Shorter communication

Convergence and divergence in the delivery of cognitive therapy in two randomized clinical trials

Christian A. Webb [a,*], Robert J. DeRubeis [a], Steven D. Hollon [b], Sona Dimidjian [c], Jay D. Amsterdam [d], Richard C. Shelton [e]

[a] University of Pennsylvania, Department of Psychology, 3720 Walnut Street, Solomon Lab Bldg, Philadelphia, PA 19104, USA
[b] Vanderbilt University, Department of Psychology, 2301 Vanderbilt Place, Nashville, TN 37240, USA
[c] University of Colorado at Boulder, Department of Psychology, 345 UCB, Boulder, CO 80309, USA
[d] University of Pennsylvania, Depression Research Unit, 3535 Market Street, 3rd Floor, Philadelphia, PA 19104, USA
[e] Vanderbilt School of Medicine, Department of Psychiatry, Vanderbilt Psychiatric Hospital, 1601 23rd Ave. S., Nashville, TN 37212, USA

ARTICLE INFO

ABSTRACT

Objective: Research indicates that cognitive therapy (CT) can be differentiated from other treatment modalities based on in-session therapist behavior. However, to our knowledge, consistency in the implementation of individual CT across clinical trials has not been tested. We compared therapist adherence to CT, as well as the therapeutic alliance, in two randomized clinical trials (RCTs) of depression treatment.
Method: Data were drawn from two highly cited RCTs of CT for major depression, representing a total of three sites. Trained raters coded sessions for therapist adherence to CT and the therapeutic alliance.
Results: Significant differences were obtained between sites in overall level of adherence to CT, therapist emphasis on cognitive vs behavioral strategies, and therapist focus on homework. In contrast, no significant differences emerged in the collaborative structure of CT and in the therapeutic alliance.
Conclusions: Despite efforts to maximize the consistency of CT implementation (e.g., via the use of the same treatment manuals, delivered by carefully-selected and experienced therapists), differences in the implementation of CT can result. Although preliminary, these findings raise questions regarding the uniformity of CT delivery across published clinical trials, and underline the importance of assessing treatment integrity, both across clinical trials and in dissemination research.

© 2013 Elsevier Ltd. All rights reserved.

In an effort to ensure that treatments are implemented as intended, psychotherapy researchers have emphasized the importance of monitoring *treatment integrity* (Perepletchikova & Kazdin, 2005), which has been defined as consisting of three components: therapist *adherence* (i.e., the extent to which therapists deliver the prescribed procedures of a given treatment modality, and avoid proscribed procedures), therapist *competence* (i.e., the skill with which these procedures are implemented); and *treatment differentiation* (i.e., when more than one treatment is being investigated, demonstrating that the conditions can be distinguished from one another along critical dimensions; Waltz, Addis, Koerner, & Jacobson, 1993).

Treatment differentiation in regard to theory-specified therapist behavior has been documented in comparative studies of psychotherapies. CT has been differentiated from other treatment modalities based on assessments of transcript material or audio or video recordings of sessions, including interpersonal therapy (e.g., DeRubeis, Hollon, Evans, & Bemis, 1982; Hill, O'Gray, & Elkin, 1992), psychodynamic psychotherapy (e.g., Watzke, Rueddel, Koch, Rudolph, & Schulz, 2008; for reviews, see Blagys & Hilsenroth, 2000, 2002), supportive-expressive psychotherapy (e.g., Luborsky, Woody, McLellan, O'Brien, & Rosenzweig, 1982) and drug counseling (e.g., Luborsky et al., 1982). However, we are aware of no published study that has examined the consistency of the implementation of individual CT across different clinical trials. Treatment outcome studies may assess and report therapist adherence/competence data in their individual outcome trials. However, conclusions about the consistency of treatment delivery across studies are limited until adherence/competence are assessed using the same set of raters (who undergo the same training), using the same measures and assessing adherence/competence at the same time points across studies. Moreover, in addition to comparing overall levels of therapist adherence/competence (i.e., overall mean

* Corresponding author. Tel.: +1 617 855 4429.
  E-mail address: cwebb@mclean.harvard.edu (C.A. Webb).

on an adherence or competence scale) across studies, it is also important to examine adherence to the components of treatment (e.g., two studies may report similar mean scores on a measure of therapist adherence to CT, yet significantly differ in the emphasis therapists' placed on the differing components of CT, such as the use of cognitive vs behavioral techniques, homework assignment and review, etc.).

Although Malik, Beutler, Alimohamed, Gallagher-Thompson, and Thompson (2003) did not address the question of the consistency of individual CT across settings, they did compare levels of psychotherapy process variables in three different manual-based formats of CT (individual CT for major depression, group CT for major depression and couples CT for alcohol dependence and depression) as well as in six theoretically-diverse but noncognitive treatments. Using the observer-rated Systematic Treatment Selection Therapy Process Rating Scale (TPRS; Beutler, Clarkin, & Bongar, 2000), the authors compared the treatments on dimensions designed to assess: (1) therapist directiveness, (2) the extent to which the therapist made efforts to deepen the client's in-session emotional experience/arousal, (3) the degree of behavioral vs insight-focused interventions, and (4) the quality of the therapeutic alliance. Malik et al. (2003) found significant differences across the different variants of CT on all of the dimensions except therapist directiveness. As the authors note, the differences in levels of process variables across the CTs could have resulted, at least in part, from the fact that the treatments represented different formats of CT, and they targeted different populations. Moreover, they did not assess adherence to the CT protocol per se, but rather examined a range of broader process variables (e.g., therapist directiveness, the alliance) using a measure (TPRS) designed to differentiate a variety of different forms of therapy.

In the current exploratory study, we directly compared treatment processes in CT for depression at three sites, from two large randomized clinical trials (RCTs; DeRubeis et al., 2005; Dimidjian et al., 2006). Both studies were of individual CT, provided for a total of 16 weeks, for adults diagnosed with DSM-IV Major Depressive Disorder (MDD). Therapists in both studies followed the same CT treatment manuals (Beck, 1995; Beck, Rush, Shaw, & Emery, 1979). In the current study, by applying the Collaborative Study Psychotherapy Rating Scale-Cognitive-Behavioral subscale (CSPRS-CB; Hollon et al., 1988), a commonly-employed measure of therapist adherence to CT, to tape recordings of sessions from these studies, we were able to test for differences in the provision of CT across sites on central CT dimensions, including overall adherence to the CT protocol, delivery of cognitive and behavioral techniques, collaborative structure, and homework assignment. In addition, given the substantial attention the therapeutic alliance has received in the psychotherapy process literature (Horvath, Del Re, Fluckiger, & Symonds, 2011), sites were also compared for the quality of the alliance. To control and allow for tests of the influence of therapist effects on treatment adherence and alliance, therapists were included as a term in our statistical models testing for site differences (see below).

Method

Participants

Patients

All patients from the CT conditions of the Cognitive Pharmacotherapy-II (CPT-II; $N = 60$; DeRubeis et al., 2005) and University of Washington (UW; $N = 45$; Dimidjian et al., 2006) studies were included. The CPT-II study consisted of two sites: one at the University of Pennsylvania and the other at Vanderbilt University ($N = 30$ patients at each site). Both studies targeted adults with MDD, and CT was provided for a total of 16 weeks.

Therapists

In the CPT-II study, four male and two female clinicians (three therapists at each site) served as cognitive therapists. In the UW study, two male and one female clinician served as cognitive therapists (see [1] for additional information on study therapists).

Measures

Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon et al., 1988)

We utilized the CSPRS' Cognitive-Behavioral (CB) scale, which has been used to rate the extent to which therapists adhere to CBT procedures (e.g., Feeley, DeRubeis, & Gelfand, 1999). The CB scale consists of 28 items organized into 6 subscales, three of which (CB1-CB3) assess therapist use of cognitive methods: Cognitive Rationale (CB1; 3 items), Assessing Cognitive Processes (CB2; 5 items), Evaluating/Changing Beliefs (CB3; 7 items), Behavioral Focus (CB4; 4 items), Homework (CB5; 3 items), and Collaborative Structure (CB6; 6 items). Items are rated on a 7-point scale. The overall means on the CSPRS adherence scale reported in the current study are similar to those reported in previous studies (e.g., Shaw et al., 1999). Previous research suggests that the CB scale can be rated reliably (Hill et al., 1992).[2]

Working Alliance Inventory observer-rated version, short form (WAI-O-S; Tracey & Kokotovic, 1989)

The WAI-O-S is a 12-item observer-rated measure designed to assess the therapeutic alliance. It is a shortened and modified version of the original WAI scales. Based on Bordin's (1979) conceptualization of the alliance, the WAI-O-S consists of three subscales, each with four items: 1) the *bond* between therapist and patient, 2) agreement about the *goals* of therapy and 3) agreement about the *tasks*. Items are rated on a 7-point scale (0 = never to 6 = always). Previous research suggests that the WAI-O-S, referred to henceforth as the WAI, can be rated reliably (Intraclass Correlation Coefficient [ICC] = .77; Strunk, Brotman, & DeRubeis, 2010).

Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960)

The modified 17-item HRSD is a commonly-employed interview-based measure of depressive severity. This modified version includes atypical sleep, appetite, and weight symptoms. It was administered by clinical evaluators weekly for the first 4 weeks of treatment, and biweekly from week 6 to week 16 in the CPT-II

---

study; and at pretreatment, midtreatment, post-treatment, and nonstandard assessments as required (e.g., early termination) in the UW study.

Procedures

Therapy sessions in both studies were video and audiotaped. Session 3 was rated for each therapist–patient dyad.[3] Raters, who took notes while viewing each session, provided ratings after the end of the session. If a videotape was unavailable, we substituted either a videotape from an adjacent session or an audiotape of the session. One patient from each of the studies was not included because they dropped out of treatment prior to session 2. Recordings of sessions for 3 patients were unavailable from the UW study. Thus, a total of 100 sessions was rated.

Random numbers were assigned to the tapes, from which all other identifying information was removed. Similar to previous published studies assessing therapist adherence to CT and the alliance (DeRubeis & Feeley, 1990; Feeley et al., 1999; Strunk, Cooper, Ryan, DeRubeis, & Hollon, 2012; Strunk, DeRubeis, Chui, & Alvarez, 2007), five advanced undergraduates served as raters. Each tape was coded by 2 raters independently according to a balanced incomplete block design (Fleiss, 1981), such that each rater was paired with each of the other 4 raters an equal number of times. Raters were blind to treatment outcome, site, and the study aims, as well as to participant and session number. Raters read Beck (1995), reviewed all available manuals for the rating scales, and received approximately 30 h of training prior to the initiation of the study. In addition, in order to reduce rater drift, raters met approximately once each week with the study supervisor (CAW) to rate a tape independently and to discuss any discrepancies in ratings. Interrater reliabilities are detailed below.

Results

Means, standard deviations (SDs), intercorrelations and interrater reliabilities (ICCs) for all rated variables are listed in Table 1. The ICCs were estimated with a random effects model (Shrout & Fleiss, 1979) for the pooled (averaged) ratings of two raters. ICCs were adequate and similar to those reported in studies assessing therapist adherence to CT and the alliance (e.g., DeRubeis & Feeley, 1990; Feeley et al., 1999; Strunk et al., 2010), with the exception of the ICC of .49 for the Collaborative Structure subscale of the CPRS, which was lower than desired.

Outcome differences between sites

Intake HRSD scores differed as a function of site ($F(2,102) = 17.90$, $p < .0001$; Penn: 24.80 ± 3.69 SD, Vanderbilt: 23.20 ± 2.86, UW: 20.02 ± 3.80). Pairwise contrasts (Tukey–Kramer method) revealed higher scores at Penn and at Vanderbilt, each relative to UW (Cohen's $d = 1.27$; $p < .0001$, and $d = .92$; $p < .001$, respectively), consistent with entry criteria in each study (see DeRubeis et al., 2005; Dimidjian et al., 2006 for details). Site differences on post-treatment HRSD were not significant when we adjusted for intake scores and included 'therapist' as a fixed factor (nested within site), in a General Linear Model (GLM; $F(2,94) = 1.26$,

$p = .289$). Post-treatment HRSD least-squares (LS) means ($\pm$SE) were 9.10 ± 1.40 (Penn), 11.45 ± 1.31 (Vanderbilt) and 12.05 ± 1.19 (UW).

Differences in process variables between sites

To test for site differences in the level of investigated process variables, we applied a series of GLMs, in each of which site was a factor, as was therapist (nested within site); intake HRSD was a covariate. When the 'site' term in a model was significant, we conducted pairwise contrasts (Tukey–Kramer method). Table 2 reports LS means and SDs for each variable at each site, as well as significance levels associated with tests of site differences, and the associated effect size estimates (Cohen's d). All analyses were performed using SAS Version 9.2 PROC GLM (SAS Institute, Cary, NC).

Significant site differences emerged in overall CT adherence (CSPRS-CB; $F(2,90) = 6.11$, $p = .003$). Vanderbilt therapists received significantly higher mean adherence scores in comparison to both the UW and Penn therapists. With respect to the cognitive strategies (i.e., average of CB1 [Cognitive Rationale], CB2 [Assessing Cognitive Processes], CB3 [Evaluating and Changing Beliefs] items), there was a significant difference across sites, $F(2,90) = 9.07$, $p < .001$. The Vanderbilt and Penn means were higher than the UW mean. As shown in Fig. 2, the observed cognitive strategies mean for each of the UW therapists was numerically lower than the observed mean of each of the therapists from either of the CPT-II sites (Vanderbilt or Penn).

There were significant site differences on five of the six CSPRS subscales (not CB6, see below; and see Fig. 1). F values (2,90) ranged from 3.89 to 6.75, ps from .024 to .002. As shown in Table 2, Vanderbilt therapists received significantly higher ratings than UW therapists on all three cognitive subscales (CB1–CB3). The CB3 Penn mean was also higher than the UW mean. Indeed, on CB2 and CB3 methods, the numeric mean for each of the UW therapists was lower than those of any of the CPT-II therapists. In contrast, the CB4 (Behavioral) mean at UW was significantly higher than the Penn mean. In fact, the numeric CB4 mean for all but one of the CPT-II therapists was lower than the means of any of the UW therapists (see Fig. 3).[4] Vanderbilt therapists were rated more highly on CB5 (Homework) than their Penn counterparts. There were no site differences on either CB6 (Collaborative Structure; $F(2,90) = 1.57$, $p = .214$) or in WAI scores ($F(2,90) = .85$, $p = .430$). A significant therapist effect emerged for Homework (CB5; $F(9,90) = 2.34$, $p = .038$).

Discussion

The present study investigated several process variables in an effort to examine the consistency in the implementation of CT across three different sites, from two large RCTs (DeRubeis et al., 2005; Dimidjian et al., 2006). Both of these studies delivered 16 weeks of the same manualized treatment modality (i.e., individual CT) and targeted the same disorder (i.e., adult patients with DSM-IV MDD). One of the purposes of treatment manuals is to enhance the uniformity of treatment implementation. Given that the CT conditions in each study were based on the same two CT manuals (Beck,

---

[3] Early assessments (e.g., sessions 2, 3 or 4) of therapist adherence are common in studies investigating adherence–outcome relations. Moreover, and importantly, DeRubeis and Feeley (1990) and Feeley et al. (1999) found, in two separates studies of CT for depression, that therapists were dedicating at least as much time and effort to CT techniques early in treatment (session 2) in comparison to several randomly selected sessions later in treatment.

[4] A similar pattern of between-site differences emerged when the above analyses were rerun using the "cognitive methods" ($F(2,90) = 6.49$, $p = .002$) and "behavioral methods/homework" ($F(2,90) = 11.06$, $p < .0001$) categories derived from a recent factor analysis (Strunk et al., 2012; also see, Strunk et al., 2010), rather than using the cognitive (average of CB1–CB3 items) and the behavioral focus (CB4) variables from the current study. (The only differences being that, in the latter analysis, the Penn therapists were rated as using significantly fewer behavioral methods than their Vanderbilt [$p = .002$] counterparts.)

Table 1
Raw means (M), standard deviations (SD), intraclass correlation coefficients (ICCs) and intercorrelations for all process variables.

Variable	M	SD	ICC	1	2	3	4	5	6	7	8	9
1. CB1	2.10	.72	.71	–	.37**	.43**	.68**	−.15	.30**	.31**	.61**	.24*
2. CB2	2.38	.54	.74		–	.56**	.80**	−.20*	.09	.35**	.66**	.21*
3. CB3	1.88	.57	.72			–	.89**	−.20*	.10	.33**	.73**	.17†
4. CB-Cog	2.09	.48	.81				–	−.23*	.18†	.41**	.83**	.24*
5. CB4	2.30	.60	.74					–	.24*	.40**	.24*	.21*
6. CB5	2.88	.72	.67						–	.36**	.51**	.36**
7. CB6	3.02	.45	.49							–	.76**	.35**
8. CB-Total	2.40	.35	.68								–	.41**
9. WAI	4.38	.46	.73									–

Note: N = 100 for all variables. CB1–CB6 = Subscales from Collaborative Study Psychotherapy Rating Scale (CSPRS): CB1 = Cognitive Rationale subscale; CB2 = Assessing Cognitive Processes subscale; CB3 = Evaluating and Changing Beliefs subscale; CB4 = Behavioral Focus subscale; CB5 = Homework subscale; CB6 = Collaborative Structure subscale; CB-Cog = Average of the items from the CB1–CB3 subscales (i.e., cognitive items from the CSPRS). CB-Total = Average of all of the items from the CSPRS; WAI = Working Alliance Inventory.
† < .10; *p < .05; **p < .01.

1995; Beck et al., 1979), the present study allowed us to test for differences in treatment implementation in the context of manualized individual CT for clinical depression delivered by carefully-selected and experienced therapists.

Significant differences emerged between sites on several of the process variables examined, including in the overall level of therapist adherence to CT methods, with the highest levels observed at the Vanderbilt site. When examining adherence to cognitive vs behavioral techniques, the CPT-II therapists (Penn and Vanderbilt sites) were judged to deliver significantly greater levels of cognitive techniques than their UW counterparts. In contrast, the UW therapists were judged to deliver significantly greater behavioral strategies, relative to the Penn therapists. Differences between sites were not observed in the quality of the therapeutic alliance or in therapist adherence to the collaborative structure of CT. Overall, these findings may reflect the fact that therapists were relatively consistent across sites in their ability to keep to the basic session framework of CT, as well as in their efforts to develop and maintain an alliance with patients. However, within this basic structure, therapists at different sites did differ in the emphases they placed on specific CT procedures. Of course, it is important to note that one interpretation of the differences in adherence between sites is that

Table 2
Least-squares means, standard deviations and effect size estimates (Cohen's d) for differences in process variables between sites.

Variable	Penn (P)		Vanderbilt (V)		UW (W)		Effect sizes (d)		
	Mean	SD	Mean	SD	Mean	SD	P vs V	P vs W	V vs W
CB1	2.04	.76	2.45	.71	1.88	.77	−.57†	.21	.78**
CB2	2.44	.57	2.62	.53	2.15	.58	−.33	.51	.85**
CB3	2.04	.59	2.09	.55	1.60	.60	−.09	.75*	.86**
CB-Cog	2.18	.49	2.34	.46	1.84	.50	−.34	.70*	1.05**
CB4	1.97	.63	2.34	.58	2.48	.64	−.62†	−.81**	−.23
CB5	2.59	.70	3.03	.65	3.01	.71	−.66*	−.60†	.03
CB6	2.98	.49	3.14	.46	2.96	.50	−.34	.04	.38
CB-Total	2.36	.37	2.59	.34	2.30	.38	−.66*	.16	.81**
WAI	4.31	.50	4.46	.46	4.36	.50	−.32	−.10	.21

Note: Penn N = 29; Vanderbilt N = 30; UW N = 41. Least-squares means are displayed, adjusting for intake HRSD and the effect of therapist. Penn = University of Pennsylvania (Site A from DeRubeis et al., 2005); Vanderbilt = Vanderbilt University (Site B from DeRubeis et al.); UW = University of Washington study (Dimidjian et al., 2006). CB1–CB6 = Subscales from Collaborative Study Psychotherapy Rating Scale (CSPRS): CB1 = Cognitive Rationale subscale; CB2 = Assessing Cognitive Processes subscale; CB3 = Evaluating and Changing Beliefs subscale; CB4 = Behavioral Focus subscale; CB5 = Homework subscale; CB6 = Collaborative Structure subscale; CB-Cog = Average of the items from the CB1–CB3 subscales (i.e., cognitive items from the CSPRS). CB-Total = Average of all of the items from the CSPRS; WAI = Working Alliance Inventory.
† < .10; *p < .05; **p < .01.

they reflect therapists responding *appropriately* to between-site differences in patient variables. In other words, there may be certain unmeasured patient characteristics that help account for site differences in therapist adherence. There were, for example, differences in intake criteria between the CPT-II and UW studies, resulting in, on average, more severely depressed patients being included in the former study. Pretreatment depression severity was controlled for in the analyses in this study.

It is also important to highlight that although the two aforementioned CT texts (Beck, 1995; Beck et al., 1979) are commonly referred to as "treatment manuals", they are not highly structured manuals with step-by-step instructions for each therapy session (c.f. CBT manuals for anxiety disorders, e.g., Barlow, 2008; Hope, Heimberg, & Turk, 2010; and for eating disorders, Fairburn, 2008). Rather, they are clinical guides that describe a series of core principles, treatment procedures and interventions, rooted in and informed by the cognitive model. The texts provide therapists with a cognitive-behavioral framework for guiding their sessions, allowing clinicians substantial flexibility to tailor interventions to the idiosyncratic needs of each patient. Thus, differences across therapists and sites in therapist adherence may be to a certain extent expected. It may be that the more structured manuals, such as those cited above, yield greater consistency in implementation across therapists and clinical trials.

Only one significant therapist effect emerged, suggesting that the observed differences in levels of process variables were generally accounted for by differences in emphasis placed on techniques between sites, rather than due to differences across therapists. Indeed, there were intriguing consistencies in mean levels between therapists within a site for several of the variables we examined. However, power to detect therapist effects was limited.

It is not uncommon for clinical trials to report some form of treatment integrity data (e.g., therapist adherence or competence ratings). However, even when the same measures of treatment integrity are utilized across studies, the interpretation of between-study differences is constrained by the fact that different sets of raters, who are trained independently of one another, are typically employed. One strength of the current study is that the same set of raters coded variables at all three sites, while applying the same measure of therapist adherence to the same sessions across sites, facilitating the interpretation of between-site differences.

A limitation of the present study is that investigated process variables were only assessed at a single session, because observer-rated assessments of in-session processes are labor- and time-intensive. It may be that a different pattern of findings would have emerged if the process variables were assessed at different time points in treatment. For example, in contrast to the current pattern
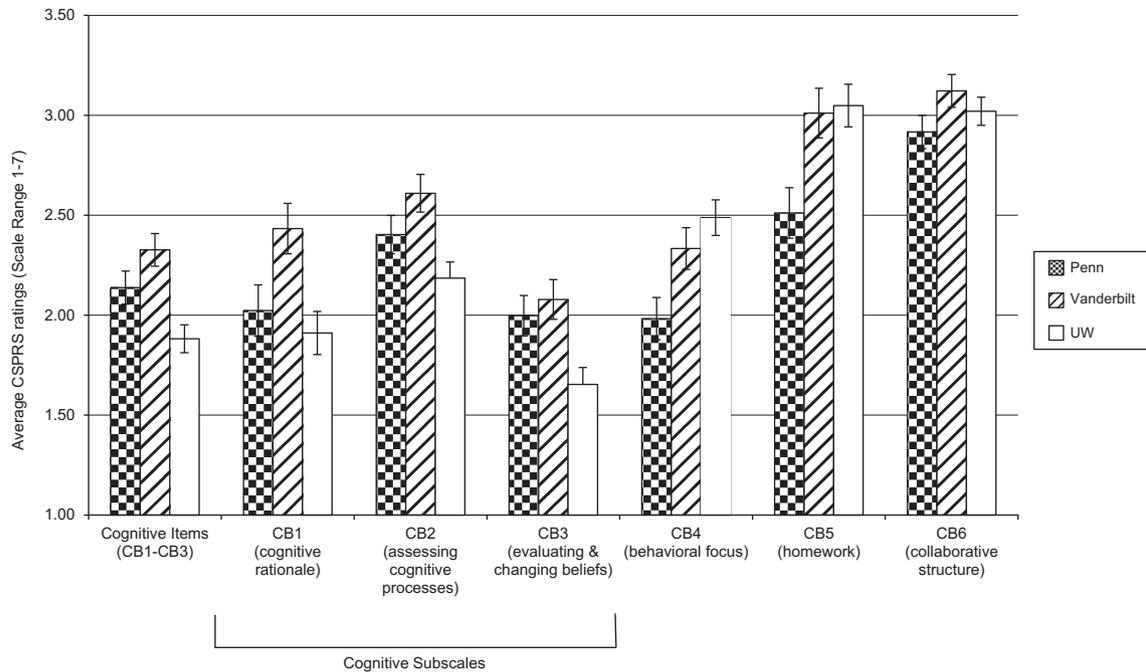
Fig. 1. Site means for therapist adherence to CT variables, as rated by the Collaborative Study Psychotherapy Rating Scale (CSPRS). Raw means are displayed. For least-squares means, adjusting for intake HRSD scores and the effect of therapist, see Results and Table 2. Error bars represent standard error. Penn = University of Pennsylvania (Site A from DeRubeis et al., 2005); Vanderbilt = Vanderbilt University (Site B from DeRubeis et al.); UW = University of Washington study (Dimidjian et al., 2006).

of findings, Dimidjian et al. (2006), utilizing a modified version of the CSPRS applied to a different set of CT sessions, and to a slightly smaller sample of therapist–patient dyads from their trial, reported (numerically) higher mean ratings for cognitive items in comparison to behavioral items. In addition, in the current study, interrater reliability for the Collaborative Structure subscale of the CPSRS (CB6) was lower than desired. Thus, the findings related to this subscale should be interpreted with caution. Increased training in coding CSPRS items or the use of graduate, rather than undergraduate, raters may help improve interrater reliability.

It will be important to replicate this inquiry in other contexts. For example, these variables could be assessed at different time points (e.g., randomly select two sessions per therapist–patient dyad to achieve a broader sampling of sessions), in different samples of patients treated with CT, or in studies that target disorders other than major depression. Moreover, future research may benefit from examining variables that were not assessed in the current study. For example, in addition to assessing therapist *adherence* to the techniques of CT, it will be important to measure the skill or appropriateness of the delivery of these techniques (i.e., *competence*). The results of future investigations may help clarify areas of convergence and divergence in the implementation of CT across clinical trials, and in turn inform efforts to enhance fidelity to CT.
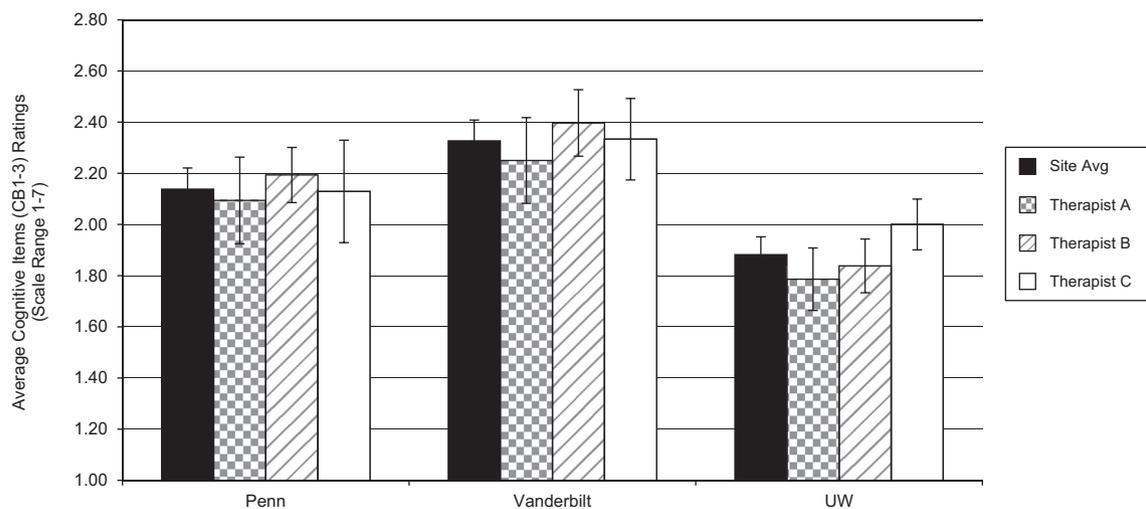

Fig. 2. Site and therapist (raw) means for cognitive items from the Collaborative Study Psychotherapy Rating Scale (CSPRS; i.e., average of CB1-CB3 subscale items). Error bars represent standard error. Penn = University of Pennsylvania (Site A from DeRubeis et al., 2005); Vanderbilt = Vanderbilt University (Site B from DeRubeis et al.); UW = University of Washington study (Dimidjian et al., 2006).
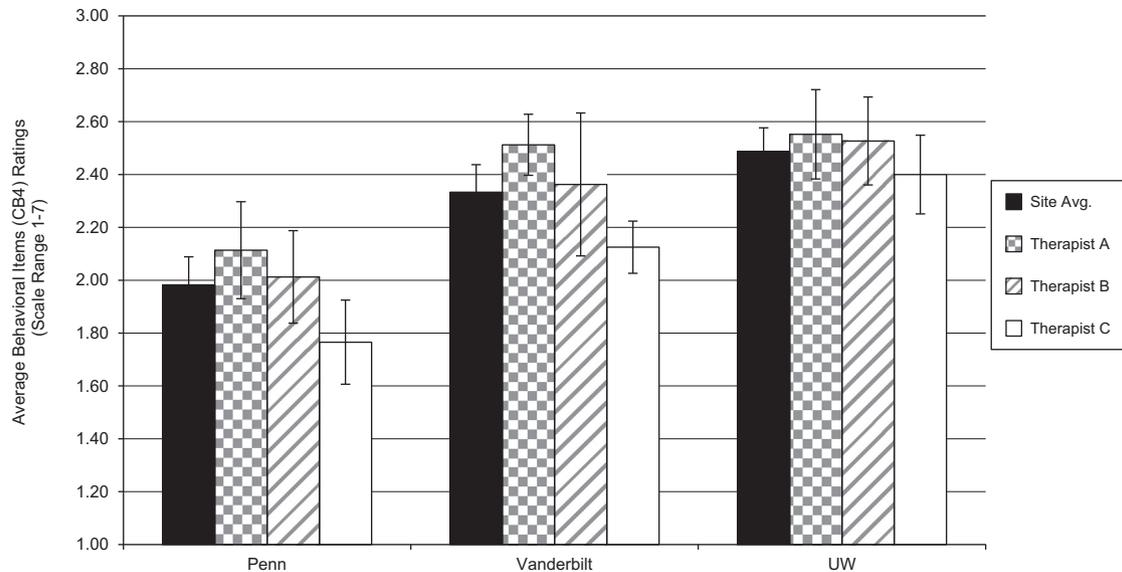
Fig. 3. Site and therapist (raw) means for behavioral focus (CB4 subscale) items from the Collaborative Study Psychotherapy Rating Scale (CSPRS). Error bars represent standard error. Penn = University of Pennsylvania (Site A from DeRubeis et al., 2005); Vanderbilt = Vanderbilt University (Site B from DeRubeis et al.); UW = University of Washington study (Dimidjian et al., 2006).

Acknowledgments

References

Barlow, D. H. (2008). *Clinical handbook of psychological disorders: A step-by-step treatment manual.* Guilford Publication.

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond.* New York: Guilford Press.

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

Beutler, L. E., Clarkin, J. E., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient.* New York: Oxford University Press.

Blagys, M. D., & Hilsenroth, M. J. (2000). Distinctive feature of short-term psychodynamic-interpersonal psychotherapy: a review of the comparative psychotherapy process literature. *Clinical Psychology: Science and Practice, 7,* 167–188.

Blagys, M. D., & Hilsenroth, M. J. (2002). Distinctive activities of cognitive-behavioral therapy: a review of the comparative psychotherapy process literature. *Clinical Psychology Review, 22,* 671–706.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice, 16,* 252–260.

DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14,* 469–482.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., et al. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62,* 409–416.

DeRubeis, R. J., Hollon, S. D., Evans, M. D., & Bemis, K. M. (1982). Can psychotherapies for depression be discriminated? A systematic investigation of cognitive therapy and interpersonal therapy. *Journal of Consulting and Clinical Psychology, 50,* 744–756.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M., et al. (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.

Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders.* Guilford Publication.

Feeley, M., DeRubeis, R. J., & Gelfand, L. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 67,* 578–582.

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement, 5,* 105–112.

Hamilton, M. A. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56–61.

Hill, C. E., O'Grady, K. E., & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in CBT, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology, 60,* 73–79.

Hollon, S. D., Evans, M. D., Auerbach, A., DeRubeis, R. J., Elkin, I., Lowery, A., et al. (1988). *Development of a system for rating therapies for depression: Differentiating cognitive therapy, interpersonal psychotherapy, and clinical management.* Unpublished manuscript. Nashville, TN: Vanderbilt University.

Hope, D. A., Heimberg, R. G., & Turk, C. L. (2010). *Managing social anxiety: A cognitive-behavioral therapy guide.* USA: Oxford University Press.

Horvath, A. O., Del Re, A. C., Fluckiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48,* 9–16.

Luborsky, L., Woody, G. E., McLellan, A. T., O'Brien, C. P., & Rosenzweig, J. (1982). Can independent judges recognize different psychotherapies? An experience with manual-guided therapies. *Journal of Consulting and Clinical Psychology, 50,* 49–62.

Malik, M. L., Beutler, L. E., Alimohamed, S., Gallagher-Thompson, D., & Thompson, L. (2003). Are all cognitive therapies alike? A comparison of cognitive and noncognitive therapy process and implications for the application of empirically supported treatments. *Journal of Consulting and Clinical Psychology, 71,* 150–158.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: issues and research recommendations. *Clinical Psychology: Science and Practice, 12,* 365–383.

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67,* 837–846.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 2,* 420–428.

Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: predictors of early inter-session symptom gains. *Behavior Research and Therapy, 48,* 599–606.

Strunk, D. R., Cooper, A. A., Ryan, E. T., DeRubeis, R. J., & Hollon, S. D. (2012). The process of change in cognitive therapy for depression when combined with antidepressant medication: predictors of early intersession symptom gains. *Journal of Consulting and Clinical Psychology, 80*(5), 730–738.

Strunk, D. R., DeRubeis, R. J., Chui, A., & Alvarez, J. A. (2007). Patients' competence in and performance of cognitive therapy skills: relation to the reduction of relapse risk following treatment for depression. *Journal of Consulting and Clinical Psychology, 75,* 523–530.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the working alliance inventory. *Psychological Assessment, 1,* 207–210.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620–630.

Watzke, B., Rueddel, H., Koch, U., Rudolph, M., & Schulz, H. (2008). Comparison of therapeutic action, style and content in cognitive-behavioral and psychodynamic group therapy under clinically representative conditions. *Clinical Psychology & Psychotherapy, 15,* 404–417.