Published in final edited form as:

J Consult Clin Psychol. 2010 June ; 78(3): 420–428. doi:10.1037/a0019549.

Interpersonal Accuracy of Interventions and the Outcome of Cognitive and Interpersonal Therapies for Depression

Paul Crits-Christoph,

Department of Psychiatry, University of Pennsylvania

Mary Beth Connolly Gibbons,

Department of Psychiatry, University of Pennsylvania

Christina M. Temes,

Department of Psychiatry, University of Pennsylvania

Irene Elkin, and

School of Social Service Administration, University of Chicago

Robert Gallop, Ph.D.

West Chester University

Abstract

Objective—The purpose of the current investigation was to examine the interpersonal accuracy of interventions in cognitive and interpersonal therapy as a predictor of the outcome of treatment for patients with major depressive disorder.

Method—The interpersonal accuracy of interventions was rated using transcripts of treatment sessions for 72 patients who were being treated with cognitive or interpersonal therapy for major depressive disorder through the NIMH Treatment of Depression Collaborative Research Program. Interpersonal accuracy of interventions was assessed by first identifying core conflictual relationship themes for each patient and then having judges rate therapist intervention statements for the extent to which each statement addressed each component of the patient-specific interpersonal theme.

Results—Using early in treatment sessions, statistically significant interactions of interpersonal accuracy of interventions and treatment group in relation to outcome were evident. These findings included significant accuracy by treatment group interactions in the prediction of subsequent change of depressive symptoms and social adjustment from Week 4 to Week 16, with higher levels of interpersonal accuracy associated with relatively poorer outcomes for patients receiving cognitive therapy but relatively better outcomes for patients in interpersonal therapy.

Conclusion—The process of interpersonal and cognitive therapies may differ in important ways. Accurately addressing interpersonal themes may be particularly important to the process of interpersonal therapy, but not cognitive therapy.

Correspondence concerning this article should be addressed to Paul Crits-Christoph, Ph.D., University of Pennsylvania, Center for Psychotherapy Research, Department of Psychiatry, 3535 Market Street, Room 650, Philadelphia, Pennsylvania 19104-3309, crits@mail.med.upenn.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/ccp

Keywords

interpersonal therapy; cognitive therapy; depression; therapeutic process; outcome

A number of studies have suggested that the quality, or accuracy, of therapist interventions predicts treatment outcome (Crits-Christoph, Cooper, & Luborsky, 1988; Piper, Joyce, McCallum, & Azim, 1993; Norville, Sampson, & Weiss, 1996) and the development of the therapeutic alliance (Crits-Christoph, Barber, & Kurcias, 1993) in psychodynamically-oriented psychotherapy. While targeting therapist interventions to core interpersonal themes is a mainstay of focal psychodynamic therapies (e.g., Luborsky, 1984; Strupp & Binder, 1984), the concept of tailoring treatment to the individual problems and themes of each patient is also relevant to therapeutic approaches other than psychodynamic therapy. For example, in cognitive therapy (CT), there has been attention to the role of addressing interpersonal schemas in treatment (Safran & Segal, 1990).

The purpose of the current study was to examine the interpersonal accuracy of therapist interventions in interpersonal therapy (IPT) and CT using transcripts of treatment sessions from the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989). In the TDCRP, patients were randomized to CT, IPT, imipramine plus clinical management, or placebo plus clinical management. Patients in all treatments improved significantly across treatment. There was no evidence that across the full sample either psychotherapy was more or less effective than the other (although severely depressed patients may have fared better in the interpersonal psychotherapy) (Elkin et al., 1995).

Our hypothesis was that greater therapist accuracy in addressing patient-specific interpersonal issues would predict treatment outcome across both IPT and CT; however, we were also interested in exploring any potential differences between IPT and CT. We examined accurate interventions in relation to two different types of outcomes: (1) change in depressive symptoms and (2) improvement in interpersonal problems and functioning.

In testing these hypotheses our goal was also to address a limitation of the interpretability of previous findings (e.g., Crits-Christoph et al., 1988; Piper et al., 1993; Norville et al., 1996) linking therapist accuracy with treatment outcome. The limitation concerns the direction of causal influence. Previous studies of therapist accuracy in relation to treatment outcome, as is typical of almost all studies linking treatment process to outcome, measured outcome from pre- to posttreatment. However, improvement from baseline to the point in treatment that the therapeutic process is evaluated may influence that therapeutic process. For example, therapists may find it easier to be accurate (or empathic) with patients who have begun to improve since baseline and have consequently increased their motivation for therapy and engagement in the treatment process. Initial improvement in treatment is often highly associated with final outcome (Crits-Christoph et al., 2001), thereby producing a potential spurious finding between process and outcome if outcome is assessed from baseline to termination.

Method

Participants

The current study used a sample of patients from the NIMH TDCRP (Elkin et al., 1989). The TDCRP examined the comparative efficacy of CT, IPT, pharmacotherapy (plus clinical management), and placebo (plus clinical management) for the treatment of major depressive disorder. Patients gave informed consent for participation in the TDCRP and specifically

agreed to allow audiotape records of sessions to be made for research purposes. Transcripts of psychotherapy sessions for patients who received either CT or IPT from two of the three sites were used in the current investigation (tapes from the third site were not permitted to be released from that location). Of the 80 participants who entered CT or IPT at these two sites, transcripts were made for 72 patients. Seven of the 80 were excluded because they did not attend at least 14 treatment sessions, and tapes for another patient were not audible.

Each of the 72 participants had four sessions drawn from the early phase of therapy (Sessions 2, 3, 4, and 5 whenever possible; Session 1 was excluded because it often might have had the quality of an initial interview) that were transcribed. Sixty-five of the 72 participants also had four late-in-treatment sessions that were transcribed for this investigation. The late-in-treatment sessions were typically Sessions 11, 12, 13, and 14, but there were deviations from this standard due to attrition. However, in order to differentiate the “late” phase of therapy from the “early” phase, we required at least three intervening sessions between the last early session and the first late session. The late sessions also excluded the final two therapy sessions because the focus of those sessions was presumed to be the treatment termination. Although it was planned that prediction of treatment outcome would be conducted using only the early treatment sessions, late sessions were included because it was possible that therapist interpersonal accuracy would only be apparent in late sessions (once the therapist got to know the patient or had moved beyond the standard CT or IPT techniques after patients had shown symptomatic improvement). Interpersonal accuracy in early compared to late sessions was therefore evaluated to check on this possibility.

The average age of the 72 patients was 34.5 years ($SD = 8.5$). Twenty-four percent were male, and 42% were married or living with a significant other. Ninety-three percent were Caucasian, and 7% were African American, Asian, Hispanic, or members of another minority group. Sixty-nine percent had received some college education.

Treatments

Treatments used in the TDCRP were manual-guided CT (Beck, Rush, Shaw, & Emery, 1979) and IPT (Klerman et al., 1984). CT emphasizes techniques that focus on correcting the maladaptive beliefs that underlie a patient’s distorted, negative thoughts about themselves, the world, and the future. IPT is oriented towards helping the patient understand his/her interpersonal problems and improve his/her social functioning. Treatment focuses on role transitions, interpersonal deficits, interpersonal disputes, and/or interpersonal loss/grief.

Therapists

Twelve therapists (5 CT; 7 IPT) treated the 72 patients. On average, therapists (67% male) had 13 years ($SD = 6$) of clinical experience and were 43 years old ($SD = 8$).

Transcripts of Sessions

Therapy sessions were transcribed from audiotapes using the transcription rules described by Mergenthaler and Stinson (1992). The transcripts were then checked and corrected during which time all identifying information in the sessions was changed to protect confidentiality. A total of 288 early sessions and 260 late sessions were transcribed.

Measures

Assessing the interpersonal accuracy of therapist interventions during treatment sessions involved several steps. Using the transcribed sessions, these steps included: (1) identification of the narratives (“relationship episodes”) that patients told about their interactions with other people, (2) rating of the interpersonal themes apparent in these narratives, (3) identification of the main therapist interventions in each session, and (4) rating of the extent

to which the salient interpersonal themes were addressed by the therapist within each main intervention. Each of these steps was performed by a different set of trained independent judges. The procedures for accomplishing each of these steps are described below.

Identifying narratives—Two judges (research assistants) were trained to locate relationship episodes. Training consisted of 20 practice sessions and extensive review of the principles for locating relationship episodes, as described in Luborsky and Crits-Christoph (1998). Periodic recalibration sessions were scheduled throughout the period the judges worked on the actual study materials.

Relationship episodes are sections of psychotherapy sessions in which patients describe specific interactions between themselves and another person in their life (Luborsky & Crits-Christoph, 1998). For each session, judges independently identified the beginning and end of each relationship episode. Judges used a 1-to-5 Likert scale to rate the episodes on “completeness,” which was defined in terms of extent to which components of the relationship theme were present. Higher scores indicated higher levels of completeness. Interjudge reliability assessed via the intraclass correlation coefficient (ICC (2,2); Shrout & Fleiss, 1979) was found to be .75 for completeness ratings (average of two judges); the average completeness rating was 2.4 ($SD = 0.9$). As in previous research (Luborsky & Crits-Christoph, 1998), only those episodes with an average completeness rating of 2.5 or greater across judges were retained. There was almost complete agreement in identifying these episodes. Of those episodes identified by Judge 1 as having a completeness of 2.5 or greater, Judge 2 located the same episode 95% of the time. Of those episodes identified by Judge 2 as having a completeness rating of 2.5 or greater, Judge 1 located the same episode 98% of the time. Since there were some discrepancies between the two judges in the start and end points identified for each episode, a third consensus judge was used to resolve these discrepancies.

Assessment of interpersonal themes—Interpersonal patterns were assessed from the psychotherapy narratives using the Quantitative Assessment of Interpersonal Themes method (QUAINT; Crits-Christoph et al., 1994), an integration of Luborsky’s and Crits-Christoph’s (1998) CCRT method and Benjamin’s (1974) Structural Analysis of Social Behavior (SASB).

In the QUAINT method, judges rate the presence of the three components of the CCRT—wishes, responses of self, and responses of other—using a standard language derived from the cluster model of Benjamin’s Structural Analysis of Social Behavior (SASB; Benjamin, 1974) and the parallel Structural Analysis of Social Affect (SASA; Crits-Christoph and Demorest, 1989; Benjamin, 1986). The SASB and SASA models produce a total of 32 wishes (e.g., to be freeing and forgetting, for other to be affirming and forgetting, to be loving and approaching, for other to be disclosing and expressing to me), 32 responses of other (same content areas as with wishes), and 40 responses of self (e.g., hostile and angry, loving, sad, disgusted, fearful, joy, helpless). Each of these 104 items is rated on a 1-to-5 scale capturing the extent to which the item is present within each relationship episode.

For the current study, QUAINT judges were trained using 30 practice cases and regular meetings to review ratings. Once trained, monthly recalibration meetings were conducted using non-study sessions throughout the duration of time judges were rating study episodes. All relationship episodes were extracted from the psychotherapy sessions and rated in random order by the judges. A pool of five PhD-level judges and one advanced graduate student were used as QUAINT judges, with three randomly paired judges used to rate each relationship episode (balanced incomplete block design).

The interrater reliabilities for the QUAINT items in this sample are presented by Crits-Christoph, Connolly, and Shaffer (1999). ICCs for the three judges pooled were derived from a generalizability model including random factors for treatment group, therapist, patient, session, narrative, and judge. Sixty-six of the 104 QUAINT items achieved interjudge reliabilities of at least .50 and were used to construct final QUAINT profiles for each patient after averaging scores over the three raters. (In six cases, the items were actually composite scores of two items, where it seemed that raters had difficulty distinguishing between two neighboring items on the SASB circumplex.)

We conducted analyses of the QUAINT data, separately for each patient, with the goal of identifying the most common wishes, responses from others, and responses of self, that occur most repetitively across various relationship episodes for each patient. We use the term “episode profile” to refer to the QUAINT ratings (i.e., scores on 66 separate QUAINT items) for a given relationship episode. When the same, or similar profile, is found in many episodes for a given patient, we refer to this as that patient’s “main theme.”

Following Connolly, Crits-Christoph, Demorest, Azarian, Muenz, and Chittams (1996) and Connolly, Crits-Christoph, Barber, and Luborsky (2000), we used cluster analysis to extract themes from the episode profiles for each patient. As a first step, within each patient, the 66-item QUAINT profiles were collapsed across episodes about the same other person if the interpersonal themes were redundant (i.e., if the median correlation between episodes across items was at least .50). Episodes about the same other person that correlated less than .50 were retained as separate QUAINT profiles. Next, the QUAINT profiles were intercorrelated across episodes for each patient separately and then cluster analyzed using the method of average linkage.

We next identified episodes which were similar in terms of QUAINT themes by examining the resulting cluster analysis dendrogram for each patient. We identified the level of the dendrogram hierarchy in which all clusters had a median within cluster correlation of profiles of QUAINT ratings of at least .30. An average QUAINT profile for each cluster was derived by averaging the QUAINT profiles across the episodes included in each cluster. The cluster that contained the most episodes was used to identify the main theme for each patient. The single wish, response from other, and response of self that occurred in the most episodes in this cluster was identified to be rated for accuracy of intervention.

Coding of therapist statements—Each “therapist speaking turn” was identified from the transcribed sessions. Therapist speaking turns were combined into a single statement if the statement was interrupted mid-sentence by a brief patient remark and continued in the next statement. Brief therapist statements that occurred during a patient speaking turn (such as an “Mm-hm” on the therapist’s part while the patient was speaking) were counted as individual therapist speaking turns. A total of 135,352 therapist statements were coded across the 548 treatment sessions.

Therapist statements were coded using a system modified from the response mode categories described by Elliott et al. (1987) as common across six previously published coding systems. Our final system included eight statement categories: learning statement, clarification, question, restatement, role play, informational, self disclosure, or other (Connolly, Crits-Christoph, Levinson, Gladis, Siqueland, & Barber, 2002).

We did not use the interpretation category used by most existing therapist response mode systems because this category is associated closely with a psychodynamic perspective. Instead, we defined a general category called learning statements that could capture the main therapeutic statements implemented by both interpersonal and cognitive therapists. These

learning statements were used to represent the therapist interventions for the rating of interpersonal accuracy in the current study. The definition of a learning statement as “any statement that helped the patient become aware of a thought, feeling, or behavior” was general enough to identify the central interventions used in each treatment modality. Learning statements might only point out a patient’s thoughts, feelings, or behaviors (e.g., “It sounds like you are beginning to accept yourself for who you presently are, with your abilities, but also with your limitations”) or might describe a causal link between a thought, feeling, or behavior (e.g., “So, you will not approach women because you believe you will be rejected”). Learning statements might also describe a pattern of behaviors or link past behavior to present behavior.

Seven advanced graduate students were trained in our statement classification system, and randomized in a balanced incomplete block design so that three independent judges coded the therapist statements for each session transcript. Judges were trained using three training sessions followed by discussion of their ratings in a group. Ratings of therapist statements were made over the course of a year, during which time judges participated in monthly recalibration sessions to maximize reliability and prevent rater drift. A descriptive analysis of all statement categories for these sessions, including the learning statements used in this study, is provided by Connolly et al. (2002). The median pooled judge intraclass correlation, ICC (2,3), for the identification of learning statement category was .77 across five successive, random sub-samples of 2,000 therapist statements.

All statements identified by at least two of the three judges as learning statements were extracted from the transcripts and used for the final step of assessing the interpersonal accuracy of therapist interventions. Learning statements represented 4% of cognitive session statements and 6% of the interpersonal therapist statements. The average CT session contained 10 learning statements, and the average IPT session contained 11 learning statements. A total of 5,603 learning statements were extracted from the 548 sessions and rated for their interpersonal accuracy.

Ratings of interpersonal accuracy of interventions—Accuracy of interventions was assessed using a 1 (“not at all” addressed in the therapist statement) to 7 (“definitely” addressed) rating scale that measured the extent to which the interpersonal content of the patient’s main theme, as measured by the QUAINT method, matched the content of the therapists’ learning statements. The term “accuracy” is not used here to convey that the therapist was right or wrong in their interventions, but rather that a given therapist statement “hit the mark” in terms of an independently derived evaluation of each patient’s interpersonal themes.

Each of the three CCRT components (wish, response from other, response of self) contained in the QUAINT system was separately rated for accuracy by a set of three judges. An example will help illustrate the accuracy ratings. The following is an edited version of a learning statement provided by a therapist in the TDCRP: “Really what’s bothering you, what’s upset you with (the woman at work), is you really don’t like being left out. You want to be part of the group...you want them to invite you in. That’s what you seem to want, but then you see her as having no interest in you. What I’m saying is, what would it take at work, what could you do there?” The main theme for this patient consisted of the wish “for the other to be loving and approaching”; the response from other was “ignoring and neglecting”; the response of self was “feels sad.” This learning statement would have been rated moderately high on accuracy with regard to the main theme wish and response from other, but rated as a “1” (*not accurate*) for the main theme response of self.

The intraclass correlation coefficients ([ICC 2,3]; Shrout & Fleiss, 1979) indicated good interjudge reliability for accuracy ratings. Across the 72 patients, the coefficients for accuracy on the wish, response from other, and response of self component were .81, .81, and .91, respectively.

The final scores for accuracy on each CCRT component were computed by first averaging across the three judges. Average session accuracy scores were then created by averaging accuracy ratings across all learning statements rated within each session. Finally, early-in-treatment accuracy scores were then created by averaging over the 4 early sessions, and late-in-treatment accuracy scores were created by averaging over the 4 late sessions. The early-in-treatment wish, response from other, and response of self accuracy scores were relatively uncorrelated, Wish-RO: $r(70) = -.16$, *ns*; Wish-RS: $r(70) = -.13$, *ns*; RO-RS: $r(70) = .00$, *ns*. These measures were therefore retained as separate predictors rather than creating an overall accuracy score.

Depressive symptomatology—The 17-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), administered at baseline and Weeks 4, 8, 12, and 16, was used as an interview-based measure of depressive symptoms. Intraclass correlations, calculated based on ratings done by interviewers at each site rating tapes from all sites for the TDCRP, ranged from .93 to .96.

Measure of social adjustment—The Social Adjustment Scale (SAS; Weissman & Pakel, 1974) was used to measure interpersonal problems and functioning. The SAS consists of 59 items rated by a clinical evaluator covering the domains of work, social and leisure activities, extended family relationships, marital or partner relationships, parental relationships, and family unit functioning. All items are rates on a 1-to-5 scale, with a high score indicating greater impairment in social adjustment. Reliabilities (ICCs) for the global SAS score ranged from .69 to .75, and for the social and leisure score from .90 to .97, based on clinical evaluators from each site rating tapes of SAS interviews drawn from all sites. Because we were interested in the impact of accurate interpersonal interventions on interpersonal functioning in current relationships, we used a modified scoring of the SAS developed by Crits-Christoph et al. (1999) that focused only on a subset of 12 items that reflected problems in ongoing relationships (items that captured a lack of interpersonal relationships were excluded). These 12 items included two from the social leisure domain (friction in relationships and hypersensitivity to others), four items from the extended family domain (friction in family relationships, worry, guilt, and resentment), and six items from the marital/partner domain (friction, submissive behavior, lack of affection, disinterest in sex, diminished sexual intercourse, and sexual problems). An average of the 12 items was calculated (if an item was not applicable the average of the non-missing items was obtained). Internal consistency of this 12-item scale, calculated with a split-half coefficient, was .65 (a split-half coefficient was used because not all items applied to all patients, particularly if a patient did not have a current partner). The SAS was administered at baseline and Weeks 4, 8, 12, and 16.

Data Analyses

Two sets of preliminary analyses were conducted. The first was to examine whether there were mean differences in accuracy between early and late sessions. Our primary interest was in using early-in-treatment accuracy scores as predictors of outcome so that we could predict outcome subsequent to the assessment of the process measure. However, before doing so, it was important to examine whether accuracy occurred in early sessions at similar levels as evident in later sessions. To test for this, the three accuracy scores were analyzed using a repeated measures analysis of variance with one within-group factor (early vs. late sessions)

and one between-group factor (CT vs. IPT) to examine mean differences over time and treatment group.

A second preliminary analysis was examination of mean differences between therapists in their accuracy scores. This was done using analyses of variance specifying therapist as a random effect nested within treatment condition.

The primary predictive analyses for the HRSD were conducted using Hierarchical Linear Modeling (HLM) techniques, specifying the HRSD over time (Weeks 4, 8, 12, and 16) as the dependent variable. As per Gibbons et al. (1993), $\log[\text{weeks}+1]$ was entered as time. The analysis modeled the linear slope over time for each patient, with early-in-treatment accuracy scores as second level predictor variables of the slopes. Change in scores from baseline to Week 4 on the HRSD along with marital status were entered as covariates. Using change from baseline to Week 4 as a covariate controlled for the impact of early improvement on the outcome measures. Marital status was used as a covariate because it had been previously found to predict outcome in the TDCRP and was used as a covariate in the primary outcome analyses (Elkin et al., 1989). A parallel analysis was conducted for the modified SAS, with change from baseline to Week 4 on the SAS used as the covariate to control for early improvement on this measure. Marital status was also used as a covariate in the model predicting the SAS. In these HLM models, a random intercept and random slope was specified. The covariance structure between the two random effects was specified as unstructured. Effect sizes (converted to Cohen's d), derived from the F -test for the mixed

effects model, were calculated as $d = 2 \sqrt{\frac{F}{df}}$, where F is the F -test statistic for the effect of interest in the repeated measures model as well as other multilevel designs (Rosenthal & Rosnow, 1991; Verbeke & Molenberghs, 2000).

Results

Mean Level of Interpersonal Accuracy of Interventions in CT and IPT and Therapist Effects

The mean accuracy scores for early and late sessions are presented in Table 1 separately by treatment modality. Results of the repeated measures analysis of variance revealed no significant effects for time, treatment, or treatment by time for accuracy on the wish or response from other. However, we found significant effects for treatment for accuracy on the response of self ($F(1, 62) = 4.45, p = .039, d = .79$). The CT group had higher ratings on accuracy on the response of self than did the IPT group. There were no significant effects for time, or treatment by time, for these accuracy measures. Correlations between early in treatment accuracy scores and later in treatment scores varied depending on the CCRT components, with the early accuracy on the wish scores moderately correlated with the late in treatment accuracy on the wish scores ($r = .55, p < .0001$), but response from other accuracy scores uncorrelated ($r = .03$) between early and later sessions. The correlation between accuracy on the response of self from early to late sessions was $.38$ ($p = .002$).

Using the three early-in-treatment accuracy scores, analyses testing for mean differences between therapists revealed no significant findings (all p -values $> .25$).

Prediction of Change in Depression and Social Adjustment from Early Interpersonal Accuracy of Interventions

The two dependent variables used in the analyses of prediction of outcome showed only relatively small intercorrelations. The correlation between individual differences in slope coefficients for the HRSD and the slope coefficients for the modified SAS was $r(70) = .28, p = .03$.

Predictive analyses focusing on each of the three individual early-in-treatment accuracy measures revealed no significant main effects (see Table 2). However, incorporation of cross-product (interaction) terms in the regression models to test for differential patterns between the two treatment modalities revealed several significant findings (Table 2).

There was a significant interaction between treatment and accuracy on the wish in predicting symptom course on the HRSD, $F(1, 63) = 4.4, p = .04, d = .51$. The interaction effect is displayed in Figure 1. Within each treatment group, the average slope of change on HRSD per unit increase in accuracy was 6.60 ($SE = 3.43, p = .05$) for CT and -2.40 ($SE = 2.55, p = .35$) for IPT. For CT, lower accuracy scores were indicative of greater reduction in HRSD scores whereas for IPT, higher accuracy scores were related to more symptom reduction. There was also a significant interaction of treatment and accuracy on the wish in the prediction of change on the modified SAS, $F(1, 63) = 5.1, p = .02, d = .57$. The average slope of change on the modified SAS per unit increase in accuracy was .44 ($SE = .35, p = .20$) for CT and $-.54$ ($SE = .26, p = .04$) for IPT. As seen in Figure 2, for CT, lower accuracy scores were again associated with greater improvement in modified SAS scores, whereas for IPT, higher accuracy scores were associated with greater improvement in modified SAS scores. A significant interaction of treatment and accuracy on the response of self in the prediction of modified SAS scores was also apparent, $F(1, 64) = 3.8, p = .05, d = .49$. The direction of this effect was the same as above (see Figure 3). Average slope was $-.35$ ($SE = .13, p = .008$) for CT and $-.04$ ($SE = .15, p = .77$) for IPT.

Discussion

Our findings give some support the hypothesis that the interpersonal accuracy of learning statement interventions is important to treatment outcome especially as measured by social adjustment. However, our results suggest that the direction of the relation between interpersonal accuracy of learning statement interventions and treatment outcome differs by type of treatment. Specifically, higher accuracy scores were related to poorer outcomes in CT but relatively better outcomes in IPT. These differential effects were found in the prediction of various outcomes, including depressive symptoms and social adjustment over the course of active treatment.

The accuracy scores for the three CCRT components (wish, response from other, response of self) were relatively uncorrelated. This indicates that when therapists in the current study were accurate on one aspect of a relationship pattern they were not necessarily accurate on other facets of the pattern. Perhaps this result is not unexpected given that the therapists in the TDCRP were not trained in the CCRT method and therefore did not likely think about relationship themes in the same coherent way as contained within the CCRT method. However, because prediction of outcome was achieved for accuracy scores on both the wish and response of self elements, our findings suggest that addressing each of these aspects of a relationship theme may be useful. In addition to low correlations among the accuracy scores, there was no evidence of therapist effects on the accuracy scores. These findings suggest that it is not the case that some therapists are consistently “accurate” and others not “accurate.” In the TDCRP sample, interpersonal accuracy of learning statements varied from dyad to dyad.

This study addresses a major methodological weakness of previous studies of interpersonal accuracy of therapists’ interventions by including data from multiple assessment points across treatment. Previous studies of interpersonal accuracy (Crits-Christoph et al., 1988; Norville et al., 1996; Piper et al., 1993) only assessed outcome pre- and posttreatment. The existence of the 4-week HRSD and SAS assessments in the TDCRP permitted early improvement (from baseline to Week 4) to be statistically controlled; the assessments at

Weeks 4, 8, 12, and 16 permitted a better estimate of the trajectory of change over treatment. By controlling for early improvement on the HRSD and SAS, the current findings suggest that the relation between interpersonally accurate learning statement interventions and treatment outcome is not a spurious correlation produced by early improvements leading to both more accurate interventions and better treatment outcomes.

The finding of a differential relationship between interpersonal accuracy and treatment outcome in IPT and CT suggests that these treatments may, in part, achieve comparable results through different mechanisms. Within the context of a psychotherapy that focuses on interpersonal relationships, it is perhaps not surprising that more accurate learning statement interventions lead to better treatment outcomes. Such accurate interventions are likely to be seen as more empathic and have the potential to increase patient insight into the nature of their interpersonal difficulties. What is surprising is that accurate learning statement interventions led to relatively poorer outcomes in CT.

One factor to consider in understanding the relation between interpersonal accuracy and outcome in CT is that a previous report documented that CT tends to have fewer relationship episodes than IPT as well as more therapist words spoken while patients are describing a relationship episode (Crits-Christoph et al., 1999). Because of the greater amount of interpersonal material in sessions, it might be expected that IPT therapists would be able to have a better understanding of patients' interpersonal themes and therefore achieve relatively higher levels of accuracy in their interventions. However, on most of the accuracy measures, no mean differences between CT and IPT were apparent. On one measure (accuracy on the response of self), CT was found to have higher mean levels of accuracy compared to IPT. In addition, the standard deviations for the accuracy measures were similar or higher in CT compared to IPT. Thus, the differential prediction of outcome across IPT and CT for some of the accuracy measures does not appear to be due to relatively lower levels, or lower variability, of accuracy scores in CT preventing a similar relation emerging with outcome as was found in IPT.

Our finding of differential prediction in CT versus IPT is discrepant from previous process studies conducted using sessions from the TDCRP. Krupnick et al. (1996) reported that the therapeutic alliance predicted outcome across all four treatment conditions in the TDCRP, with virtually no differences between treatments in the nature of the relationship between alliance and outcome. Ablon and Jones (1999) found that patient process qualities (e.g., "patient feels helped," "patient achieves a new understanding") were significantly related to treatment outcome across IPT and CT. Based on their findings, these authors have emphasized that the common elements in these psychotherapies are the factors responsible for the general finding of no difference between the psychotherapies in treatment outcome. Although common factors may well have contributed to the result of no differences between the psychotherapies in treatment outcomes, the current findings suggest that processes unique to each treatment may also be contributing. Interpersonal accuracy, as defined here, may be facilitating positive outcomes in IPT but hindering positive outcomes in CT. In CT, the quality of implementation of CT techniques may be leading to better outcomes (Shaw et al., 1992).

The negative relationship between interpersonal accuracy of interventions and treatment outcomes in CT may appear discrepant from other studies that have pointed to the positive impact of an interpersonal or psychodynamic focus in CT. In particular, Hayes, Castonguay, and Goldfried (1996) found that interventions that addressed the interpersonal and developmental domains were associated with greater improvement in CT, and Jones and Pulos (1993) found that greater use of psychodynamic techniques in CT was associated with relatively more favorable outcomes. The different findings of these studies compared to the

current one may be a function of the type of process measures used. Neither the Hayes, Castonguay, and Goldfried (1996) nor the Jones and Pulos (1993) study assessed the interpersonal content of relationship themes for each patient so that therapist accuracy in addressing these themes could be examined. It may be that doing the work of CT within interpersonal domains is helpful in CT but that focusing extensively on the content of interpersonal wish and response components, rather than on automatic thoughts and beliefs, is distracting from the primary task of CT. In addition, alternative measures of interpersonal focus in CT may be more appropriate to studying interpersonal aspects of therapist techniques. Methodological differences between the studies, such as our use of early improvement as a covariate and predicting the slope of change from Week 4 to Week 16, may also be responsible for the divergent findings.

Our predictive findings occurred with the ratings of accuracy in regard to recurrent themes that were apparent across multiple relationship episodes. Thus, it appears to be particularly important within IPT to address recurrent relationship themes rather than only addressing situational specific relationship problems. Our findings therefore suggest that the IPT model could likely be enhanced by having therapists focus more on formulating and addressing the unique recurrent interpersonal themes of each patient. IPT therapists are trained to address interpersonal issues within the context of the four main domains of interpersonal problems (role transitions, interpersonal deficits, interpersonal disputes, and/or interpersonal loss/grief) described in the IPT manual, but no specific system for formulating individual patient recurrent themes is provided for IPT therapists (although case examples are given in the manual). Therefore, to enhance IPT, the integration of formulation systems for interpersonal themes such as the CCRT method might be considered.

A number of limitations of the findings presented here should be noted. First, the specific direction of the interactions between accuracy of interpersonal interventions and treatment type were not hypothesized in advance and therefore these findings should be replicated. This is especially important because only some of the accuracy measures predicted outcome and no correction for multiple analyses was applied. Thus, the current findings need to be considered preliminary. Second, although we attempted to avoid a spurious finding that can result from predicting change from baseline to termination from process measures that are sampled after baseline, we have not ruled out the influence of other "third variables" on our correlational findings. Third, it is likely that aspects of therapist statements other than their interpersonal accuracy are relevant to treatment outcome. This appears to be especially true in CT, where interpersonal accuracy may have a negative impact. Not only are other aspects of therapist interventions important, but a complete picture of the change process would incorporate various patient, therapist, and process variables to account for treatment outcome differences across patients. Fourth, as mentioned, the measure of interpersonal accuracy of interventions may not capture the way in which interpersonal issues are addressed within the context of CT. Fifth, some of the QUAINT items retained for analysis had marginal reliabilities and many items not retained for analysis had weak reliabilities. More reliable methods of assessing an expanded set of interpersonal wishes and responses may yield stronger, or different, findings than presented here. Despite these limitations, the current study provides some clues about how the process of IPT and CT may be different and how aspects of these treatments beyond "common factors" may in part be responsible for therapeutic change.

Acknowledgments

This project was supported in part by National Institute of Mental Health grants P50-MH-45178, K02-MH00756, and R01-MH40472. The National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program was a multisite program initiated and sponsored by the Psychosocial Treatments Research Branch, Division of Extramural Research Programs (and later by the Mood, Anxiety, and Personality Disorders

Research Branch, Division of Clinical Research), NIMH, Rockville, Maryland. The program was funded by Cooperative Agreements to six participating sites (George Washington University, Washington, DC – MH 33762; University of Pittsburgh MH – 33753; University of Oklahoma, Oklahoma City – MH 33760; Yale University, New Haven, Connecticut – MH 33827; Clarke Institute of Psychiatry, Toronto, Ontario – MH 38231; and Rush Presbyterian-St. Luke's Medical Center, Chicago, Illinois – MH 35017). The principal NIMH collaborators were Irene Elkin, Ph.D., Coordinator (now at the University of Chicago); M. Tracie Shea, Ph.D., Associate Coordinator (now at Brown University, Providence Rhode Island); John P. Docherty, M.D. (now at The New York Hospital-Cornell University Medical College, White Plains, New York); and Morris B. Parloff, Ph.D. (now at Washington School of Psychiatry, Washington, DC). The Principal Investigators and Project Coordinators at the three participating research sites were as follows: George Washington University – Stuart M. Sotsky, M.D., and David Glass, Ph.D.; University of Pittsburgh – Stanley D. Imber, Ph.D., and Paul A. Pilkonis, Ph.D.; and the University of Oklahoma – John T. Watkins, Ph.D. (now at Atlanta [GA] Center for Cognitive Therapy) and William Leber, Ph.D. The Principal Investigators and Project Coordinators at the three sites responsible for training therapist were as follows: Yale University – Myrna Weissman, Ph.D. (now at Columbia University), Eve Chevron, M.S. and Bruce J. Rounsaville, M.D.; Clarke Institute of Psychiatry – Brian F. Shaw, Ph.D., (now at Hospital for Sick Children, University of Toronto), T. Michael Vallis, Ph.D., (now at Dalhousie University), K. Dobson, Ph.D., (now at the University of Calgary), and Rush Presbyterian-St. Luke's medical center – Jan A. Fawcett, M.D., and Phillip Epstein, M.D. Collaborators in the data management and data analysis aspects of the program were as follows: C. James Klett, Ph.D., Joseph F. Collins, Scd., and Roderic Gillis of the Veterans Affairs Cooperative Studies Program, Perry Point, Md.

References

- Ablon JS, Jones EE. Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*. 1999; 67:64–75. [PubMed: 10028210]
- Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. *Cognitive Therapy of Depression*. New York: Guilford; 1979.
- Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years later. *Clinical Psychology Review*. 1988; 8:77–100.
- Benjamin, LS. Adding social and intrapsychic descriptors to Axis I of DSM-III. In: Millon, T.; Klerman, G., editors. *Contemporary Directions in Psychopathology*. New York: Guilford Press; 1986.
- Caspar F, Pessier J, Stuart J, Safran JD, Samstag LW, Guirguis M. One step further in assessing how interpretations influence the process of psychotherapy. *Psychotherapy Research*. 2000; 10:309–320.
- Connolly MB, Crits-Christoph P, Demorest A, Azarian K, Muenz L, Chittams J. The varieties of transference patterns in psychotherapy. *Journal of Consulting and Clinical Psychology*. 1996; 64:1213–1221. [PubMed: 8991307]
- Connolly MB, Crits-Christoph P, Barber JP, Luborsky L. Transference Patterns in the therapeutic relationship in supportive-expressive psychotherapy for depression. *Psychotherapy Research*. 2000; 10:356–372.
- Connolly MB, Crits-Christoph P, Levinson J, Gladis L, Siqueland L, Barber JP, et al. Therapist interventions in the interpersonal and cognitive therapy sessions of the Treatment of Depression Collaborative Research Program (TDCRP). *American Journal of Psychotherapy*. 2002; 56:3–26. [PubMed: 11977782]
- Connolly MB, Crits-Christoph P, Shappell S, Barber JP, Luborsky L, Shaffer C. Relation of transference interpretations to outcome in the early sessions of brief supportive-expressive psychotherapy. *Psychotherapy Research*. 1999; 9:485–495.
- Crits-Christoph, P.; Demorest, A. Structural Analysis of Social Affect. Paper presented at the meeting of the Society for Psychotherapy Research; Toronto. 1989.
- Crits-Christoph P, Barber J, Kurcias J. The accuracy of therapists' interpretations and the development of the therapeutic alliance. *Psychotherapy Research*. 1993; 3:25–35.
- Crits-Christoph P, Connolly MB, Shaffer C. Reliability and base rates of interpersonal themes in narratives from psychotherapy sessions. *Journal of Clinical Psychology*. 1999; 55:1227–1242. [PubMed: 11045773]
- Crits-Christoph P, Connolly MB, Gallop R, Barber JP, Tu X, Gladis M, et al. Early improvement during manual-guided cognitive and dynamic psychotherapies predicts 16-week remission status. *Journal of Psychotherapy Research and Practice*. 2001; 10:145–154.

- Crits-Christoph P, Connolly MB, Shappell S, Elkin I, Krupnick J, Sotsky S. Interpersonal narratives in cognitive and interpersonal psychotherapies. *Psychotherapy Research*. 1999; 9:22–35.
- Crits-Christoph P, Cooper A, Luborsky L. The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *Journal of Consulting and Clinical Psychology*. 1988; 56:490–495. [PubMed: 3198804]
- Crits-Christoph P, Demorest A, Muenz LR, Baranackie K. Consistency of interpersonal themes for patients in psychotherapy. *Journal of Personality*. 1994; 62:499–526. [PubMed: 7861303]
- Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, et al. Initial severity and differential treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*. 1995; 63:841–847. [PubMed: 7593878]
- Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*. 1989; 46:971–982. [PubMed: 2684085]
- Elliott R, Hill CE, Stiles WB, Friedlander ML, Mahrer AR, Margison FR. Primary therapist response modes: Comparison of six rating systems. *Journal of Consulting and Clinical Psychology*. 1987; 55(2):218–223. [PubMed: 3571676]
- Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry*. 1993; 50:739–750. [PubMed: 8357299]
- Hamilton MA. A rating scale for depression. *Journal of Neurology Neurosurgery and Psychiatry*. 1960; 23:56–62.
- Hayes AM, Castonguay LG, Goldfried MR. Effectiveness of targeting the vulnerability factors of depression in cognitive therapy. *Journal of Consulting and Clinical Psychology*. 1996; 64:623–627. [PubMed: 8698959]
- Jones EE, Pulos SM. Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology*. 1993; 61:306–316. [PubMed: 8473585]
- Klerman, GL.; Weissman, MM.; Rounsaville, BJ.; Chevron, E. Interpersonal psychotherapy of depression. New York: Basic Books; 1984.
- Krupnick J, Sotsky SM, Elkin I, Watkins J, Pilkonis PA. The role of 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*. 1996; 64:532–539. [PubMed: 8698947]
- Luborsky, L. Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment. New York: Basic Books; 1984.
- Luborsky, L.; Crits-Christoph, P. Understanding transference: The Core Conflictual Relationship Theme Method. Washington, DC: American Psychological Association; 1998.
- Mergenthaler E, Stinson CH. Psychotherapy transcription standards. *Psychotherapy Research*. 1992; 2:125–142.
- Norville R, Sampson H, Weiss J. Accurate interpretations and brief psychotherapy outcome. *Psychotherapy Research*. 1996; 6:16–29.
- Persons JB. Psychotherapy outcome studies do not accurately represent current models of psychotherapy. A proposed remedy. *American Psychologist*. 1991; 46:99–106. [PubMed: 2014942]
- Piper WE, Joyce AS, McCallum M, Azim HFA. Concentration and correspondence of transference interpretations in short-term psychotherapy. *Journal of Consulting and Clinical Psychology*. 1993; 61:586–595. [PubMed: 8370854]
- Rosenthal, R.; Rosnow, RL. Essentials of behavioral research: Methods and data analysis. 2. New York: McGraw Hill; 1991.
- Safran, JD.; Segal, ZV. Interpersonal process in cognitive therapy. New York: Basic Books; 1990.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86:420–428. [PubMed: 18839484]
- Shaw BF, Elkin I, Yamaguchi J, Olmsted M, Vallis TM, et al. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*. 1999; 67:837–846. [PubMed: 10596506]

- Silberschatz G, Fretter PB, Curtis JT. How do interpretations influence the process of psychotherapy? *Journal of Consulting and Clinical Psychology*. 1986; 54:646–652. [PubMed: 3771882]
- Strupp, HH.; Binder, JL. *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy*. New York: Basic Books; 1984.
- Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data*. New York: Springer; 2000.
- Weissman, MM.; Paykel, ES. *The depressed woman: A study of social relationships*. Chicago: University of Chicago Press; 1974.

Figure 1. Interaction of treatment group with accuracy of intervention on the interpersonal wish component in the prediction of slope on the Hamilton Rating Scale for Depression (HRSD) from Week 4 to Week 16.

Figure 2. Interaction of treatment group with accuracy of interventions on interpersonal wish component in the prediction of the modified Social Adjustment Scale (SAS) from Week 4 to Week 16.

Figure 3. Interaction of treatment group with accuracy of interventions on the response of self in the prediction of the modified Social Adjustment Scale (SAS) from Week 4 to Week 16.

Table 1

Mean Interpersonal Accuracy of Intervention Scores for Early and Late Sessions

Accuracy Variable	Early-In-Treatment		Late-in-Treatment	
	Cognitive Therapy	Interpersonal Therapy	Cognitive Therapy	Interpersonal Therapy
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Wish	1.21 (0.24)	1.32 (0.29)	1.24 (0.27)	1.29 (0.26)
Response from Other	1.28 (0.33)	1.28 (0.19)	1.21 (0.22)	1.21 (0.20)
Response of Self	2.08 (0.66)	1.73 (0.46)	1.87 (0.62)	1.74 (0.46)

Table 2

Predicting Outcome from Interpersonal Accuracy of Intervention Scores

Accuracy Measure	Main Effect		Interaction with Treatment Group	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Hamilton Rating Scale for Depression				
Wish	.1	.74	4.4	.04
Response from Other	.3	.59	.9	.35
Response of Self	.1	.82	.0	.87
Social Adjustment Scale				
Wish	.6	.45	5.1	.02
Response from Other	1.6	.22	3.3	.07
Response of Self	2.5	.11	3.8	.05

Note. $DF = 1, 65$ for main effects; $DF = 1, 235$ for interactions. Separate models performed for each accuracy measure. For the Hamilton Rating Scale for Depression and the Social Adjustment Scale, the analyses involved the prediction of the slope from Week 4 to Week 16 (controlling for change from baseline to Week 4 and marital status).