

Validity of Controlled Clinical Trials of Psychotherapy: Findings From the NIMH Treatment of Depression Collaborative Research Program

J. Stuart Ablon, Ph.D.

Enrico E. Jones, Ph.D.

Objective: This research extends a series of studies that have examined the process of psychotherapy. The authors hypothesized that manualized regimens of psychotherapy compared in a controlled clinical trial would overlap considerably in process and technique and that intervention strategies common to both treatments would be responsible for promoting patient change.

Method: Expert therapists developed prototypes of the ideal regimens of brief interpersonal psychotherapy and cognitive behavior therapy using the Psychotherapy Process Q-Set, an instrument designed to provide a standard language for describing treatment processes. A separate set of clinical judges then used the Psychotherapy Process Q-Set to score the actual transcripts of interpersonal psychotherapy and cognitive behavior therapy sessions conducted as part of the NIMH Treatment of Depression Collaborative Research Program. The expert prototypes were then compared with actual therapy

administered in order to determine the extent to which each form of therapy conformed to its ideal prototype.

Results: Both the interpersonal psychotherapy and cognitive behavior therapy sessions adhered most strongly to the ideal prototype of cognitive behavior therapy. In addition, adherence to the cognitive behavior therapy prototype yielded more positive correlations with outcome measures across both types of treatment.

Conclusions: Relying on brand names of therapy can be misleading. These findings suggest that the basic premise of controlled clinical trials (i.e., that the compared interventions represent separate and distinct treatments) may not have been met in the NIMH Treatment of Depression Collaborative Research Program. The implications of these findings for using controlled clinical trials to study psychotherapy are discussed.

(*Am J Psychiatry* 2002; 159:775–783)

There is currently a sharp debate in the clinical research literature concerning what constitutes the best method for empirically validating treatments. Some researchers have argued that it is most important to determine the efficacy of a treatment (1), while others have suggested that the notion of effectiveness is more important (2, 3). Efficacy studies emphasize the importance of being able to draw a causal inference between the treatment provided and outcome (4). Such studies rely on clinical trial methods in which participants are randomly assigned to different treatment conditions that are closely controlled. It is then assumed that any change that occurs differentially across treatment conditions can be attributed directly to the causal effect of the treatment. The cost of being able to draw causal inferences lies in the fact that the study must be so closely controlled that the treatments and the patients may no longer resemble the reality of clinical practice (5, 6). Randomized clinical trials test a somewhat artificial treatment in an artificially controlled setting with atypical patients, so they have little generalizability to the real world of mental health care delivery.

Effectiveness studies, on the other hand, emphasize external validity and are less constrained by research protocols. Participants in effectiveness studies are usually selected without using exclusion criteria, and they are free to choose whatever treatment they wish to receive. The treatments are not standardized; that is to say, they are conducted as they would be in clinical practice without necessarily having a set number of sessions or a treatment manual to guide interventions. Treatments studied in this way are obviously closer to what is actually practiced in the real world. However, without random assignment to treatment conditions, the causal influence of the treatment cannot be assessed because the greater latitude for individual treatment differences can confound treatment effects. Despite the fact that both efficacy and effectiveness designs have significant shortcomings, controlled clinical trials have been embraced by the scientific community as representing the state-of-the-art research design for empirically validating treatments.

Efficacy studies of psychotherapy tend to focus on the outcome of different interventions relative to each other.

Less attention is typically paid to the link between the process of therapy and outcome. When a participant in a controlled clinical trial improves after undergoing psychotherapy, it is assumed that the improvement was caused by the specific interventions that were prescribed by a manual and monitored for adherence. However, this assumption relies heavily on the clinician's ability to apply certain techniques without using others and adhere to a particular treatment approach. It is impossible to say what factors were associated with improvement after treatment unless the treatment process itself is studied.

A recent series of studies suggests that the basic assumption of controlled clinical trials of psychotherapy can be questioned. We have shown that even with manualized regimens of psychotherapy, elements are "borrowed" from different treatment approaches and that these common techniques can be among the active ingredients responsible for promoting positive patient change. For example, we previously demonstrated that brief psychodynamic treatments include a diverse set of interventions and that therapists, in addition to applying strategies considered to be psychodynamic in nature, also apply to a considerable degree intervention techniques that are usually associated with cognitive behavior approaches (e.g., examining "faulty thinking" and irrational beliefs) (7). In other words, there was significant overlap in how therapists conducted treatment between theoretical models widely assumed to be distinct intervention strategies. Consistent with our own results, comparison studies conducted by other investigators have also found extensive overlap between psychodynamic-interpersonal therapy and cognitive behavior therapy when conducted by master therapists (8). In one well-studied treatment sample (9), investigators determined that cognitive behavior therapists occasionally used psychodynamic strategies and that it was these techniques that were responsible for promoting patient change (10). The use of techniques not prescribed by the cognitive behavior therapy manual probably escaped the detection of adherence checks; nevertheless, they proved to be significant correlates of patient change.

These provocative results raised serious questions about the conclusions typically drawn from controlled clinical trials. The present study hypothesized 1) that manualized regimens of psychotherapy studied in a controlled clinical trial would overlap considerably in process and technique and that 2) intervention strategies common to both treatments would be responsible for promoting patient change. The NIMH Treatment of Depression Collaborative Research Program was chosen as the best data set to attempt to test these hypotheses. To date, it remains the most carefully conducted and methodologically sound randomized, controlled clinical trial comparing different forms of brief psychological therapy.

A further aim of the present study was to clarify the treatment mechanisms by which patients in the NIMH Treatment of Depression Collaborative Research Program

improved. While both interpersonal psychotherapy and cognitive behavior therapy were shown to be generally effective (11), the specific active ingredients in these therapies have yet to be identified. For interpersonal psychotherapy and cognitive behavior therapy, there have only been a few studies of the treatment process, and these studies have focused on the therapeutic alliance (12) and therapist adherence (13). This study extends our recently reported investigation of treatment processes in the NIMH Treatment of Depression Collaborative Research Program (14). In that study, a comparison of therapist technique and the overall nature of interpersonal transactions in interpersonal psychotherapy and cognitive behavior therapy identified some key differences in therapist activity level and technique. However, there was also a great deal of similarity in therapists' authoritative and supportive stance, in the use of reassurance, and in offering advice or counsel regarding alternative ways of relating to others. The present study attempts to demonstrate whether and how the theories of interpersonal psychotherapy and cognitive behavior therapy were translated into actual practice. Expert therapists from these two theoretical orientations used the Psychotherapy Process Q-Set (15), a rating system designed to provide a standard language for describing therapy process, to generate templates or prototypes of an ideal treatment hour conducted from the perspective of each of these schools of therapy. These prototypes were then compared with Psychotherapy Process Q-Set ratings of the verbatim transcripts, made by independent clinical judges, of interpersonal psychotherapy and cognitive behavior therapy sessions from the NIMH Treatment of Depression Collaborative Research Program. The extent to which the actual treatments conformed to the expert-generated prototypes was calculated quantitatively and then correlated with outcome to determine whether the data support the specific theories of change represented by these two treatments.

Method

Subjects

Participants were outpatients between the ages of 21 and 60 who met Research Diagnostic Criteria (16) for a current episode of major depressive disorder and who scored 14 or higher on a revised 17-item Hamilton Depression Rating Scale (17). Exclusion criteria included certain other psychiatric disorders (bipolar and psychotic disorders), concurrent psychiatric treatment, medical conditions that contraindicated use of imipramine, and the need for immediate intervention (e.g., active suicide potential). Two hundred fifty potential participants met these criteria and were randomly assigned to one of four treatment conditions. The 239 participants who entered treatment were primarily women (70%) and Caucasian (89%). The average age was 35. Seventy-seven participants terminated prematurely (before completing at least 12 sessions of therapy and 15 weeks of treatment). There were no significant differences among the psychosocial treatments in the number of premature terminations (11). The remaining 162 participants are referred to as the "completer group." The mean

number of sessions for subjects in the completer group was 16.2, the maximum was 20.

Treatments

Participants were randomly assigned to one of four brief treatments for depression: interpersonal psychotherapy, cognitive behavior therapy, imipramine plus clinical management as a standard reference treatment, and pill placebo plus clinical management as a double-blind control group. Patients improved in all treatment conditions, and the effect sizes of the psychotherapeutic treatments were found to be consistent with those reported in the literature (11, 18). When differences between the treatments outcomes were observed (19, 20), they were inconsistent, not robust, and appeared only in specific areas. It is generally accepted that the treatments in the NIMH Treatment of Depression Collaborative Research Program were effective and equivalent overall.

Interpersonal psychotherapy and cognitive behavior therapy were carried out in accordance with detailed manuals (21, 22). Ten therapists delivered interpersonal psychotherapy, eight delivered cognitive behavior therapy. The therapists, 72% (N=13) of whom were men, averaged over 11 years of clinical experience. They were all trained in the treatment they provided, and tapes of sessions were reviewed to monitor adherence to treatment protocols (15).

Treatment outcome was measured from a variety of perspectives and with an array of different measures. For example, the clinical evaluator-rated Hamilton Depression Rating Scale (17) and the client-rated Beck Depression Inventory (23) measured depressive symptoms; overall functioning was measured by the clinical evaluator-rated Global Assessment Scale (GAS) (24) and the client-rated HSCL-90 (25). Information on the procedures and methods of the NIMH Treatment of Depression Collaborative Research Program beyond what is presented here is available in Elkin et al. (11, 26).

The Psychotherapy Process Q-Set

This 100-item instrument furnishes a language and rating procedure for the comprehensive description, in clinically relevant terms, of the therapist-patient interaction in a form suitable for quantitative comparison and analysis (15). The Psychotherapy Process Q-Set uses an entire hour as the unit of observation, allowing a greater opportunity to capture events of importance. Clinical judges read the verbatim transcript of a therapy session and then sort the 100 items in the Q-set on a continuum from least characteristic or negatively salient (category 1) to most characteristic or salient (category 9). The middle pile (category 5) is used for items deemed either neutral or irrelevant to the particular hour being rated. Most items have specific instructions that provide examples of the distinction between uncharacteristic and neutral ratings. For example, one item describes the therapist as "distant or aloof" when rated in the characteristic range. However, when rated in the uncharacteristic range, the item indicates that the therapist was "genuinely responsive or affectively involved" (the opposite of "distant or aloof"). Only if the item were irrelevant to the description of the hour would it be placed in the neutral range. The number of cards sorted into each category of the Q-sort (from five at the extremes to 18 in the middle or neutral category) conforms to a normal distribution, requiring judges to make multiple evaluations among items, and thereby avoiding halo effects and response sets (27). Judges rate the frequency, intensity, and estimated importance of each of the 100 statements. A detailed coding manual provides the Q-items and their descriptions as well as operational examples. When rating, judges are asked to take the position of a "generalized other," i.e., an observer who stands midway between patient and therapist and who views the interaction from the outside. In placing each item, judges are instructed to ask themselves: Is this attitude, behavior, or experience clearly present (or absent)? If the evidence is not

compelling, the judge is asked to search for specific evidence of the extent to which it is present or absent. The rating procedure allows judges to form impressions and hypotheses and study the material for confirmation or alternative conceptualizations. The Psychotherapy Process Q-Set was developed pantheoretically to assess therapist actions in different types of therapy, so it is especially useful for comparing the process of different forms of therapy (28, 29). The Psychotherapy Process Q-Set has been used to rate hours of psychodynamic, cognitive behavior, client-centered, Gestalt, and rational-emotive therapies (29). The Psychotherapy Process Q-Set has demonstrated both reliability and strong discriminant validity across a variety of studies and treatment samples (10, 29-31).

Development and Application of Prototypes

Responses by expert therapists of different orientations to a questionnaire form of the Psychotherapy Process Q-Set were used to develop prototypes of ideal treatments. Therapist panels of experts in cognitive behavior therapy (N=10) and interpersonal psychotherapy (N=11) were formed; the panels comprised leading theoreticians and practitioners of each perspective. All of the therapists were highly experienced and internationally recognized for their expertise. Each member of the expert panels had trained therapists in their orientation, and most had published work concerning their approach to psychotherapy. Many of the members of the expert panels were involved in the creation of the treatment modality.

Members of the expert therapist panels were asked simply to rate each of the 100 items of the Q-set questionnaire, on a scale from 1 to 9, according to how characteristic each item was of their understanding of an ideally conducted course of therapy that adheres to the principles of their theoretical perspective. Each questionnaire yielded one score for each of the 100 items of the Q-set. Each expert was also asked to suggest additional items if they felt that the item set was not sufficient to capture adequately the important aspects of their approach to treatment. There were no consistent suggestions for additional items.

Coefficient alpha reliabilities demonstrated that the level of agreement of the expert therapists' ratings was high for both orientations (cognitive behavior therapy=0.95; interpersonal psychotherapy=0.96). The prototypes were created by using a small sample statistical method for studying points of view, sometimes called Q-technique (32-35) (see our earlier article [7] for a detailed description). The responses of the therapist experts in interpersonal psychotherapy and cognitive behavior therapy to the Psychotherapy Process Q-Set were subjected to a principal-components factor analysis. As expected, the factor analysis yielded distinct theoretical orientation factors with eigenvalues above 1.0 after varimax rotation, which together explained 70.9% of the variation in the correlations among the expert therapists. In other words, the interpersonal and cognitive behavior therapists had distinct points of view about an ideal treatment process.

Linear regressions were calculated to determine the contribution of each Q-item to the experts' factors. Factor scores represent the weighted sum of each Q-item. The items with the highest factor scores are most defining of the factor, and the items with the lowest factor scores are least defining of the factor. Table 1 and Table 2 present the 20 items with the highest factor scores for the cognitive behavior and interpersonal therapy techniques, respectively. In the interest of space, only these 20 items are displayed, even though the entire array of factor scores on all 100 items of the Psychotherapy Process Q-Set was included in the Q-prototypes. In other words, each prototype contains all of the 100 items of the Psychotherapy Process Q-Set (not only those appearing in the tables). All 100 items were included in the prototype because it is important that the description of an ideal treatment capture not only those items that are characteristic, but also those that are

TABLE 1. Rank Ordering, by Factor Score, of the 20 Most Characteristic Psychotherapy Process Items in an Ideal Regimen of Cognitive Behavior Therapy

Psychotherapy Process Item ^a	Factor Score ^b
There is discussion of specific activities or tasks for the patient to attempt outside of session	1.93
Discussion centers on cognitive themes, i.e., about ideas or belief systems	1.68
Patient's treatment goals are discussed	1.51
Therapist encourages patient to try new ways of behaving with others	1.49
Therapist actively exerts control over the interaction (e.g., structuring, introducing new topics)	1.45
Therapist adopts supportive stance	1.43
Dialogue has a specific focus	1.38
Therapist asks for more information or elaboration	1.37
Patient's current or recent life situation is emphasized in discussion	1.35
Therapist gives explicit advice and guidance	1.32
Therapist presents an experience or event in a different perspective	1.28
Therapist is confident or self-assured (versus uncertain or defensive)	1.21
Therapist behaves in a teacher-like (didactic) manner	1.17
Patient is committed to work of therapy	1.14
Therapist explains rationale behind technique or approach to treatment	1.13
Patient brings up significant issues and material	1.09
Patient understands the nature of therapy and what is expected	1.08
Patient feels helped	1.06
Therapist accurately perceives the therapeutic process	1.05
Therapist encourages independence of action or opinion in patient	1.02

^a From the Psychotherapy Process Q-Set, 100 statements describing the therapist-patient interaction during the treatment hour designed to provide a standard language for the therapeutic process and allow quantitative comparison and analysis of different forms of psychotherapy. Each statement is rated on a 9-point scale (1=item least characteristic or negatively salient, 9=item most characteristic or salient).

^b Derived from ratings of the 100 Psychotherapy Process Q-Set items by 10 expert cognitive behavior therapists.

uncharacteristic and not relevant. That is, a prototype should reflect what is absent from an ideal treatment as well as what is prominent. This is important to consider when evaluating the face validity of the prototypes, since items contributing significantly to a prototype may not be listed among the 20 items with the highest factor loadings. For example, one psychotherapy process item that describes the therapist as challenging the patient's view or core beliefs received a high factor loading (0.66), indicating that the experts felt it was an important component of cognitive behavior therapy. However, it does not appear in Table 1 because it was not among the 20 items with the highest factor loadings. Because the prototypes consist of all 100 items of the Psychotherapy Process Q-Set, they account for the relative contribution of all items in the item set.

As seen in Table 1 and Table 2, both the cognitive behavior therapy and the interpersonal psychotherapy experts rated six of the same items as being among the 20 most defining aspects of their respective treatments. These items have different factor loadings, so they each contribute to the prototype to a varying degree. For example, the item "therapist adopts supportive stance" has a higher factor loading on the cognitive behavior therapy scale, indicating that the experts judged the item to be more characteristic of ideal cognitive behavior therapy than ideal interpersonal psychotherapy.

TABLE 2. Rank Ordering, by Factor Score, of the 20 Most Characteristic Psychotherapy Process Items in an Ideal Regimen of Interpersonal Psychotherapy

Psychotherapy Process Item ^a	Factor Score ^b
Patient's interpersonal relationships are a major theme	2.22
Therapist emphasizes patient's feelings in order to help him/her experience them more deeply	1.65
Patient talks of feelings about being close to or needing someone	1.62
Love or romantic relationships are a topic of discussion	1.58
Therapist explains rationale behind technique or approach to treatment	1.55
Dialogue has a specific focus	1.39
Termination of therapy is discussed	1.32
Therapist is directly reassuring	1.29
Therapist draws attention to patient's nonverbal behavior	1.27
Therapist makes interpretations referring to actual people in patient's life	1.25
There is discussion of body functions, physical symptoms, or health	1.20
Therapist's remarks are aimed at facilitating patient's speech	1.19
Therapist clarifies, restates, or rephrases patient's communication	1.15
Therapist comments on changes in patient's mood or affect	1.13
Patient's treatment goals are discussed	1.10
Patient's current or recent life situation is emphasized in discussion	1.09
Therapist adopts supportive stance	1.09
Patient experiences discomforting or troublesome (painful) affect	1.05
There is discussion of scheduling of hours or fees	0.94
Therapist accurately perceives the therapeutic process	0.91

^a From the Psychotherapy Process Q-Set, 100 statements describing the therapist-patient interaction during the treatment hour designed to provide a standard language for the therapeutic process and allow quantitative comparison and analysis of different forms of psychotherapy. Each statement is rated on a 9-point scale (1=item least characteristic or negatively salient, 9=item most characteristic or salient).

^b Derived from ratings of the 100 Psychotherapy Process Q-Set items by 11 expert interpersonal psychotherapists.

Q-Sort Ratings of Actual Therapy Administered

A pool of nine research-oriented psychologists and master's-level graduate students in clinical psychology, all trained in the application of the Q-technique, completed Q-sort ratings of interpersonal psychotherapy and cognitive behavior sessions from the NIMH Treatment of Depression Collaborative Research Program. The raters represented a range of theoretical perspectives, although most were eclectic in their clinical orientations. Two verbatim transcripts, a session early in treatment (session 4) and a session late in treatment (session 12) were selected for each patient in the interpersonal psychotherapy and cognitive behavior therapy groups. When transcripts of these sessions were not available, the next closest session was selected. The length of treatments for these patients ranged from 12 to 20 sessions, with an average length of 16.2 sessions. Transcripts were only available from the archive for two of the three sites that participated in the NIMH Treatment of Depression Collaborative Research Program. Therefore, transcripts were available for only 35 of the 47 participants who received interpersonal psychotherapy and 29 of the 37 participants who received cognitive behavior therapy. All of the transcripts (N=128) were randomly assigned, and independent ratings were completed by at least two judges who were blind to treatment type and session number. Average interrater reliability for Q-sorts achieved 0.82. There were no significant differences in reliability between treatment modalities. When agreement was

below $r=0.50$, the author of the Psychotherapy Process Q-Set, a licensed and practicing clinical psychologist (E.E.J.), was added as a third rater. Periodic calibration meetings were conducted to correct rater drift. Statistical analyses comparing Q-sort ratings from sessions 4 and 12 revealed little change in process over time. The independent Q-sort ratings (sessions 4 and 12) of the judges for each transcript were therefore composited as they have been in previous studies that used the Psychotherapy Process Q-Set. The composite Q-sort ratings were averaged across both sessions to obtain one score per Q-item for each patient.

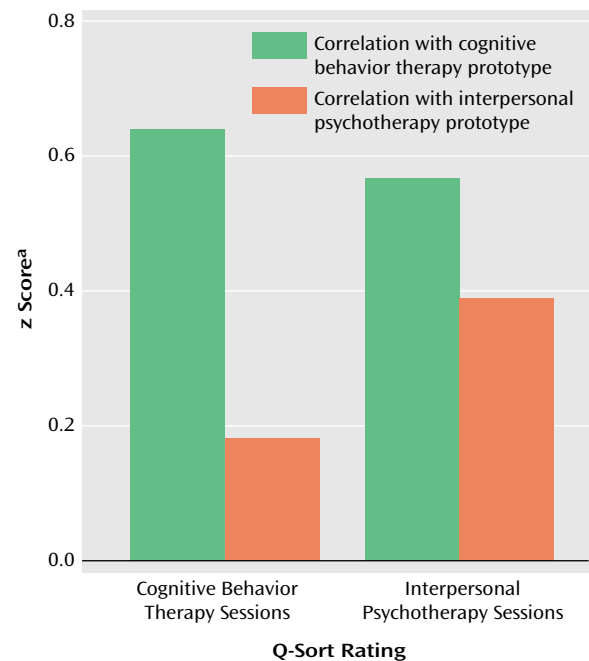
Analyses

The Q-prototypes of ideal interpersonal psychotherapy and cognitive behavior therapy were then correlated with the Q-sort ratings of the actual psychotherapy sessions from the NIMH Treatment of Depression Collaborative Research Program in order to determine how close the treatments were to the ideal prototypes. That is, for each patient the composite Q-rating of each of the 100 items was correlated with the factor score for that item from the prototype. These correlations with the Q-prototypes then were transformed into z scores by using the Fisher r-to-z transformation. The z scores represent the degree to which an hour of psychotherapy was correlated with the Q-prototypes. The z scores were then correlated with outcome to determine whether those aspects of the therapy process that in theory should promote patient change were in fact responsible for predicting positive outcome. Partial correlations of the z scores and the outcome measures were calculated in order to control for pretreatment scores.

Results

Figure 1 depicts the mean correlations between the Q-prototypes and the Q-sort ratings of the cognitive behavior therapy and interpersonal psychotherapy sessions. The same pattern of correlations emerged for both the cognitive behavior therapy and interpersonal psychotherapy groups. The Q-sort ratings of the session transcripts had significantly higher correlations with the cognitive behavior therapy prototype than with the interpersonal psychotherapy prototype for both the cognitive behavior therapy group ($t=11.92$, $df=28$, $p<0.001$) and the interpersonal psychotherapy group ($t=5.39$, $df=34$, $p<0.001$). In other words, both forms of therapy administered as part of the NIMH Treatment of Depression Collaborative Research Program adhered more to the cognitive behavior therapy prototype than the interpersonal therapy prototype. In the group assigned to cognitive behavior therapy, there was a high correlation between the Q-sort ratings of the session transcripts and the cognitive behavior therapy prototype (z score mean=0.64, $SD=0.28$) and only a small correlation between the Q-sort ratings and the interpersonal psychotherapy prototype (z score mean=0.18, $SD=0.10$). In the group assigned to interpersonal psychotherapy, there again was a high correlation between the Q-sort ratings of the session transcripts and the cognitive behavior therapy prototype (z score mean=0.57, $SD=0.24$) and a moderate correlation between the Q-sort ratings and the interpersonal psychotherapy prototype (z score mean=0.39, $SD=0.11$). The correlation between the Q-sort ratings of the session transcripts and the interpersonal psychotherapy pro-

FIGURE 1. Correlations Between Ideal Psychotherapy Regimens (Q-Prototypes) and Actual Therapy Administered (Q-Sort Ratings) for Patients Randomly Assigned to Receive Cognitive Behavior Therapy (N=29) or Interpersonal Psychotherapy (N=35) as Part of the NIMH Treatment of Depression Collaborative Research Program



^a Correlations were transformed from Pearson correlation coefficients into z scores by using the Fisher r-to-z transformation before averaging. The z scores represent the degree to which an hour of psychotherapy was correlated with an ideal regimen of either cognitive behavior therapy or interpersonal psychotherapy.

tototype was significantly higher in the interpersonal psychotherapy group than in the cognitive behavior therapy group ($t=-8.24$, $df=62$, $p<0.001$). However, there was no significant difference between the correlations of the Q-sort ratings of the session transcripts and the cognitive behavior therapy prototype across the two treatment forms. The only significant difference between the two treatments was in adherence to the interpersonal prototype.

Table 3 displays the partial correlation coefficients, with pretreatment scores controlled, of the outcome measures and adherence to the Q-prototypes for both interpersonal psychotherapy and cognitive behavior therapy. Positive correlations reflect a favorable association with outcome. For the interpersonal psychotherapy treatment group (N=35), adherence to the cognitive behavior therapy prototype was associated significantly with positive outcome on five of the six outcome measures. Adherence to the interpersonal psychotherapy prototype was associated significantly with positive outcome on three of the six outcome measures. For the cognitive behavior therapy group (N=29), adherence to the cognitive behavior therapy prototype was associated significantly with positive outcome on all six of the outcome measures. Adherence to the interpersonal prototype was associated significantly with positive outcome on two of the six outcome measures.

TABLE 3. Correlations Between Outcome and Adherence to Ideal Psychotherapy Regimens (Q-Prototypes) for Patients Randomly Assigned to Receive Cognitive Behavior Therapy (N=29) or Interpersonal Psychotherapy (N=35) as Part of the NIMH Treatment of Depression Collaborative Research Program

Treatment Group and Outcome Measure	Correlation (r^a)	
	Adherence to Interpersonal Psychotherapy Prototype	Adherence to Cognitive Behavior Therapy Prototype
Interpersonal psychotherapy group		
23-item Hamilton Depression Rating Scale	0.32*	0.45**
Global score on the Modified Social Adjustment Scale	0.27	0.51**
HSCL-90	0.29*	0.36*
Beck Depression Inventory	0.37*	0.52**
Dysfunctional Attitudes Scale	0.24	0.13
Global Assessment Scale	0.24	0.47**
Cognitive behavior therapy group		
23-item Hamilton Depression Rating Scale	0.26	0.49**
Global score on the Modified Social Adjustment Scale	0.11	0.38*
HSCL-90	0.30	0.36*
Beck Depression Inventory	0.41*	0.53**
Dysfunctional Attitudes Scale	0.48**	0.46**
Global Assessment Scale	0.22	0.45*

^a Positive correlations reflect a favorable association with outcome. All Pearson correlations are partial correlations controlling for patient pretreatment scores.

* $p < 0.05$. ** $p < 0.01$.

In summary, adherence to the cognitive behavior therapy prototype correlated with positive outcome in a consistent and robust manner for both the interpersonal psychotherapy and the cognitive behavior therapy groups. The correlations were particularly strong ($r=0.36$ – 0.53) relative to other process-outcome linkages found in the literature. Adherence to the interpersonal psychotherapy prototype also predicted positive outcome for both the interpersonal psychotherapy and the cognitive behavior therapy groups but in a less consistent and robust manner ($r=0.29$ – 0.48). It is unlikely that differences in the variability of prototype scores across treatment conditions contributed to this pattern of correlations with outcome. Levene tests for equality of variances revealed no significant differences in the variance of interpersonal psychotherapy prototype adherence scores for the interpersonal psychotherapy (variance=0.01) and cognitive behavior therapy (variance=0.01) groups ($F=0.05$, $df=1, 62$, $p=0.83$) or in the variance of cognitive behavior therapy prototype adherence scores for the interpersonal psychotherapy (variance=0.06) and cognitive behavior therapy (variance=0.08) groups ($F=2.05$, $df=1, 62$, $p=0.16$).

Discussion

Brand Names of Therapy Can Be Misleading

A remarkable finding of this study was that the nature of the process fostered by therapists administering interper-

sonal psychotherapy in the NIMH Treatment of Depression Collaborative Research Program was so similar to that fostered by therapists administering cognitive behavior therapy. In fact, interpersonal psychotherapy conformed more to the expert-generated prototype of cognitive behavior therapy than it did the prototype of interpersonal psychotherapy. Interpersonal psychotherapy included a more diverse set of therapeutic processes, including significant aspects of the kind found in cognitive behavior therapy. Cognitive behavior therapy, on the other hand, was more pure in its emphasis relative to its prototype. An earlier study of the NIMH Treatment of Depression Collaborative Research Program showed that the two treatments could be reliably distinguished on a rating scale meant to assess features of therapist activity specific to the respective treatments (13). That study also showed that both treatments contained an equivalent degree of what has been termed common "facilitative conditions." Our own results suggest that the two treatments were by no means pure and distinct despite adherence checks. The overall process of interpersonal psychotherapy was startlingly similar to cognitive behavior therapy in actual practice, challenging the presumption of large differences across these two brand names of therapy.

Since the interpersonal psychotherapy and cognitive behavior therapy processes can be quite similar, it is not at all surprising that the two forms of therapy could achieve such strikingly similar outcomes in the Treatment of Depression Collaborative Research Program (11). Because the treatments were manualized, it was generally concluded, without having examined the relationship between process and outcome, that the techniques of interpersonal psychotherapy and cognitive behavior therapy were at the same time very different but equally effective. Our results suggest that both interpersonal psychotherapy and cognitive behavior therapy were characterized by a significant degree of overlap, particularly in what experts described as cognitive behavior therapy processes. It was these shared cognitive behavior processes that most robustly predicted positive outcome in both the interpersonal psychotherapy and the cognitive behavior therapy groups. Nevertheless, the interpersonal psychotherapy process also contributed to positive outcome in both treatment modalities. Similarities in process across treatments help explain the lack of differential outcomes frequently observed in trials of psychotherapy. What are often referred to as nonspecific effects in the literature may actually represent the specific effects of shared processes.

One possible explanation for how the process of treatments with different brand names and different prescribed techniques can be so similar lies in the theoretical language used by proponents of different forms of psychotherapy. Clinicians from different orientations may use very different terminology to describe psychological constructs and processes that are actually quite similar. Because of these differences in language and ways of con-

ceptualizing, treatments may appear quite different on the surface. However, once we move beyond theory-specific language and examine what actually occurs during treatment sessions, these two treatments in fact share many common features.

Another likely explanation for the similarity between the two forms of therapy, at least as they were actually conducted, concerns the contribution of the patient to the treatment process. Previous reports on the NIMH Treatment of Depression Collaborative Research Program documenting distinctions between interpersonal psychotherapy and cognitive behavior therapy focused on therapist technique (13). Indeed, most observations of distinctions between treatments tend to focus only on the therapists' interventions and adherence to prescribed techniques in isolation from the larger context in which they occur. It is usually assumed that influence flows in one direction, i.e., therapist interventions cause patient change. However, the nature of the therapeutic process is inherently interactional. Patient and therapist mutually affect one another and the nature of what occurs in treatment sessions. Every clinician knows that beyond patients' specific diagnoses, their personal qualities, affect states, and style of relating to others enormously influence what may or may not occur in a treatment session. Our research suggests that the patient's contribution to the process in interpersonal psychotherapy and cognitive behavior therapy was remarkably similar (14). That is, patients in both brief treatment forms fostered a similar overall process. The overlap in overall process of interpersonal psychotherapy and cognitive behavior therapy is likely the result of strong similarities in patients' contribution to the process coupled with important areas of overlap in therapists' stance toward patients.

Contributing to the appearance of large differences across these treatments is a focus on content rather than process. The content of what the cognitive behavior therapist takes up with his patient (dysfunctional attitudes or irrational beliefs) is often quite different from the content on which the interpersonal psychotherapy therapist might focus (e.g., disruptions in personal relationships). However, the nature of the interaction between therapist and patient might be quite similar. For example, a cognitive behavior therapist may coach or educate a patient about more effective ways of thinking (e.g., to not "catastrophize"), while an interpersonal psychotherapy therapist might teach a patient how to manage relationships with co-workers more effectively. The content is different, but the nature of the interaction is similar. In both treatments, the therapist assumes an authoritative role and coaches patients to think or conduct themselves differently and encourages them to test out these new ways of thinking and behavior in everyday life. A narrow focus on the different content of these manualized treatments makes it easy to overlook the high degree of correspondence in process. The Psychotherapy Process Q-Set, which was used to de-

scribe treatment processes in the current study, largely avoids the use of theory-specific language and focuses on process rather than content (30). Although the Psychotherapy Process Q-Set has strong discriminant validity (10), these features of the measure may have helped to reveal the similarities across the study treatments.

Correlation Between Process and Outcome in Brief Therapy

Our previous study of the process of psychotherapy in the NIMH Treatment of Depression Collaborative Research Program (14), which used individual Q-items as the unit of analysis, suggested more specifically how these treatments helped patients with depressive symptoms. The majority of the individual Q-items that predicted positive outcome in both interpersonal psychotherapy and cognitive behavior therapy described the patient's contribution to the process rather than the therapists' actions and interventions. The Q-items associated with change across both interpersonal psychotherapy and cognitive behavior therapy described patients' experience of a positive sense of self and an unambivalent, deferential, even idealized view of their therapist. This constellation of Q-items captured the presence of a positive attachment to a benevolent, supportive, and reassuring authority figure that apparently facilitated improvement in both forms of brief treatment. Therapists for both the interpersonal psychotherapy and the cognitive behavior therapy groups relied heavily on interpersonal influence; they offered advice and guidance, suggested patients change their behavior, reassured patients that they would feel better soon, and did not, as a rule, discuss features of their interaction with the patient. In spite of their different theoretical orientations, our results suggest that the two short-term treatments shared these common characteristics, essentially supportive in nature, and it is these features that are associated with positive outcome at termination. Taken together, the results of our two studies suggest that what was shared between the two forms of therapy in the NIMH Treatment of Depression Collaborative Research Program was more salient and defining of the treatments than what was different.

Controlled Clinical Trials and Empirical Validation of Treatments

The NIMH Treatment of Depression Collaborative Research Program represents a state-of-the-art controlled clinical trial of psychotherapy in which therapists were trained in treatments that were prescribed by manuals and then checked for adherence (13). The controlled clinical trial paradigm enforces all of these experimental controls on the therapy process in an effort to maintain pure and distinct treatments. With such controls in place, improvement during the course of treatment is presumed to be the result of the techniques described in the treatment manual. The results of this study and others suggest that

this can be a flawed assumption. It may in fact not be possible to control, to a sufficient degree, how treatments are actually conducted. If it is impossible to fully control the process of psychotherapy, then a basic premise of a randomized, controlled clinical trial is not met. Such studies can only reveal whether a treatment is effective. It is not possible to draw conclusions about what promotes patient change, let alone establish the validity of a treatment's underlying theory of therapy.

This has, of course, implications for the movement to empirically validate treatments. The primary means for establishing the validity of therapies is the comparative clinical trial. However, if the nature of the treatment cannot be controlled despite the use of manuals and adherence checks, the usefulness of the clinical trials model for validating particular psychological therapies must be questioned. The controlled clinical trial method was initially designed by medical science for use in studies of medications. A physician administers a specific medication knowing it is the only medication being administered in order to compare the results to a placebo condition. Unlike medication, psychotherapy cannot be administered in such pure form, and adherence is much more difficult to measure. The controlled clinical trial method is effective in investigating medical interventions for comparing psychotherapy to pharmacotherapy or their combination. It is limited, however, when imposed on the studies of psychotherapy alone. The reality of clinical practice is that therapists do not exclusively rely on certain techniques or adhere strictly to a particular treatment as manualized. In actuality, therapists often consider disparate intervention strategies to be compatible with their own theoretical orientation.

Limitations

Our findings must be considered in the context of the limitations of our study design and methods. As in all studies, it is important to consider issues of replicability and generalizability. Because this study relied on a relatively small number of subjects, it is possible that some bias exists in the sample despite random assignment to treatment conditions. It is possible that with a different set of therapists and a different set of patients, less overlap may have been found between the process components of interpersonal psychotherapy and cognitive behavior therapy. Future process studies with other treatment groups are needed to address the generalizability of our findings.

Future Directions

The external validity that is sacrificed in controlled clinical trials of psychotherapy does not seem warranted if the primary goal of establishing causal relationships cannot be attained because of the shortcomings we have described. Our findings raise questions about the usefulness of clinical trials for establishing the validity of psychological treatments. Alternate models are needed to establish which

therapies might earn an imprimatur as an "empirically validated treatment." Many prominent psychotherapy researchers have called for a shift in focus from the study of comparative effectiveness of treatments to the empirical validation of change processes (5, 36, 37). Psychotherapy research would profit from the study of change processes as they occur naturalistically, rather than focusing on the empirical validation of brand names of therapy. Comparative treatment studies emphasize treatment procedures, and it is important to include more fully the role of the individual patient in a treatment's outcome. It is assumed that if the patient sample is homogeneous in terms of diagnosis (e.g., major depression in the NIMH Treatment of Depression Collaborative Research Program), the most important source of patient variance has been controlled. Our studies have demonstrated that patient in-session characteristics were far more important correlates of outcome than treatment type. This conclusion is consistent with the long-standing finding (38, 39) that patient qualities, including personality characteristics, are far more robust predictors of outcome than are treatment techniques. Naturalistic studies that assess relevant patient characteristics, and their interaction with therapists and intervention strategies, would be a fruitful complement to randomized clinical trials in therapy research.

Received Dec. 8, 2000; revision received June 18, 2001; accepted Dec. 18, 2001. From Massachusetts General Hospital, Harvard Medical School; and the University of California at Berkeley. Address reprint requests to Dr. Ablon, Massachusetts General Hospital, WAC 812, 15 Parkman St., Boston, MA 02114; sablon@partners.org (e-mail).

The authors thank the investigators and oversight committee of the NIMH Treatment of Depression Collaborative Research Program for making data available for further analyses and providing transcripts of the treatment sessions.

Supported by an NIH National Research Service Award (MH-11356); the American Psychological Association Scott Mesh Honorary Scholarship; grants from the Humanities Department, University of California, Berkeley; and funding from the Luria Foundation.

References

1. Jacobson NS, Christensen A: Studying the effectiveness of psychotherapy: how well can clinical trials do the job? *Am Psychol* 1996; 51:1031-1039
2. Seligman MEP: The effectiveness of psychotherapy. *Am Psychol* 1995; 50:965-974
3. Seligman M: Science as an ally of practice. *Am Psychol* 1996; 51:1072-1079
4. Hollon SD: The efficacy and effectiveness of psychotherapy relative to medications. *Am Psychol* 1996; 51:1025-1030
5. Goldfried MR, Wolfe BE: Psychotherapy practice and research: repairing a strained alliance. *Am Psychol* 1996; 51:1007-1016
6. Goldfried MR, Wolfe BE: Toward a more clinically valid approach to therapy research. *J Consult Clin Psychol* 1998; 66: 143-150
7. Ablon JS, Jones EE: How expert clinicians' prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychother Res* 1998; 8:71-83
8. Goldfried M, Raue PJ, Castonguay LG: The therapeutic focus in significant sessions of master therapists: a comparison of cog-

- nitive-behavioral and psychodynamic-interpersonal interventions. *J Consult Clin Psychol* 1998; 66:803–810
9. Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove MW, Tuason VB: Cognitive therapy and pharmacotherapy for depression: singly and in combination. *Arch Gen Psychiatry* 1992; 49:774–781
 10. Jones EE, Pulos SM: Comparing the process in psychodynamic and cognitive-behavioral therapies. *J Consult Clin Psychol* 1993; 61:306–316
 11. Elkin I, Shea T, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB: National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry* 1989; 46:971–982
 12. 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. *J Consult Clin Psychol* 1996; 64:532–539
 13. Hill CE, O'Grady KE, Elkin I: Applying the Collaborative Study Psychotherapy Rating Scale to rate adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *J Consult Clin Psychol* 1992; 60:73–79
 14. Ablon JS, Jones EE: Psychotherapy process in the NIMH Collaborative Study. *J Consult Clin Psychol* 1999; 67:64–75
 15. Jones EE: *Therapeutic Action: A Guide to Psychoanalytic Therapy*. Northvale, NJ, Jason Aronson, 2000
 16. Spitzer RL, Endicott J, Robins E: *Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders*, 3rd ed. New York, New York State Psychiatric Institute, Biometrics Research, 1978
 17. Hamilton M: Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967; 6:278–296
 18. Elkin I: The NIMH Treatment of Depression Collaborative Research Program: where we began and where we are now, in *Handbook of Psychotherapy and Behavior Change*, 4th ed. Edited by Bergin AE, Garfield SL. New York, John Wiley & Sons, 1994, pp 114–135
 19. Elkin I, Gibbons R, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, Hedeker D: Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol* 1995; 63:841–847
 20. Ogles BM, Lambert MJ, Sawyer JD: Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *J Consult Clin Psychol* 1995; 63:321–326
 21. Beck AT, Rush AJ, Shaw BF, Emery G: *Cognitive Therapy of Depression*. New York, Guilford, 1979
 22. Klerman GL, Weissman MM, Rounsaville B, Chevron ES: *Interpersonal Psychotherapy for Depression*. New York, Basic Books, 1984
 23. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J: An inventory for measuring depression. *Arch Gen Psychiatry* 1961; 4: 561–571
 24. Endicott J, Spitzer RL, Fleiss JL, Cohen J: The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 1976; 33:766–771
 25. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L: The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 1974; 19:1–15
 26. Elkin I, Parloff MB, Hadley SW, Autry JH: National Institute of Mental Health Treatment of Depression Collaborative Research Program: background and research plan. *Arch Gen Psychiatry* 1985; 42:305–316
 27. Block J: *The Q-Sort Method in Personality Assessment and Psychiatric Research*. Springfield, Ill, Charles C Thomas, 1961
 28. Lambert MJ, Hill CE: Assessing psychotherapy outcomes and processes, in *Handbook of Psychotherapy and Behavior Change*, 4th ed. Edited by Bergin AE, Garfield SL. New York, John Wiley & Sons, 1994, pp 72–113
 29. Jones EE, Cumming JD, Horowitz MJ: Another look at the non-specific hypothesis of therapeutic effectiveness. *J Consult Clin Psychol* 1988; 56:48–55
 30. Jones EE, Hall S, Parke LA: The process of change: the Berkeley Psychotherapy Research Group, in *Psychotherapy Research: An International Review of Programmatic Studies*. Edited by Beutler L, Crago M. Washington, DC, American Psychological Association, 1991, pp 98–107
 31. Jones EE, Krupnick JH, Kerig PK: Some gender effects in a brief psychotherapy. *Psychotherapy* 1987; 24:336–352
 32. Cattell RB: *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist*. New York, Harper, 1952
 33. Stephenson W: *The Study of Behavior: Q-Technique and Its Methodology*. Chicago, University of Chicago Press, 1953
 34. Nunnally JC: *Psychometric Theory*. New York, McGraw-Hill, 1978
 35. McKeown B, Thomas D: *Q Methodology*. Newbury Park, Calif, Sage Publications, 1988
 36. Howard KI, Moras K, Brill PB, Martinovich Z, Lutz W: Evaluation of psychotherapy: efficacy, effectiveness, and patient progress. *Am Psychol* 1996; 51:1059–1064
 37. Garfield SL: Some comments on empirically supported treatments. *J Consult Clin Psychol* 1998; 66:121–125
 38. Blatt SJ, Quinlan DM, Pilkonis PA, Shea TM: Impact of perfectionism and need for approval on the brief treatment of depression: the National Institute of Mental Health Treatment of Depression Collaborative Research Program revisited. *J Consult Clin Psychol* 1995; 63:125–132
 39. Garfield SL: *Research on Client Variables in Psychotherapy*, 2nd ed. New York, John Wiley & Sons, 1978