Between-Therapist and Within-Therapist Differences in the Quality of the Therapeutic Relationship: Effects on Maladjustment and Self-Critical Perfectionism

David C. Zuroff, Allison C. Kelly, and Michelle J. Leybman
McGill University

Sidney J. Blatt
Yale University

Bruce E. Wampold
University of Wisconsin – Madison

The relationship between therapeutic outcome and a patient-reported measure of the Rogerian conditions of positive regard, empathy, and genuineness was decomposed into between-therapist effects and within-therapist effects using multilevel modeling. Data were available for 157 depressed outpatients treated by 27 therapists in the cognitive-behavioral therapy, interpersonal therapy, or placebo with clinical management conditions of the Treatment of Depression

We thank Scott Baldwin for his advice on conducting the multilevel modeling in SAS.

The NIMH-TDCRP was a multisite program initiated and sponsored by the Psychosocial Treatments Research Branch, Division of Extramural Research Programs, NIMH. The program was funded by cooperative agreements to six participating sites: George Washington University, MH 33762; University of Pittsburgh, MH 33753; University of Oklahoma, MH 33760; Yale University, MH 33827; Clarke Institute of Psychiatry, MH 38231; and Rush Presbyterian-St. Luke’s Medical Center, MH 35017. The principal NIMH collaborators were Irene Elkin, Coordinator; Tracie Shea, Associate Coordinator; John P. Docherty; and Morris B. Parloff. The principal investigators and project coordinators at the three participating research sites were, at George Washington University, Stuart M. Sotsky and David Glass; at the University of Pittsburgh, Stanley D. Imber and Paul A. Pilkonis; and at the University of Oklahoma, John T. Watkins and William Leber. The principal investigators and project coordinators at the three research sites responsible for training therapists were, at Yale University, Myrna Weissman (now at Columbia University), Eve Chevron, and Bruce J. Rounsaville; at Clarke Institute of Psychiatry, Brian F. Shaw and T. Michael Vallis; and at Rush Presbyterian-St. Luke’s Medical Center, Jan A. Fawcett and Phillip Epstein. Collaborators in the data management and data analysis aspects of the program were C. James Klett, Joseph F. Collins, and Roderic Gillis of the Veterans Administration Studies Program, Perry Point, Maryland.

Correspondence concerning this article should be addressed to: David C. Zuroff, McGill University, Department of Psychology, 1205 Dr. Penfield Ave., Montreal, Quebec, Canada H3A 1B1; e-mail: zuroff@ego.psych.mcgill.ca

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jclp.20683
A growing body of evidence suggests that psychotherapists differ substantially in their average effectiveness; indeed, variability in outcomes attributable to differences between therapists appears to be considerably larger than variability attributable to differences between treatments (Wampold, 2001). In an important early review, Crits-Christoph and Mintz (1991) examined 10 clinical trials that included 27 treatment conditions for a variety of disorders, concluding that on average about 8% of the variance in outcomes was attributable to differences between therapists. Two subsequent studies examined data from randomized clinical trials of treatments for depression (Blatt, Sanislow, Zuroff, & Pilkonis, 1996) and panic disorder (Huppert et al., 2001). Although both studies relied on analysis of variance, in which therapists were treated as a fixed effect, suggestive evidence was obtained that therapists differed in their average effectiveness.

Recent studies have employed more sophisticated multilevel modeling strategies, which treat therapists as a random rather than a fixed effect. Kim, Wampold, and Bolt (2006) reanalyzed data from the cognitive-behavior therapy (CBT) and interpersonal therapy (IPT) conditions in the Treatment of Depression Collaborative Research Program (TDCRP; Elkin, 1994), reporting that the proportion of variance in outcome within the completer sample, attributable to differences between therapists, ranged from about 8% to 12% for different measures of outcome. In a similar set of analyses using pill placebo with clinical management (PLA-CM) and imipramine with clinical management (IMI-CM) conditions in the TDCRP, McKay, Imel, and Wampold (2006) found that the proportion of variance in outcome attributable to differences between psychiatrists was 6.7%, using the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967), and 9.1%, using the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, & Erbaugh, 1961).

Recent studies of psychotherapy in naturalistic settings have uncovered therapist effects similar to those observed in controlled trials. Using multilevel modeling to examine rates of improvements (slopes) in a university counseling center, Okiishi, Lambert, Nielsen, and Ogles (2003) found that there was significant variability across therapists in their clients’ average rate of improvement. This finding was replicated in a larger sample of students and therapists at the same counseling center (Okiishi et al., 2006). In a study of outcome in a wide variety of disorders treated in a managed care context, Wampold and Brown (2005) found that about 5% of the variance in outcomes was attributable to differences between therapists. Lutz, Leon, Martinovich, Lyons, and Stiles (2007) also studied a managed care sample, finding
that 17% of the variance in clients’ rates of change could be attributed to differences between therapists. Dinger, Strack, Leichsenring, Wilmers, and Schauenberg (2008) found that outcome variance attributable to therapists in a naturalistic study of inpatient treatment was significant (3%), but smaller than that reported in the outpatient studies. Despite these consistent therapist effects in studies of diverse disorders, in numerous forms of treatment, and using multiple outcome measures, researchers have remained more interested in putative differences between treatments (e.g., CBT vs. Behavioral Activation) than differences between treatment providers (Blatt & Zuroff, 2005; Kim et al., 2006; Okiishi et al., 2003).

A number of possible mechanisms may account for the differences between therapists that lead to differential outcomes. Given the significant relation between dimensions of the therapeutic relationship and outcome (Norcross, 2002), one possibility is that therapists differ in their capacity for forming constructive psychotherapeutic relationships, and these differences in relational capacity give rise to differences in outcome. Examining this hypothesis requires that the overall effect of a relationship variable (e.g., the therapeutic alliance) on outcome be decomposed into the component attributable to average differences between therapists and the component attributable to differences between patients within a given therapist’s caseload (Baldwin, Wampold, & Imel, 2007). Conceptually, the first (between-therapists) component reflects the degree to which the quality of the therapeutic relationship is attributable to characteristics of the therapist. The second (within-therapists) component reflects both the patient’s general capacity to form a therapeutic relationship and the specific or interactive effect on the relationship of the pairing of the patient with a particular therapist. This decomposition into the therapist’s and patient’s contribution to the therapeutic alliance is effected by conducting multilevel analyses in which patients are modeled as nested within therapists, and the within-therapist (level-1) effects are tested separately from the between-therapist (level-2) effects.

Baldwin et al. (2007) carried out such analyses using data from the Research Consortium of Counseling and Psychological Services in Higher Education study (Brownson, 2004). Pretreatment and posttreatment scores on the Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004) were available for 331 patients seen by 80 therapists for an average of about seven sessions. The strength of the therapeutic alliance was assessed using the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989) administered prior to the fourth therapy session. Baldwin et al.’s primary finding was that therapist variability in the alliance was related to outcome; that is, “therapists whose patients, on average, rated their alliance high, also had better outcomes than therapists whose patients, on average, rated their alliance low” (p. 846). Patient variability in the alliance was not significantly related to outcome, suggesting that the causal pathways from alliance to outcome primarily reflect the therapist’s capacity for establishing a constructive relationship with a range of patients rather than the patient’s capacity to establish or respond to a constructive relationship. In summary, Baldwin et al.’s results provide a new line of evidence for the importance of therapists in determining the outcomes achieved by their patients.

The theoretical import of Baldwin et al.’s (2007) findings warrants efforts to assess their replicability and generalizability. Their analyses were conducted with data from a naturalistic study of a heterogeneous population of college counseling clients who received a heterogeneous array of (mostly) brief interventions. One wonders whether similar results would be obtained with samples of psychiatric patients with specific disorders receiving specific, manualized treatments in the context of a clinical trial.
The generalizability of Baldwin et al.’s findings over different dimensions of the therapeutic relationship and different measures of outcome also remains unknown. The TDCRP data set is an ideal context for investigating these questions, as it involved the random assignment of carefully diagnosed depressed outpatients to receive one of several well-defined treatments, and included a wealth of measures of process and outcome. Thus, we undertook to replicate and extend Baldwin et al.’s results using publicly available data from the TDCRP.

Several previous analyses of TDCRP data are relevant to the present article. Zuroff and Blatt (2006) conducted a stringent test of the relations to outcome of three dimensions of the therapeutic relationship: the therapist contribution to the alliance, the patient contribution to the alliance, and the aggregate of patient-perceived empathy, positive regard, and congruence in the therapist. Empathy, positive regard, and congruence are the qualities that Carl Rogers (1951, 1957) considered to be the necessary and sufficient conditions for therapeutic change, and are frequently referred to as the Rogerian conditions. Patient and therapist contributions to the therapeutic alliance were assessed using Krupnick et al.’s (1996) observer ratings of video recordings of the third treatment session; patients’ perceptions of the Rogerian conditions were assessed using the Barrett-Lennard Relationship Inventory (B-L RI; Barrett-Lennard, 1962) administered after the second treatment session. Outcome was assessed using both patient-reported and clinician-rated measures. Early treatment response (from intake to the fourth treatment session) was partialled from the relationship measures, and the resulting residualized measures were used to predict subsequent change in outcome measures. Multilevel modeling consistently demonstrated effects of the residualized B-L RI on outcome, demonstrating that the impact of perceived Rogerian conditions on outcome was independent of the degree of early therapeutic change. The most robust effects were obtained with a composite measure of outcome that combined five measures of patient-reported and clinician-rated symptoms and overall life adjustment (Blatt, Zuroff, Bondi, Sanislow, & Pilkonis, 1998). The analyses reported here, therefore, focused on the B-L RI as a predictor of this composite measure of outcome.

The Rogerian conditions and the therapeutic alliance are conceptually distinct dimensions of the therapeutic relationship (Hatcher & Barends, 2006; Horvath & Bedi, 2002), as the former captures the patient’s perceptions of a facet of the therapist’s behavior and the latter addresses the active collaboration between the therapist and patient. Measures of perceived Rogerian conditions and the working alliance are only weakly correlated (Zuroff & Blatt, 2006). Consequently, the present analyses are not a strict replication of Baldwin et al.’s analyses, but rather an exploration of another facet of the therapeutic relationship.

A comprehensive assessment of outcome should include measures of reduction in underlying vulnerabilities as well as symptom and adjustment-focused measures (Blatt & Zuroff, 2005; Blatt, Zuroff, Hawley & Auerbach, 2010). Fortunately, the architects of the TDCRP included the Dysfunctional Attitudes Scale (DAS; Weissman & Beck, 1978) at each assessment point during treatment. The DAS was designed to measure cognitive vulnerability to depression. Research with the Perfectionism subscale of the DAS and related measures such as the Self-Criticism scale of the Depressive Experiences Questionnaire (DEQ; Blatt, D’Afflitti, & Quinlan, 1976), has shown that self-critical perfectionism (SC-PFT) both confers vulnerability to depression (Blatt, 2004; Zuroff, Mongrain, & Santor, 2004) and interferes with the treatment of depression (Blatt & Zuroff, 2005). Blatt and Zuroff
(2005) found that, across the four treatment conditions of the TDCRP, higher levels of the Rogerian conditions were associated with greater reduction in SC-PFT, but they did not distinguish the effects of therapist variability and patient variability in the B-L RI. We, therefore, extended Blatt and Zuroff’s and Baldwin et al.’s (2007) investigations by examining the contributions of between-therapist and within-therapist variability in perceived Rogerian conditions to the reduction of SC-PFT.

Because of differences between the TDCRP and Research Consortium datasets, our statistical procedures differed from those of Baldwin et al. (2007). Outcome data were available in the TDCRP at five points from the beginning to the end of treatment, whereas the Research Consortium collected only pretreatment and posttreatment data. Baldwin et al., therefore, focused on prediction of posttreatment scores, adjusted for pretreatment scores, by between-therapist and within-therapist alliance scores. We instead modelled rates of change over time (slopes) in the dependent variables (cf. Lutz et al., 2007; Zuroff & Blatt, 2006), and examined the degree to which these slopes were determined by the between-therapist and within-therapist B-L RI scores.

We included the two psychotherapy conditions in the TDCRP (CBT and IPT) as well as the PLA-CM condition. The PLA-CM condition was included because clinical management was designed to provide the basic elements of supportive psychotherapy (Elkin, Parloff, Hadley, & Autry, 1985). Additionally, previous analyses of variables related to the therapeutic relationship found few differences between the formal psychotherapies and the CM conditions (e.g., Blatt, Zuroff, Quinlan, & Pilkonis, 1996; Zuroff et al., 2000). Patients in the IMI-CM condition were omitted from the analyses for two reasons. First, the administration of imipramine introduced a treatment element that clearly could not be construed as psychotherapy. Second, the fact that each psychiatrist treated patients in both the IMI-CM and PLA-CM conditions would have made statistical modeling very difficult, as treatment providers were nested within the CBT and IPT conditions but crossed the two CM conditions.

Based on Baldwin et al.’s (2007) findings, we hypothesized that more rapid reductions in both the composite maladjustment measure and SC-PFT would be achieved by patients whose therapists were perceived to offer, on average, higher levels of the Rogerian conditions. We thought it was possible that the more sensitive tests afforded by the repeated measurements in the TDCRP would reveal additional within-therapists effects of the B-L RI, that is, that patients who experienced higher levels of the Rogerian conditions than their therapist’s average patient would report more rapid improvement.

Method

Participants in the TDCRP were outpatients with nonbipolar, nonpsychotic, major depressive disorders. Two hundred and fifty patients were randomly assigned to treatments, 239 patients began treatment, and 162 were defined as completers, having received at least 12 treatment sessions over at least a 15-week period. Inclusion and exclusion criteria, sample characteristics, treatment procedures, and assessment procedures have been described previously (Elkin et al., 1985; Elkin et al., 1989; Shea et al., 1992). Patients scored 14 or higher on the 17-item version of the HRSD and met Research Diagnostic Criteria for a current episode of definite major depression. Patients in the IMI-CM condition were excluded from the present analyses, as were patients with missing data for the B-L RI at session two. One
therapist in the CBT condition saw only one patient, and that dyad was excluded as well. These criteria resulted in a sample of 157 (43 male, 114 female) patients, with 53 receiving IPT, 54 receiving CBT, and 50 receiving PLA-CM. They had an average age of 35.1 (SD = 8.60). Their racial/ethnic backgrounds were 86.6% white, 10.2% black, 2.6% Hispanic, and 0.6% other. Their religions were 33.8% Catholic, 40.8% Protestant, 5.7% Jewish, 1.3% Moslem, 3.2% other, and 15.3% unaffiliated. Their mean occupational status was 4.60 (SD = 1.25), on a 0–7 scale anchored by never employed and executive/professional.

These 157 patients were treated by 27 MD-level or PhD-level therapists who saw between 3 and 11 patients each. Twenty of the 27 therapists were male, and 18 of the 27 were MDs. IPT was provided by eight therapists; CBT was provided by nine therapists; and PLA-CM was provided by 10 therapists. The average age of the therapists was 41.51 (SD = 8.23) years, and their average years of experience was 11.33 (SD = 7.09).

All therapists received training and met competency criteria in the treatment they provided. The treatments were manualized and monitored by experts. Tapes of sessions were reviewed to assure adherence to treatment protocols, and therapists received consultation during the study (Elkin, 1994). The manuals for clinical management in the medication conditions provided “guidelines for providing support and encouragement to the patient and giving direct advice when necessary. This CM component thus approximates a “minimal supportive therapy” condition” (Elkin et al., 1985, p. 311).

Evaluations were conducted at intake (0 weeks), 4 weeks, 8 weeks, 12 weeks, and 16 weeks. At each assessment, patients completed self-report measures and were interviewed by a PhD-level clinical evaluator (CE).

**Measures**

*Maladjustment*. The principal outcome measures in the TDCRP were as follows: the BDI, the HRSD, the Hopkins Symptom Checklist (HSCL-90; Derogatis, Lipman, & Covi, 1973), the Global Adjustment Scale (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976), and the Social Adjustment Scale (SAS; Weissman & Paykel, 1974). Despite varying in both item content and rater (self vs. CE), residual change scores on the five outcome measures loaded on a single factor that accounted for 75.5% of the variance (Blatt, Zuroff et al., 1996). We used a Composite Maladjustment Index (CMI; Blatt et al., 1998), that was based on z-scores for each of the five variables calculated using the pooled mean over the five observation points and the pooled within-time period standard deviation. The CMI was computed for each observation point by taking the mean of the five z-scores with appropriate reversals. The CMI has been used in several reanalyses of the TDCRP (Shahar, Blatt, Zuroff, & Pilkonis, 2003; Shahar, Blatt, Zuroff, Krupnick, & Sotsky, 2004; Zuroff et al., 2000; Zuroff & Blatt, 2006). Because of the higher reliability afforded by aggregating correlated measures, more sensitive tests of effects are possible using this composite. Moreover, parameter estimates are likely to be more precise as a result of reduced random error.

*Self-critical perfectionism*. The DAS (Weisssman & Beck, 1978) was designed to measure cognitive vulnerability to depression. Consistent with prior research (e.g., Cane, Olinger, Gotlib, & Kuiper, 1986), Perfectionism and Need for Approval dimensions emerged in a principal components analysis with a varimax rotation of TDCRP intake data (Imber et al., 1990). Eleven items loaded substantially (> .40)
on Need for Approval, and 15 items loaded substantially (> .40) on Perfectionism. Summing the items with high loadings yielded composites with high internal consistency, \( \alpha = .91 \) for Need for Approval and .82 for Perfectionism. Because perfectionism has been found to be related to outcome at the end of treatment and to functioning during the follow-up period (Blatt et al., 1998; Zuroff & Blatt, 2002), we focused on Perfectionism rather than Need for Approval. The two highest loading items for Perfectionism were as follows: “If I do not do as well as other people, it means that I am an inferior human being” and “If I fail at my work, then I am a failure as a person.” In the present article, we refer to the DAS Perfectionism scale as self-critical perfectionism (SC-PFT) rather than simply perfectionism. Recent research has demonstrated that perfectionism is a multidimensional construct, and that the DAS-SC scale loads on the self-critical perfectionism factor, the more maladaptive of the two major factors of perfectionism (Dunkley & Kyparissis, 2008; Powers, Zuroff, & Topciu, 2004). The mean score for SC-PFT was 49.34 (SD = 17.22). In comparison, Dunkley and Kyparissis’ (2008) community sample of employed adults reported a mean score of 38.81 (SD = 15.15).

Perceived Rogerian conditions in the therapeutic relationship. The Barrett-Lennard Relationship Inventory (Barrett-Lennard, 1962) includes four subscales that assess the patient’s perception of the extent to which the therapist provides positive regard, empathy, unconditional regard, and congruence. The B-L RI has been shown to predict outcome in psychotherapy and counselling (Barrett-Lennard, 1986; Gurman, 1977). We used the B-L RI from the second treatment session.

Blatt, Sanislow et al. (1996) found that the four scales of the B-L RI formed a single factor and constructed a measure of the perceived Rogerian conditions by summing scores on the three high-loading scales, Empathy, Level of Regard, and Congruence. Each subscale comprises 16 items rated on a 6-point scale. Cronbach’s \( \alpha \) for the composite of 48 items was .95. Examples of items include: “… nearly always knows what I mean” (Empathy), “… feels a true liking for me” (Level of Regard), and “… is comfortable and at ease in our relationship” (Congruence). The mean B-L RI score was 20.78 (SD = 11.66).

The mean score over the patients in each therapist’s caseload was designated the Between-Therapist B-L RI (BT-BLRI). The SD of the BT-BLRI was 5.73, with a range from –14.19 to 10.59. Each patient’s B-L RI was centered by subtracting from it the mean score for that patient’s therapist. The resulting centered score was designated the Within-Therapist B-L RI (WT-BLRI). To simplify calculations of simple slopes in the multilevel models, the BT-BLRI scores were centered over the 27 therapists. WT-BLRI was negatively related to SC-PFT at intake, \( r(155) = .23 \), \( p < .01 \).

Results

Our principal data analytic strategy was multilevel modeling, using PROC MIXED, version 9.1 (SAS Institute, 2004) and maximum likelihood estimation. The dependent variables were the CMI and SC-PFT, each measured at five time points (0 weeks, 4 weeks, 8 weeks, 12 weeks, and 16 weeks). CMI scores were multiplied by 100 for ease of presentation. Descriptive statistics for the CMI and SC-PFT are presented in Table 1. The effect size \( r \) was calculated for each significant effect using the formula: \( r = [F(F+df)]^{1/2} \) (Rosnow & Rosenthal, 1996). Cohen (1988) characterized \( r = .10 \) as a “small” effect and \( r = .30 \) as a “medium” effect.
The amount of between-therapist and within-therapist variance in B-L RI scores was estimated using a multilevel model with a random intercept for therapist and no fixed effect predictors (Snijders & Bosker, 1999). The relative size of between-therapist and within-therapist variance was assessed using the intraclass correlation (ICC), which is defined as the ratio of between-therapist variance to the sum of the between-therapist and within-therapist variance. The ICC for the B-L RI was .098, indicating that about 10% of the variance in patients’ perceptions of the Rogerian conditions could be accounted for by differences between the therapists. Thus, the majority of the variance was within-therapists, but between-therapists variability was present as well.

Predictive models for changes in the CMI and SC-PFT were developed using a sequential strategy (Wallace & Green, 2002), beginning by evaluating the fixed effects part of the model while assuming a simple structure for the random effects part. The initial models included a random intercept for patients and allowed the error variance at each time point to vary (a UN(1) structure in PROC MIXED notation). In analyses for the CMI, time was log-transformed (i.e., time = log [1 + weeks]) to achieve a nearly linear relationship with time. The relationship between SC-PFT and time (i.e., week of treatment) was close to linear, so time was untransformed in those analyses. In both analyses, a negative slope for time indicated change in the desired direction, decreased maladjustment or decreased SC-PFT. The fixed effects in the initial models were as follows: treatment condition, BT-BLRI, WT-BLRI, and the interactions of these three predictors with time. Interactions with time were of particular interest, because they would indicate that the rate of improvement was predicted by treatment condition or aspects of the therapeutic relationship. The Condition × Time effect was nonsignificant for both the CMI and SC-PFT. We, therefore, deleted the effects involving Condition. We also tested for the presence of cross-level interactions, that is, BT-BLRI × WT-BLRI × Time effects. In no case was this interaction significant, so all models described below omit the cross-level interaction.

Our second step was to explore alternative error covariance structures. It is important to find the most appropriate error structure because the fixed and random components of mixed models are linked; misspecification of one can lead to inaccurate estimates in the other (Wallace & Green, 2002). In particular, we tested whether the model fit was improved according to the AIC criterion when: (a) a random intercept for therapist was added, (b) a random slope for time was added, or (c) the UN(1) error covariance structure was replaced by either an autoregressive

<table>
<thead>
<tr>
<th>Time</th>
<th>CMI M</th>
<th>CMI SD</th>
<th>CMI N</th>
<th>SC-PFT M</th>
<th>SC-PFT SD</th>
<th>SC-PFT N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 weeks</td>
<td>113.20</td>
<td>58.74</td>
<td>157</td>
<td>49.50</td>
<td>17.26</td>
<td>157</td>
</tr>
<tr>
<td>4 weeks</td>
<td>33.44</td>
<td>87.57</td>
<td>144</td>
<td>46.86</td>
<td>17.52</td>
<td>143</td>
</tr>
<tr>
<td>8 weeks</td>
<td>-11.18</td>
<td>98.34</td>
<td>129</td>
<td>43.32</td>
<td>17.19</td>
<td>126</td>
</tr>
<tr>
<td>12 weeks</td>
<td>-25.67</td>
<td>96.10</td>
<td>114</td>
<td>43.87</td>
<td>17.41</td>
<td>114</td>
</tr>
<tr>
<td>16 weeks</td>
<td>-67.64</td>
<td>94.75</td>
<td>113</td>
<td>39.94</td>
<td>16.59</td>
<td>112</td>
</tr>
</tbody>
</table>

Note: CMI = composite maladjustment index; SC-PFT = self-critical perfectionism.
(AR[1]) or autoregressive with heterogeneous error variance (ARH[1]) structures. BT-BLRI and WT-BLRI were left in raw score units in these analyses; consequently, the reported slopes indicate the effect of a difference of one unit in raw scores on the BLRI. We also explored whether the effect of patient variability in Rogerian conditions varied depending on the therapist; that is, we added a random effect for therapist for the WT-BLRI × Time interaction. No evidence was found for significant variance in that parameter, so this random effect was omitted from all models.

**Composite Maladjustment Index**

The final model for the CMI included a random slope for log-transformed time in addition to the random intercept for patients and the UN(1) error covariance structure. Significant effects were found for both the BT-BLRI × Time ($F(1, 497) = 9.79, p < .01$, effect size $r = .281$) and WT-BLRI × Time interactions ($F(1, 497) = 4.73, p < .05$, effect size $r = .195$). The regression parameters for the interaction effects were, respectively, $-1.56$ ($SE = .50$) and $-0.65$ ($SE = .30$). Although the parameter for the between-therapist effect was twice as large as the parameter for the within-therapist effect, that difference only approached significance, $t(497) = -1.60, p < .12$. To interpret the interactions, we calculated slopes of the CMI as a function of time (i.e., the rates at which the CMI decreased) for patients whose therapists fell above or below the mean (+1 or −1) for BT-BLRI and for patients whose WT-BLRI fell above or below (+1 or −1) the mean for their therapist. As expected, patients whose therapists were characterized by high mean scores on the B-L RI improved more quickly (slope $= -55.68$, $SE = 2.87$, $p < .001$) than patients whose therapists were characterized by low mean scores on the B-L RI (slope $= -52.56$, $SE = 2.94$, $p < .001$). As well, patients whose B-L RI scores were higher than the average of the other patients for their therapist improved more rapidly (slope $= -54.77$, $SE = 2.86$, $p < .001$) than did patients with B-L RI scores that were below the average for other patients of their therapist (slope $= -53.46$, $SE = 2.89$, $p < .001$).

Results expressed in terms of the effects of B-L RI units on slopes per unit of log-time are difficult to understand and distant from clinical realities. To illustrate the impact of differences in Rogerian conditions more directly, we calculated estimated values for the CMI at each time point for patients with above average (+10) or below average (−10) therapists. We focused on differences of 10 because it is close to the SD of the raw B-L RI in this population and also represents nearly the full range of BT-BLRI scores. It can be seen from Figure 1 that high scores on the BT-BLRI and, to a lesser extent, the WT-BLRI, predicted more rapid declines in the CMI. The difference in the decrease in predicted CMI scores from intake to termination for the patients of therapists with low versus high average levels of perceived Rogerian conditions was 103.4 units on the CMI, that is, a full standard deviation of the CMI at the beginning of therapy.

**Self-Critical Perfectionism**

The final model for SC-PFT replaced the UN(1) error covariance structure with an ARH(1) structure. The autoregressive correlation was .42, $p < .001$. A significant effect was found for BT-BLRI × Time ($F(1, 492) = 7.60, p < .01$, effect size $r = .249$), but not for WT-BLRI × Time ($p > .50$). The regression parameters for the interaction effects were, respectively, $-0.035$ ($SE = .013$) and $-0.004$ ($SE = .008$), and differed...
significantly, $t(492) = -2.12, p < .05$. As expected, patients whose therapists were characterized by high mean scores on the B-L RI showed a more rapid reduction in vulnerability (slope $= -.662, SE = .072, p < .001$) than patients whose therapists were characterized by low mean scores on the B-L RI (slope $= -.593, SE = .074, p < .001$). These results are illustrated in Figure 2. The difference in SC-PFT change from intake to termination for the patients of the more Rogerian versus the less Rogerian therapists was 11.06 SC-PFT units, or about two-thirds of the $SD$ of SC-PFT at intake.

Because of the association of SC-PFT with measures of depression (Zuroff, Blatt, Sanislow, Bondi, & Pilkonis, 1999), it is conceivable that the results for SC-PFT were not independent findings, but simply reflected shared variance with the CMI. We, therefore, repeated the analyses, including the CMI as a time-varying covariate. The same pattern of results was obtained. The BT-BLRI $\times$ Time effect remained significant ($F(1, 491) = 4.43, p < .05$, effect size $r = .190$), although the regression parameter was somewhat smaller ($-.025, SE = .012$). The regression coefficients for the BT-BLRI $\times$ Time and WT-BLRI $\times$ Time effects remained significantly different, $t(491) = -1.99, p < .05$.

Because initial severity moderated some of the primary TDCRP analyses (Elkin et al., 1989), the preceding analyses for the CMI and for SC-PFT were repeated,
introducing initial severity as assessed by the HRSD and then initial severity as assessed by the GAS as potential moderators. In neither case did the triple interaction of severity × BT-BLRI × Time or the triple interaction of Severity × WT-BLRI × Time achieve statistical significance. Thus, there was no evidence that the between-therapist or within-therapist components of the B-L RI had different effects on rate of improvement for initially more severely or less severely depressed patients.

**Discussion**

The fact that therapeutic outcome is predicted by dimensions of the therapeutic relationship is well-established (Norcross, 2002), and in particular, such relationships have been extensively examined within the TDCRP (e.g., Krupnick et al., 1996; Zuroff & Blatt, 2006). The importance of the present findings lies in the explication of why these associations exist. Our analyses were an attempt to replicate and extend Baldwin et al.’s (2007) finding that the relation between the therapeutic alliance and outcome is explainable in terms of differences between therapists in the average strength of the alliances they establish. Three principal results emerged. First, we found that between-therapist variability in patients’ perceptions of the Rogerian conditions was related to overall clinical improvement, as indexed by the CMI. Second, we found that between-therapist variability in perceptions of the Rogerian conditions was related to reduced vulnerability to depression, as indexed by SC-PFT. These findings applied equally across the CBT, IPT, and PLA-CM conditions, and they applied equally to patients with low and high levels of initial severity. Finally, there was a smaller but statistically significant effect demonstrating that better outcomes on the CMI were achieved by patients who experienced a higher level of the Rogerian conditions than the average patient seen by their therapist. Thus, the present analyses demonstrate that previously reported associations between Rogerian conditions and outcome in the treatment of depression are primarily, but not exclusively, explainable in terms of differences between therapists. That conclusion underlines the importance of researchers’ devoting increased attention to therapists as key determinants of outcome (Baldwin et al., 2007; Wampold, 2001).
**Between-Therapist Effects**

Between-therapist differences in aspects of the therapeutic relationship emerged as predictors of outcome in both Baldwin et al. (2007) and the present analyses, despite numerous methodological differences between the studies. Thus, there appears to be robust evidence that variability between therapists in the capacity to establish constructive therapeutic relationships is a key predictor of outcome. The present analyses extend Baldwin et al.’s findings, by focusing on a different dimension of the therapeutic relationship (Rogerian conditions vs. therapeutic alliance) and by examining changes in vulnerability as well as changes in maladjustment.

The proportion of between-therapist variance in B–L RI scores was about 10% in the TDCRP, and the proportion of between-therapist variance in working alliance scores was about 6% in the Research Consortium dataset (S.A. Baldwin, personal communication, April 16, 2009). The association of outcome and between-therapist variability is all the more striking given that the large majority of variance in therapeutic relationship measures appears to be within-therapist. In other words, it is the small fraction of between-therapist variance that does most of the “work” in predicting outcome. This may be one reason for the generally modest size of most reported correlations between measures of the therapeutic relationship and outcome (Martin, Garske, & Davis, 2000); these overall correlations confound stronger between-therapist effects with weaker within-therapist effects. The degree to which the total correlation masks the between-therapist effect is related to the ratio of patients to therapists; the more patients per therapist, the greater the extent to which the total correlation will be attenuated.

**Within-Therapist Effects**

Baldwin et al. (2007) did not find a significant effect of within-therapist differences in therapeutic alliance on outcome, whereas we found a significant effect of within-therapist differences in perceived Rogerian conditions on maladjustment. It is possible that the discrepancy between the two studies simply reflects the greater power afforded by our repeated measures design. Alternatively, the within-therapist effect on the CMI might be specific to the patient population (depressed outpatients) or the predictor variable (Rogerian conditions) used in the TDCRP. Depressed individuals are sensitive to cues of rejection (Ayduk, Downey, & Kim, 2001), and so perceiving one’s therapist as displaying unconditional positive regard might be especially therapeutic for individuals with major depression. No within-therapist effect of perceived Rogerian conditions was found for the SC-PFT; this might reflect the slower pace at which the vulnerability of SC-PFT changes compared with maladjustment (Hawley, Ho, Zuroff, & Blatt, 2006). A within-therapist effect on vulnerability measures such as SC-PFT might become apparent in longer-term treatments than the ones provided in the TDCRP.

The relatively weak or absent effects of within-therapist measures are somewhat counter-intuitive and puzzling. One possible explanation is that between-therapist differences are measured more reliably and validly because they are derived by aggregating over multiple patients. Within-therapist differences reflect ratings by a single patient at a single point in time and may contain irrelevant variance related to patients’ response biases in using the rating scales (Baldwin et al., 2007). Because between-therapist measures contain more valid variance, they would be expected to yield stronger effects. A second possibility is based on the contextual model of psychotherapy (Wampold, 2001), which suggests that it is the therapeutic context as
a whole, rather than its individual constitutive elements, that is responsible for therapeutic change. It may be that between-therapist differences in relationship measures serve as markers of differences in the global therapeutic context (Wampold, 2001); that is, therapists who are consistently perceived as high in the Rogerian conditions may also be the therapists who are able to orchestrate the full range of contextual factors (e.g., fostering hope, providing a coherent and meaningful treatment rationale, and providing a treatment that is consistent with the rationale) into a transformative therapeutic experience. If between-therapist measures are indices of the overall therapeutic context, rather than merely one part of the context, they would be expected to be strongly related to outcome.

**Generalizability**

Well-designed and comprehensive as it was, the specific design features of the TDCRP limit the generalizability of our results. Those features include the short-term, manualized nature of the treatments, the limited range of well-trained, well-supervised doctoral-level therapists, the homogeneous sample of depressed outpatients, and the random assignment of patients to conditions. Because each of these features differs in the Research Consortium dataset analyzed by Baldwin et al. (2007), we believe that the core findings are likely to be highly generalizable. Nevertheless, there is certainly a need for studies that investigate the impact of between-therapist and within-therapist components of relationship variables in other psychiatric disorders and in other forms of treatment.

**Implications**

Research and training in mental health services remain focused on differences between treatments rather than differences between treatment providers (Blatt & Zuroff, 2005; Kim et al., 2006). What are the implications of acknowledging that between-therapist differences are important (arguably, more important) sources of differences in patients’ outcomes?

First, more research effort needs to be devoted to determining the characteristics that distinguish therapists who, in general, tend to be experienced as providing low or high levels of the Rogerian conditions. A wide variety of characteristics could be pertinent, including personality traits, interpersonal styles, interpersonal skills, and preferred therapeutic strategies and interventions. Ackerman and Hilsenroth’s (2003) review of predictors of the alliance can provide guidance for investigations of therapist variability in providing Rogerian conditions. They found that several therapist attributes (e.g., flexible, respectful, warm, and open) and several types of interventions (e.g., reflection, support, affirming, accurate interpretation, and facilitating affective expression) were associated with more positive alliances. Dispositional qualities of the therapist such as the Big Five traits (Chapman, Talbot, Tatman, & Britton, 2009) and compassion (Gilbert, 2009) are also plausible antecedents of the capacity to provide high levels of the Rogerian conditions.

Second, psychotherapy researchers need to include multiple therapists in their designs whenever possible so as to permit the separation of between-therapist and within-therapist effects. Moreover, researchers should bear in mind that the conceptual distinction between within-therapist and between-therapist effects, and the statistical machinery for separating them, do not apply solely to therapeutic relationship variables. They can be applied to almost any variable related to psychotherapy process, including, for example, clients’ adherence to treatment.
clients’ motivation for treatment, clients’ interpersonal behavior towards their therapists, and therapists’ behavior towards their clients.

Finally, for graduate programs charged with training therapists and certifying their competence, the key implication is that trainees’ average levels of therapeutic relationship variables should be targets of training and ongoing monitoring. Training programs should view these outcomes as at least equally important as the acquisition of the specific skill sets associated with particular schools of therapy.

References


