.1158

Ybema, J.F., & van den Bos, K. (2010). Effects of organizational justice on depressive symptoms and sickness absence: A longitudinal perspective. *Social Science & Medicine*, *70*, 1609–1617.

Young, J.L., Waehler, C.A., Laux, J.M., McDaniel, P.S., & Hilsenroth, M.J. (2003). Four Studies extending the utility of the Schwartz Outcome Scale (SOS-10). *Journal of Personality Assessment*, *80*, 130–138. doi:10.1207/S15327752JPA8002_0