

BRIEF REPORTS

The Temporal Relation of Adherence and Alliance to Symptom Change in Cognitive Therapy for Depression

Michael Feeley, Robert J. DeRubeis, and Lois A. Gelfand
University of Pennsylvania

This study attempted to replicate an earlier study (R. J. DeRubeis & M. Feeley, 1990) of the prediction of symptom change from process variables in cognitive therapy for depressed outpatients. Measures of in-session therapist behavior and therapist–patient interactions were correlated with prior and subsequent symptom change. One of the positive findings was confirmed, but the other received only marginal support. A “concrete” subset of theory-specified therapist actions, measured early in treatment, predicted subsequent change in depression. The therapeutic alliance was predicted by prior symptom change in 1 of the 2 later assessments, but only at a trend level. Several negative findings were similar to those obtained in the earlier study. Specifically, the alliance, an “abstract” subset of theory-specified therapist actions, and facilitative conditions did not predict subsequent change. Implications for causal inferences in psychotherapy process research are discussed.

Cognitive therapy (CT) is a well-standardized, effective treatment for major depression (cf. DeRubeis & Crits-Christoph, 1998), but the processes by which it works are still not well understood (Whisman, 1993). In a previous study, DeRubeis and Feeley (1990) examined the relations of CT-specific techniques—as well as general therapy process variables—to outcome in a sample of depressed outpatients. They found that the quantity of “concrete” methods of CT in an early (second) session correlated significantly with subsequent symptom change. Other variables, including an “abstract” CT variable (therapist-offered facilitative conditions; Hollon et al., 1988; Rogers, 1957) and the therapeutic alliance (Bordin, 1979; Morgan, Luborsky, Crits-Christoph, Curtis, & Solomon, 1982), did not predict change that occurred after the second session. However, later in treatment, the alliance was predicted by prior symptom reduction, suggesting that symptom improvement led to positive alliances rather than the other way around.

DeRubeis and Feeley’s (1990) findings were at odds with conventional wisdom, which has suggested that the alliance is an important causal agent in therapy outcome. Measures of the alliance have repeatedly been found to correlate with outcome in psychodynamic and eclectic therapies (see Horvath & Symonds,

1991, for a meta-analytic review; see also Krupnick et al., 1996), as well as in CT (e.g., Castonguay, Goldfried, Wisner, Raue, & Hayes, 1996; Gaston, Thompson, Gallagher, Cournoyer, & Gagnon, 1998; Krupnick et al., 1996; Safran & Wallner, 1991).

Three conditions must be met to substantiate a causal claim about a process variable: covariation of the process variable and symptom change, nonspuriousness, and temporal precedence of the process variable (Judd & Kenny, 1981). Therapy process researchers typically use observational, rather than experimental, methods; covariation is the principal criterion used to support or to imply causal claims. Spuriousness cannot be ruled out, but researchers often attempt to identify and control for plausible third variables.

Unfortunately, temporal precedence, which can be established even with observational methods, is often ignored. In some studies, the alliance has been measured in the midst of treatment and correlated with symptom change from the beginning to the end of treatment (e.g., Castonguay et al., 1996;¹ Gaston et al., 1998; Jones, Cumming, & Horowitz, 1988; Krupnick et al., 1996; Marziali, 1984; Safran & Wallner, 1991). Assessments of the alliance have also been averaged across the duration of treatment, and the averaged measure has been related to overall symptom change

Michael Feeley, Robert J. DeRubeis, and Lois A. Gelfand, Department of Psychology, University of Pennsylvania.

Michael Feeley is now at the Treatment Research Center, Department of Psychiatry, University of Pennsylvania.

This research was supported in part by Grant R10-MH55877 from the National Institute of Mental Health and by a University of Pennsylvania Biomedical Research Support Grant. A version of this article was included in Michael Feeley’s dissertation.

Correspondence concerning this article should be addressed to Robert J. DeRubeis, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104-1696. Electronic mail may be sent to derubeis@psych.upenn.edu.

¹ Castonguay et al.’s (1996) findings were obtained using the same patient sample as used in the present study. The differences between their findings (a significant positive relation between alliance and Beck Depression Inventory [BDI] outcome) and those reported here are not surprising in light of the many differences in methods between the two studies. Among the differences were (a) we used Session 2 for predicting outcome, whereas Castonguay et al. used Session 4, 5, 6, or 7 and (b) we used residualized subsequent BDI change as the dependent variable, whereas Castonguay et al. used termination BDI as the dependent variable. Thus, Castonguay et al. did not rule out reverse causality between alliance and outcome.

(e.g., Gaston et al., 1998; Horowitz, Marmar, Weiss, DeWitt, & Rosenbaum, 1984; Krupnick et al., 1996; Marziali, 1984; Marziali, Marmar, & Krupnick, 1981).

In the aforementioned studies, the "predicted" variable, or outcome variable, incorporated symptom change that had occurred before the alliance was assessed, confounding prior and subsequent change. When the alliance is averaged across the course of therapy, there is no way to avoid this temporal confound. However, when the alliance (or any process variable) is assessed in a given session, the temporal confound can be avoided by assessing symptom change that occurs prior to and subsequent to that session. When a process variable predicts subsequent change, it lends support to the hypothesis that the variable caused the change. However, when a process variable is predicted by symptom change, the most parsimonious explanation is that symptom change was the cause and therapy process the effect.

Besides DeRubeis and Feeley (1990), we know of only one group that has specifically attempted to rule out this temporal confound in the analysis of alliance effects. In two articles based on the same sample of depressed older adults, Gaston, Marmar, and colleagues (Gaston, Marmar, Gallagher, & Thompson, 1991; Marmar, Gaston, Gallagher, & Thompson, 1989) analyzed the relation of outcome to the alliance measured in Session 5 (as well as others) in four different treatment conditions, including CT. They controlled statistically for pretreatment depression level and for initial change in depression (from pretreatment until Session 5 in the Gaston et al. study, from pretreatment until Session 10 in the Marmar et al. study). Marmar et al. reported a significant association between one aspect of the alliance (termed *patient commitment*) and outcome, and the relationship was especially strong in the CT condition. Gaston et al. found weaker relationships and did not obtain any significant alliance effects.

The present study used the same methods as those in the DeRubeis and Feeley (1990) study, with a similar sample of depressed outpatients who received CT. Unlike the patients in the original study, however, patients in the present study were randomly assigned to standardized treatment with a controlled duration.

We hypothesized that the results of this study would resemble those obtained by DeRubeis and Feeley (1990):

1a. Concrete CT methods, assessed in Session 2, will predict subsequent symptom change.

1b. The alliance, facilitative conditions, and abstract CT methods, assessed at Session 2, will not predict subsequent change.

2a. The alliance, assessed at two time points late in treatment, will be predicted by symptom change that occurred prior to those sessions.

2b. Concrete CT methods, abstract CT methods, and facilitative conditions, assessed at two time points late in treatment, will not be predicted by symptom change that occurred prior to those sessions.

Method

Participants

Patients. Patients were drawn from the 32 adults who completed a course of CT for depression as part of the Cognitive Pharmacotherapy (CPT) study that took place at the St. Paul-Ramsey Medical Center in St. Paul, Minnesota (cf. Hollon et al., 1992); they received either CT alone

($n = 16$) or CT plus imipramine pharmacotherapy ($n = 16$). Patients were included in the CPT study only if they had a "definite" diagnosis of major depressive disorder at intake using the Research Diagnostic Criteria (Spitzer, Endicott, & Robins, 1978), a score of 20 or above on the BDI (Beck, Steer, & Garbin, 1988), and a score of 14 or above on the Hamilton Rating Scale for Depression (Hamilton, 1967).

For inclusion in the present report, patients' Session 1 BDI scores had to be 15 or greater and their Session 2 audiotape needed to be available and audible. One patient was excluded because his Session 1 BDI score was less than 15; another 6 patients were excluded because the required session audiotapes were either unavailable or inaudible. Thus, 25 patients were included in the study. The mean age in this sample was 32.9 years ($SD = 11.2$); 22 were female, 3 were male. Twelve of the 25 patients came from the CT-alone group, 13 from the CT plus imipramine group. (Because there were no differences in the pattern of findings for CT alone vs. CT plus imipramine, we report all analyses with these two groups pooled.) The length of treatment was limited to 12 weeks, with a maximum of 20 sessions during that time. The average number of sessions was 14.6.

Therapists. There were 4 cognitive therapists. Three were ACSW-level clinical social workers; 1 was a PhD-level clinical psychologist. These 4 therapists had a range of 8 to 20 years of therapy experience prior to taking part in the CPT study. Three of the therapists were male. Therapists received 6 months of intensive training and supervision in CT prior to seeing their first study patient. They also received twice-weekly supervision during the first two thirds of the study; supervision was then tapered to once per week.

Measures

Depression severity. The BDI was used as the indicator of depression severity. Patients were given the BDI at the intake interview and prior to each therapy session. Two symptom change scores were calculated for each session: prior change and subsequent change. The prior change measure for a rated session is the difference between the Session 1 BDI and the rated session BDI. The difference between the rated session BDI and the 12-week BDI is referred to as the subsequent change measure for that session. All change scores are in the form of residualized change scores. Residualized prior change scores remove the effect of the Session 1 BDI; residualized subsequent change scores remove the effect of current session BDI score. Higher residualized scores always reflect greater symptom relief.

Process measures. Adherence to therapy-specific techniques was measured by two subscales taken from the Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon et al., 1988). The CSPRS CT-Concrete subscale assesses the problem-focused, specific aspects of CT. The CSPRS CT-Abstract subscale assesses more general discussions about CT, as well as discussions of patients' general beliefs (see DeRubeis & Feeley, 1990).

The alliance was assessed with the Penn Helping Alliance rating scale (Morgan et al., 1982; see also DeRubeis & Feeley, 1990).

To measure therapist-offered facilitative conditions, we used the Facilitative Conditions subscale from the CSPRS (see DeRubeis & Feeley, 1990).

Procedure

All study sessions were audiotaped. To attempt to replicate DeRubeis and Feeley's (1990) findings regarding the second session, as well as sessions from late in therapy, we sampled three sessions from each patient-therapist dyad: Session 2 plus one randomly selected session from Quadrant 3 (Weeks 7-9) and another from Quadrant 4 (Weeks 10-12).

Raters were University of Pennsylvania undergraduate psychology majors who participated in 30 hr of training. Each of 13 raters rated an average of 18 sessions; 2 raters rated each sampled tape. Ratings were made after listening to an entire session of therapy. Raters were unaware of the

Table 1
Descriptive Statistics and Correlations With Subsequent Change in Depression for Rated Variables at Session 2

Rated variable	<i>M</i>	<i>SD</i>	<i>r</i>
CT-Concrete subscale (1–7)	2.95	0.77	.39*
CT-Abstract subscale (1–7)	2.35	0.53	.06
Facilitative Conditions subscale (1–7)	4.12	0.61	–.07
Penn Helping Alliance rating scale (0–10)	4.07	0.92	–.27

Note. $N = 25$. In all cases, correlations are positive when higher ratings on a process measure are associated with greater subsequent symptom change. CT = Cognitive Therapy.

* $p < .05$, one-tailed.

identity of patient and therapist, the session number, and the eventual outcome of each case.

Psychometric Properties of the Scales

Estimates of interrater reliability and internal consistency were calculated for each of the four scales. We estimated interrater reliability from an intraclass correlation coefficient (ICC) using the random-effects estimate, with two raters pooled (ICC[2,2]; Shrout & Fleiss, 1979). Coefficients were .75 for the CT-Concrete subscale, .60 for the CT-Abstract subscale, .57 for the Facilitative Conditions subscale, and .66 for the Penn Helping Alliance rating scale. These are somewhat lower than the ICCs reported by DeRubeis and Feeley (1990) but in the range often reported for such variables (cf. Hill, O'Grady, & Elkin, 1992). Pooled ratings were used in the data analyses.

The scales also exhibited satisfactory internal consistency; alpha coefficients ranged from .76 to .92. The scales were relatively independent of one another (mean intercorrelation = .34; range = .15 to .58; the .58 intercorrelation was between the CT-Concrete and the CT-Abstract subscales), although somewhat less independent than they were in the DeRubeis and Feeley (1990) study.

Results

More than half of the total mean change in BDI scores occurred in the first 6 weeks of treatment. The average change on the BDI between Session 1 and the 12-week session was 18.2 points. By Quadrant 3, the mean BDI score had dropped 14.8 (i.e., 81%) of the 18.2-point decrease. Following DeRubeis and Feeley (1990), we present the analyses of subsequent change for Session 2. The prediction of in-session behavior from prior symptom change is presented for Quadrants 3 and 4. In all cases, correlations are positive when higher ratings on a process measure are associated with greater symptom reduction.

Prediction of Subsequent Symptom Change

Regarding Hypotheses 1a and 1b, the CT-Concrete subscale was a significant positive predictor of subsequent change in the BDI (see Table 1).² This was not true of any of the other variables; in fact, the correlations with subsequent change of two of the variables, the Penn Helping Alliance rating scale and the Facilitative Conditions subscale, were (nonsignificantly) negative. An inspection of the standard deviations suggests that range restriction was not a special problem with the scales that failed to evidence a positive predictive effect.

To investigate the possibility that the predictive relation found for the CT-Concrete subscale in the early session could be explained by the other scales, we conducted a regression analysis with all four rating scales as predictors, and subsequent change, again, as the criterion. In this analysis, the effect of each predictor was examined only after all the other predictors had been controlled for. The test of the CT-Concrete subscale yielded a significant positive effect, $\beta = 0.78$, $t(20) = 3.32$, $p < .005$. No other variable yielded a positive effect, and none of the effects were significant (all β s < 0.00 , all t s < 0.00 , all p s $> .05$).

Prediction of Process Variables From Prior Symptom Change

Regarding Hypothesis 2a, there was a nonsignificant trend for prior symptom change to predict greater Quadrant 3 Penn Helping Alliance rating scale scores (see Table 2). The correlation between prior change and the Penn Helping Alliance rating scale in Quadrant 4 was positive, but it was small and not significant. Regarding Hypothesis 2b, correlations between prior change and the other variables (the CT-Concrete subscale, the CT-Abstract subscale, and the Facilitative Conditions subscale) were small or negative and none were significant.³

Discussion

The DeRubeis and Feeley (1990) finding—that CT-Concrete subscale scores predicted symptom change—was replicated in the present study. The more the therapist delivered this subset of

² Because we conducted only one test of Hypothesis 1a (that the CT-Concrete subscale would predict subsequent change), a Bonferroni correction was not deemed appropriate. Tests of the predictive value of the Penn Helping Alliance rating scale, the Facilitative Conditions subscale, and the CT-Abstract subscale were conducted in case they might disconfirm DeRubeis and Feeley's (1990) null findings with these variables (Hypothesis 1b). None of these associations were significant, even without correction for multiple tests.

³ Because we conducted two tests of Hypothesis 2a (that prior change would predict helping alliance in the later sessions), we used a Bonferroni correction for tests of the relation of prior change to assessment of Quadrants 3 and 4 Penn Helping Alliance rating scale scores. Pertinent to Hypothesis 2b, none of the other variables assessed in Quadrant 3 or 4 were significantly correlated with prior change, even without correction for multiple tests.

Table 2
*Descriptive Statistics and Correlations With Prior Change in Depression for
 Rated Variables From Quadrants 3 and 4*

Rated variable	Quadrant 3 (n = 24)			Quadrant 4 (n = 19)		
	M	SD	r	M	SD	r
CT-Concrete subscale (1-7)	2.48	0.85	.14	2.16	0.67	-.22
CT-Abstract subscale (1-7)	2.31	0.73	.04	2.50	0.97	-.11
Facilitative Conditions subscale (1-7)	3.88	0.78	.04	4.15	0.86	.01
Penn Helping Alliance rating scale (0-10)	4.25	1.82	.39*	4.82	1.75	.15

Note. In all cases, correlations are positive when higher ratings on a process measure are associated with greater prior symptom change. CT = Cognitive Therapy.

* $p < .06$, one-tailed, with Bonferroni correction for two tests of Hypothesis 2a.

theory-specified actions early in therapy, the greater the symptom relief experienced subsequently by the patient. The CT-Abstract subscale, despite its correlation with the CT-Concrete subscale, bore no relation to symptom change. These findings suggest that the problem-focused aspects of CT play a role in reducing depressive symptoms.

By measuring therapist behaviors and depression symptoms at several time points during therapy, and by distinguishing prior symptom change from subsequent symptom change, we were able to show that the CT-Concrete subscale predicted change in depression, rather than vice versa. This temporal pattern is consistent with the cause-effect hypothesis of interest. By measuring two widely regarded alternative therapy process variables (the Penn Helping Alliance rating scale and the Facilitative Conditions subscale) and analyzing their effects together with the CT-Concrete subscale, we were able to show that the CT-Concrete subscale alone predicted change in depression. Moreover, its association with these two other variables did not account for the result.

Although the methods used both in the DeRubeis and Feeley (1990) study and in the present study can rule out a temporal confound, they cannot rule out a third-variable confound. In particular, it cannot be known from these studies whether there are patient characteristics that lead both to therapists' delivery of high levels of concrete CT and to favorable outcome.

The DeRubeis and Feeley (1990) finding that greater helping alliance in Quadrant 3 was predicted by prior change was not replicated; the result was at the level of a nonsignificant trend. Their finding that Quadrant 4 helping alliance was predicted by prior change was not replicated. As in the DeRubeis and Feeley study, early session helping alliance did not significantly predict subsequent symptom change. Overall, there was a weaker relation between the helping alliance and outcome in this study compared with the earlier study. To the extent that a relation existed, the helping alliance behaved more like a product of, rather than a cause of, symptom change.

Variability in facilitative conditions bore no direct relation to prior or subsequent outcome. It may be that there is a threshold level of facilitative conditions that is necessary to promote clinical benefit and that that level was exceeded by all therapists in this study. It may also be that an optimal level of facilitative conditions reduces dropout rates, but we did not study this possibility. It should be noted that the poorest interrater reliability was obtained on the Facilitative Conditions subscale. Improved reliability on

this, as well as the other variables, would allow for more precise estimates of the relation of these variables to outcome.

In summary, the results of this study lend support to the hypothesis that theory-specific techniques, delivered early in treatment, are important contributors to change in CT. These results also highlight the possibility that, to the extent that there is a relation between the alliance and outcome in CT, it is due to the effect of symptom change on the alliance, not the effect of the alliance on outcome. We encourage researchers to continue to investigate the role of technical and relationship aspects of CT and other psychotherapies and to collect and analyze their data in a way that will allow an examination of the temporal relations of process variables and symptom change measures. Given that the relation of CT technique to outcome has been found using the second session of therapy, we advise that very early sessions be included in the samples used in process research. This type of research has the promise of informing and improving therapeutic procedures.

References

- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77-100.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice, 16*, 252-260.
- Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology, 64*, 497-504.
- DeRubeis, R. J., & Crits-Christoph, P. (1998). Empirically supported individual and group psychological treatments for adult mental disorders. *Journal of Consulting and Clinical Psychology, 66*, 37-52.
- DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14*, 469-482.
- Gaston, L., Marmar, C., Gallagher, D., & Thompson, L. (1991). Alliance prediction of outcome beyond in-treatment symptomatic change as psychotherapy processes. *Psychotherapy Research, 1*, 104-112.
- Gaston, L., Thompson, L., Gallagher, D., Cournoyer, L.-G., & Gagnon, R. (1998). Alliance, technique, and their interactions in predicting outcome of behavioral, cognitive, and brief dynamic therapy. *Psychotherapy Research, 8*, 190-209.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6*, 276-296.

- Hill, C. E., O'Grady, K. E., & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology, 60*, 73-79.
- Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive therapy, pharmacotherapy and combined cognitive-pharmacotherapy in the treatment of depression. *Archives of General Psychiatry, 49*, 774-781.
- Hollon, S. D., Evans, M. D., Auerbach, A., DeRubeis, R. J., Elkin, I., Lowery, A., Kriss, M. R., Grove, W. M., Tuason, V. B., & Piasecki, J. M. (1988). *Development of a system for rating therapies for depression: Differentiating cognitive therapy, interpersonal psychotherapy and clinical management pharmacotherapy*. Unpublished manuscript, Vanderbilt University, Nashville, TN.
- Horowitz, M. J., Marmar, C., Weiss, D. S., DeWitt, H. N., & Rosenbaum, R. (1984). Brief psychotherapy of bereavement reactions. *Archives of General Psychiatry, 41*, 438-448.
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*, 139-149.
- Jones, E. E., Cumming, J. D., & Horowitz, M. J. (1988). Another look at the nonspecific hypothesis of therapeutic effectiveness. *Journal of Consulting and Clinical Psychology, 56*, 48-55.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.
- Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*, 532-539.
- Marmar, C. R., Gaston, L., Gallagher, D., & Thompson, L. W. (1989). Alliance and outcome in late-life depression. *The Journal of Nervous and Mental Disease, 177*, 464-472.
- Marziali, E. (1984). Three viewpoints on the therapeutic alliance. *The Journal of Nervous and Mental Disease, 172*, 417-423.
- Marziali, E., Marmar, C., & Krupnick, J. (1981). Therapeutic alliance scales: Development and relationship to psychotherapy outcome. *American Journal of Psychiatry, 138*, 361-364.
- Morgan, R., Luborsky, L., Crits-Christoph, P., Curtis, H., & Solomon, J. (1982). Predicting the outcomes of psychotherapy by the Penn Helping Alliance rating method. *Archives of General Psychiatry, 39*, 397-402.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95-103.
- Safran, J. D., & Wallner, L. K. (1991). The relative predictive validity of two therapeutic alliance measures in cognitive therapy. *Psychological Assessment, 3*, 188-195.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
- Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria: Rationale and reliability. *Archives of General Psychiatry, 35*, 773-782.
- Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*, 248-265.

Received July 14, 1998
Accepted November 3, 1998 ■