

Psychotherapy Adherence of Therapists Treating HIV-Positive Patients With Depressive Symptoms

John C. Markowitz, M.D.
Lisa A. Spielman, Ph.D.
Polly A. Scarvalone, Ph.D.
Samuel W. Perry, M.D.[†]

Adherence monitoring, a technology to specify research psychotherapies, was used in the NIMH Treatment of Depression Collaborative Research Program (TDCRP). The authors present adherence data from a similar randomized treatment trial of 56 depressed HIV-positive patients, comparing 16-week interventions with cognitive-behavioral therapy, interpersonal psychotherapy, and supportive psychotherapy alone or with imipramine. Therapists were certified in manualized treatments. Blind independent raters rated randomly selected taped sessions on an adaptation of the NIMH scale, yielding adherence scores for interventions and for therapist "facilitative conditions" (FC). All therapists were rated adherent. Interrater reliability was 0.89–0.99. The scale discriminated among the four treatments ($P < 0.0001$), with each scoring highest on its own scale. FC, which might measure therapist competence independent of treatment technique, varied by intervention but did not predict treatment outcome. This study demonstrates the ability to reliably train adherence monitors and therapists able to deliver specified treatments. Its adherence findings provide the first replication of those from the landmark NIMH TDCRP study.

(The Journal of Psychotherapy Practice and Research 2000; 9:75–80)

Monitoring therapist adherence to manualized psychotherapies is essential for research purposes. Adherence monitoring demonstrates the specificity and purity of the treatments delivered. It helps to answer the century-old question of what takes place in psychotherapists' offices. Although far more labor intensive, it is analogous to measuring serum levels of medication in a pharmacotherapy trial. Reliably trained observers blind to type of therapy use a structured instrument to rate the degree to which psychotherapists use particular interventions, and avoid using others, during recorded treatment sessions. Adherence monitoring represents the state of the art of current psychotherapy research; yet because of the effort and expense it involves, this still-developing technology remains relatively little used.

Adherence monitoring measures fidelity to a manualized technique rather than the competence,^{1,2} much less excellence, of the therapist. Adherence to the ther-

[†]Deceased.

Received October 18, 1999; revised December 13, 1999; accepted December 16, 1999. From the Department of Psychiatry, Cornell University Medical College, and the New School-Beth Israel Center for Clinical Training and Research, New York, New York. Address correspondence to Dr. Markowitz, 525 East 68th Street, Room 1322, New York, NY 10021; e-mail: jmarko@mail.med.cornell.edu

Copyright © 2000 American Psychiatric Association

apeutic model provides a basic measure of the type of treatment the therapist has delivered, the bedrock on which other therapist factors, such as competence, can be measured. Adherence thus can offer the limited but important assurance that subjects receiving a treatment (e.g., interpersonal psychotherapy [IPT], cognitive-behavioral therapy [CBT]) are in fact getting the treatment advertised. This represents a significant improvement over studies where one has no idea what occurred in therapists' offices.

Why report what might be considered the background validation of a psychotherapy trial? An important reason is that no benchmark measure exists for psychotherapy adherence. Different instruments have been developed for differing therapies on different occasions. The number of published studies of psychotherapy adherence in outcome trials is small.²⁻¹⁰ When such instruments have appeared in treatment studies, they have generally been used once, without replication. Yet as Klein said of psychotherapy research in another context, "Replications are the soul of scientific advance."¹¹

Perhaps the most highly developed measure, and a key example of adherence monitoring, is the Collaborative Study Psychotherapy Rating Scale (CSPRS) developed by Hollon¹² for the landmark, multisite National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP).^{13,14} Hill et al.,¹⁵ using an elegant balanced incomplete block design, found that blinded raters using the CSPRS could differentiate among therapists delivering IPT, CBT, and clinical management (CM, accompanying medication or placebo) to 180 depressed subjects. We now report an initial attempt to replicate the results of this key study. We used an adaptation of the CSPRS to measure psychotherapy adherence in a treatment study of patients seropositive for the human immunodeficiency virus (HIV) who were suffering from depressive symptoms.

We hypothesized that 1) the minor changes required to adjust the CSPRS to the needs of the current study would not compromise its effectiveness in discriminating among therapies. Another hypothesis, crucial to the validity of outcome data from this research, was that 2) therapists would generally adhere to their prescribed therapies, and hence that each treatment cell would score highest on its own CSPRS subscale. Because the therapeutic alliance has been frequently correlated with treatment outcome,¹⁶ we also wondered

whether 3) facilitative conditions (FC), a modality-nonspecific CSPRS subscale of therapist warmth and other supportive behaviors, might provide a predictive measure mediating treatment outcome for the study.

METHODS

Treatments and Study Participants

The design of the treatment study, begun by the late Samuel Perry, M.D., is based on the TDCRP. It compared the efficacy of four randomly assigned, 16-week interventions in treating depressive symptoms of HIV-positive patients: CBT;¹⁷ IPT;¹⁸⁻²⁰ supportive psychotherapy (SP) alone; and supportive therapy with imipramine (SWI). Supportive psychotherapy was substituted for clinical management plus placebo in the TDCRP because we did not feel comfortable offering a placebo condition to this treatment population. Thus the design randomized subjects to one of three psychotherapies (CBT, IPT, SP) or to a combined condition, SWI. To validate the appropriate delivery of these treatments, we modified the CSPRS and trained raters to independently grade audiotapes of sessions.

Treatments were delineated in manuals for each intervention. The CBT and IPT manuals modified the standard delivery of these treatments of depression to address the particular concerns of depressed HIV-positive patients.¹⁹ The supportive therapy used in SP and SWI, defined as non-IPT and non-CBT, was akin to the client-centered therapy of Carl Rogers²¹ with additional psychoeducational components about depression and HIV. Although SP therapy may have been somewhat hindered by proscription of interpersonal and cognitive techniques, it was by no means a non-treatment, and indeed was associated with lowered symptoms on outcome.²² The SWI condition combined this approach with the medical model derived from the TDCRP clinical management manual of Fawcett et al.,²³ which stresses a biochemical rationale and assesses medication side effects. Imipramine was titrated upward on a dosing schedule to a peak of 250 mg daily unless limited by side effects. CBT and IPT involved 16 sessions of 50 minutes within a 17-week envelope. The SWI and SP conditions ranged between 8 and 16 sessions, as needed, of 30 to 50 minutes' duration.

Therapists were psychiatrists, psychologists, social workers, and psychiatric nurses experienced in working

with depressed patients and knowledgeable about HIV.^{19,22} They were trained during a pilot phase by experts in each modality and monitored throughout the study to avoid drift in technique. Monitoring included individual supervision of audiotaped sessions and ratings of randomly selected sessions by independent and reliably trained adherence monitors. Data were analyzed on 2 CBT, 4 IPT, 8 SP, and 4 SWI therapists, who treated 12, 14, 15, and 15 depressed HIV subjects, respectively.

Subjects were HIV-positive depressed outpatients, mainly gay and bisexual white males, who were not too acutely medically ill to begin treatment and who had a score of ≥ 15 on the 24-item version of the Hamilton Rating Scale for Depression (Ham-D)²⁴ and a clinical diagnosis of probable DSM-III-R mood disorder.²²

Measures

Outcome measures included the 24-item version of the observer-rated Ham-D and the self-report Beck Depression Inventory (BDI).²⁵ An independent rater scored the Ham-D at intake, at midtreatment, and just before termination of the 16-week course. Therapists also rated the Ham-D at alternate sessions. Subjects completed the BDI before the start of each session.

The CSPRS, developed for the TDCRP, is the sixth version evolving from a lengthy psychometric testing process by Hollon.¹² Its 96 items, rated on a 7-point, Likert-type scale, have defined anchor points (1 = "not at all" to 7 = "extensively"). The CSPRS includes subscales for CBT (28 items), IPT (28 items), and CM (20 items); two modality-tangential scales (4 CBT, 4 IPT); and modality-nonspecific scales titled "facilitative conditions" (FC, 8 items; measures of nonspecific therapist qualities such as empathy and warmth) and "explicit directiveness" (ED, 4 items; how much the therapist guides the subject). In adapting the scale for this study, we expanded the 96 items of the CSPRS to 104 items to make the "CSPRS-HIV." The eight new items reflected the addition of SP therapy (e.g., "The therapist echoed the patient's concerns as a means of facilitating the patient's discussion of those concerns"). They were incorporated without the involved psychometric testing that heralded the original instrument.¹²

Raters

Independent raters were four predoctoral psychology graduate students who developed reliability after

approximately 40 hours of training. Training, conducted initially by the first author (J.C.M.) and subsequently by the two initial trainees (including P.A.S.), included reading the CSPRS manual¹² and rating and discussing up to 16 pilot tapes, randomly chosen but covering all interventions, with ongoing supervision and feedback. Because the design of the two studies differed, and because raters used the 104-item adaptation (CSPRS-HIV) of the 96-item CSPRS, no attempt was made to calibrate their ratings with tapes from the TDCRP. Raters met periodically with the trainer to discuss tapes and prevent drift.

Two complete audiotaped sessions were rated per therapist-subject dyad: an early session, randomly chosen from sessions 3-6, and a late one, from sessions in weeks 9-12 of treatment. (When on occasion the selected tape was unavailable or inaudible, an adjacent session was assessed.) This contrasts with the four tapes per patient (sessions #1; 4; 7 or 8; and 14, 15, or penultimate) assessed by Hill *et al.*¹⁵ We also made no attempt to replicate the design of Hill and colleagues in randomly assigning raters to tapes. Instead, in a more naturalistic design, raters were alternately assigned tapes as they became available from the treatment study. Study raters were blind to treatment modality and to early versus late session status.

Statistical Analysis

Data analysis employed intraclass correlations²⁶ to assess interrater reliability, following the model of Hill and colleagues,¹⁵ using a single random effect in which different subsets of raters rated each subject. Following the TDCRP analyses, analysis of variance (ANOVA) was used to assess for treatment adherence by intervention and multiple regression analysis to assess therapist FC effects on outcome, controlling for intake score and intervention.

RESULTS

The four adherence raters graded between 30 and 84 tapes (mean = 56, SD = 27.0) for each of the four treatment conditions and for FC and ED. Correlations exceeded 0.90 on all but one measure (Table 1). These scores suggest excellent interrater agreement (Fleiss,²⁷ p. 223).

All therapist-subject dyads were rated adherent. Overall adherence by intervention is shown in Table 2

and the HIV-Depression Study graph in Figure 1: the boldfaced descending diagonal of ratings in Table 2 indicates that each team of psychotherapists scored highest in its own psychotherapeutic modality. The pattern of mean scores for each subscale resembles those in the study by Hill and colleagues¹⁵ (compare the HIV-Depression and TDCRP graphs in Figure 1).

ANOVA among adherence ratings showed significant discrimination of therapeutic modalities (Table 3). The percentages of variance explained (η^2) resemble those Hill and colleagues reported: CBT = 0.77, IPT = 0.60, CM = 0.69, FC = 0.22, ED = 0.40. There were significant time-by-session interactions for IPT and SWI, which showed significant decreases in adherence levels between early and late sessions. IPT adherence scores decreased from a mean score of 2.83 in early sessions to 2.18 in late sessions ($F = 3.78$, $df = 3$, $P = 0.013$), whereas SWI fell from 2.75 to 2.36 ($F = 3.18$, $df = 3$, $P = 0.027$). In contrast, CBT scores fell from 3.05 to 2.90 (not significant) and SP rose from 2.01 to 2.06 (not significant).

ANOVA demonstrated a significant difference in

TABLE 1. Rater CSPRS intraclass correlations

Subscale	Early Session	Late Session
CBT	0.99	0.99
IPT	0.99	0.93
SP	0.93	0.96
SWI	0.95	0.97
FC	0.97	0.97
ED	0.97	0.89

♣ Note: CSPRS = Collaborative Study Psychotherapy Rating Scale; CBT = cognitive-behavioral therapy; IPT = interpersonal psychotherapy; SP = supportive psychotherapy; SWI = supportive psychotherapy with imipramine; FC = facilitative conditions; ED = explicit directiveness.

TABLE 2. Mean CSPRS subscale scores for each treatment, HIV-depression study

Treatment	Subscale					
	IPT	CBT	SP	SWI	FC	ED
IPT	2.4	1.3	1.9	1.3	4.3	3.9
CBT	1.4	3.0	2.0	1.3	5.4	4.4
SP	1.4	1.3	3.4	1.4	4.9	4.0
SWI	1.2	1.2	2.0	2.6	4.0	4.5

♣ Note: Highest score for each treatment appears in **boldface**. Abbreviations are defined in the note to Table 1.

FC across interventions ($F = 6.81$, $df = 52$, $P = 0.0006$), accounting for 28% of FC variance (CBT > SWI).

Because all therapists met criteria for adherence, the effect of adherence on outcome was not explored. We hypothesized that “facilitative conditions” (FC), measuring therapist qualities such as warmth, supportiveness, and empathy, might reflect general competence independent of treatment adherence, and thereby predict treatment outcome. Multiple regression analysis was employed, controlling for initial BDI score or Ham-D score and for treatment intervention. For individual therapists, there was no significant effect of FC on treatment outcome as measured by BDI and Ham-D scores, although 15% of BDI outcome ($F = 1.52$, $df = 10,41$, $P = 0.17$) and 20% of Ham-D variance ($F = 1.44$, $df = 10,41$, $P = 0.20$) were explained by FC.

DISCUSSION

This study adapted an adherence measure, the CSPRS, from the TDCRP study to related interventions used on a similar but HIV-seropositive sample of patients with depressive symptoms. The results of our smaller, single-site study replicate for the first time those Hill and colleagues found for the TDCRP program. Our adherence monitors achieved high reliability and discriminated the four interventions from one another. Therapists in the four interventions achieved excellent treatment adherence. The scoring range on the CSPRS is similar—even slightly accentuated—in our study when compared with the TDCRP.

This replication of the TDCRP findings, under slightly altered conditions, confirms our initial two hypotheses. Demonstration of interrater reliability for adherence monitors and adherence for therapists validates the outcome results from this treatment study.¹⁴ Thus adherence here demonstrates both discrete adherence of the therapist to a particular modality (i.e., treatment integrity) and avoidance of the techniques of other modalities (i.e., treatment differentiability).²⁸ Some overlap of treatments is inevitable due to the so-called common factors of psychotherapy.²⁹

Our final hypothesis, however, was not confirmed. Although some variation appeared in therapist FC, it did not affect treatment outcome. It may be that FC does not measure, and might even confound (I. Elkin, personal communication, July 1996), true therapist competence. Or perhaps when good therapists are all

adherent, outcome measures reflect differences in treatment technique rather than therapist differences.

One limitation of this research is the relatively small patient sample, although it proved more than sufficient to yield statistically significant results. Differing numbers of therapists and cases per treatment cell may have influenced the variance within cells, yet all therapists were rated adherent in each cell. Raters may have noted the shorter length of SP and SWI sessions, but raters were asked to rate 104 adherence items, not to guess the treatment modality.

Our SP subscale, psychometrically untested and measuring a relatively featureless therapy, still had high intraclass correlation scores at least comparable to those for CM on the original CSPRS. The time-by-session interaction noted for IPT and SWI probably reflects a decrement in therapist activity as recovering patients in these more potent therapies learned their treatment rationale: if the patients, rather than their therapists, voiced therapeutic shibboleths, scoring of *therapist* adherence would fall. By contrast, the SP therapists, whose subjects showed less improvement, may have redoubled their treatment efforts in later sessions.

Monitoring of psychotherapy adherence is a young science; the terms do not even appear in the index of a major recent reference text.³⁰ It is reassuring that adherence technology can be translated from a state-of-the-art, multisite NIMH-designed treatment program to a different locus and design. The similarity of the rating scores for interventions in the two studies suggests ei-

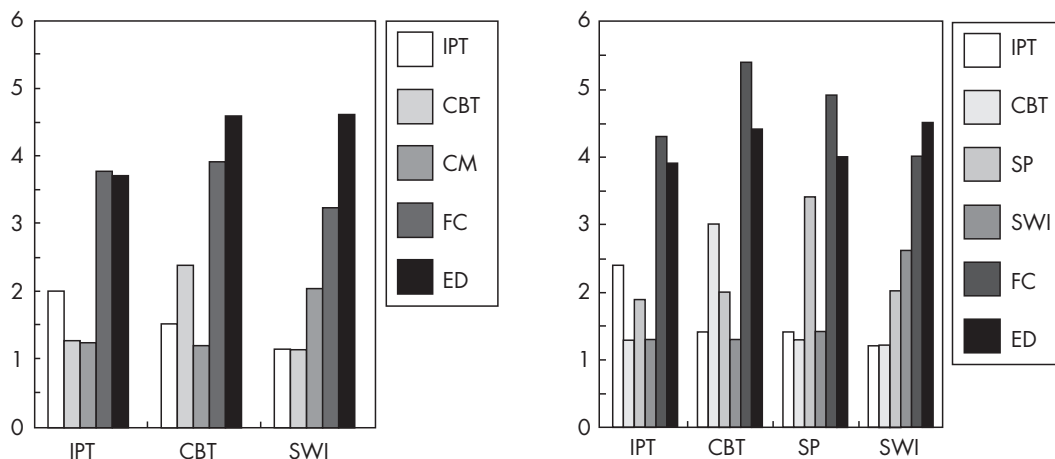
ther the ease and consistency with which these specified treatments can be learned or, perhaps, a systematic bias of the rating instrument. Hill and colleagues called the CSPRS “robust”¹⁵ (p. 77). Our clinical impression, too, is that the CSPRS accurately differentiated among the treatments our highly skilled therapists delivered.

TABLE 3. CSPRS subscale scores, by analysis of variance

Variable	F	P	eta ²
Overall (early and late session) scores			
CBT	67.6	<0.0001	0.796
IPT	40.2	<0.0001	0.699
SP	20.8	<0.0001	0.545
SWI	92.9	<0.0001	0.843
FC	6.8	0.0006	0.282
ED	3.2	0.0324	0.154
Early session scores			
CBT	39.1	<0.0001	0.697
IPT	32.6	<0.0001	0.657
SP	13.5	<0.0001	0.448
SWI	72.0	<0.0001	0.809
FC	4.3	0.0091	0.201
ED	3.6	0.0205	0.173
Late session scores			
CBT	32.8	<0.0001	0.663
IPT	14.6	<0.0001	0.467
SP	15.3	<0.0001	0.484
SWI	36.2	<0.0001	0.685
FC	7.0	0.0005	0.297
ED	1.0	0.3988	0.057

✎ Note: eta² = percentage of variance explained. Abbreviations are defined in the note to Table 1.

FIGURE 1. Comparison of HIV-depression and TDCRP study adherence ratings. HIV = human immunodeficiency virus; other abbreviations are defined in the note to Table 1.



This paper is dedicated to the late Samuel W. Perry, M.D., whose research program this was.

The authors thank Bonnie Gitlin, M.A., Stephen Hartman, M.A., and Michael Moffatt, M.A., for their superb adherence monitoring. This work was supported in part by Grant

MH46250 from the National Institute of Mental Health and by the Reader's Digest and New York Community Trust. It was presented in part at the annual meeting of the Society for Psychotherapy Research, York, England, June 1994.

REFERENCES

1. Shaw BF, Dobson KS: Competency judgments in the training and evaluation of psychotherapists. *J Consult Clin Psychol* 1988; 56:666-672
2. Waltz J, Addis ME, Koerner K, et al: Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *J Consult Clin Psychol* 1993; 4:620-630
3. Luborsky L, McLellan AT, Woody GE, et al: Therapist success and its determinants. *Arch Gen Psychiatry* 1985; 42:602-611
4. Woody G, Luborsky L, McLellan AT, et al: Psychotherapy for opiate addicts: does it help? *Arch Gen Psychiatry* 1983; 40:639-645
5. Svartberg M: Manualization and competence monitoring of short-term anxiety-provoking psychotherapy. *Psychotherapy* 1989; 26:564-571
6. Shapiro DA, Startup M: Measuring therapist adherence in exploratory psychotherapy. *Psychotherapy Research* 1992; 2:193-203
7. Hoglend P, Piper WE: Focal adherence in brief dynamic psychotherapy: a comparison of findings from two independent studies. *Psychotherapy* 1995; 32:618-628
8. Butler SF, Henry WP, Strupp HH: Measuring adherence in time-limited dynamic psychotherapy. *Psychotherapy* 1995; 32:629-638
9. Barber JP, Crits-Christoph P: Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: a preliminary report. *Psychotherapy Research* 1996; 6:81-94
10. Barber JP, Krakauer I, Calvo N, et al: Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence. *J Psychother Pract Res* 1997; 6:12-24
11. Klein DF: Preventing hung juries about therapy studies. *J Consult Clin Psychology* 1996; 64:81-87
12. Hollon SD: Final Report: System for Rating Psychotherapy Audiotapes. Rockville, MD, U.S. Department of Health and Human Services, 1984
13. Elkin I, Parloff MB, Hadley SW, et al: NIMH Treatment of Depression Collaborative Research Program. *Arch Gen Psychiatry* 1985; 42:305-316
14. Elkin I, Shea MT, Watkins JT, et al: National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry* 1989; 46:971-982
15. Hill CE, O'Grady KE, Elkin I: Applying the collaborative study psychotherapy rating scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *J Consult Clin Psychol* 1992; 60:73-79
16. Gaston L: The concept of the alliance and its role in psychotherapy: theoretical and empirical considerations. *Psychotherapy* 1990; 27:143-153
17. Beck AT, Rush AJ, Shaw BF, et al: *Cognitive Therapy of Depression*. New York, Guilford, 1979
18. Klerman GL, Weissman MM, Rounsaville BJ, et al: *Interpersonal Psychotherapy of Depression*. New York, Basic Books, 1984
19. Markowitz JC, Klerman GL, Perry SW: Interpersonal psychotherapy of depressed HIV-seropositive patients. *Hospital and Community Psychiatry* 1992; 43:885-890
20. Pergami A, Grassi L, Markowitz JC, et al: *Il Trattamento Psicologico della Depressione nell'Infezione da HIV: La Psicoterapia Interpersonale [The psychological treatment of depression in HIV infection: interpersonal psychotherapy]*. Milan, Franco Angeli, 1999
21. Rogers CR: *Client-Centered Therapy*. Boston, Houghton Mifflin, 1951
22. Markowitz JC, Kocsis JH, Fishman B, et al: Treatment of HIV-positive patients with depressive symptoms. *Arch Gen Psychiatry* 1998; 55:452-457
23. Fawcett J, Epstein P, Fiester SJ, et al: Clinical management: imipramine/placebo administration manual. NIMH Treatment of Depression Collaborative Research Program. *Psychopharmacol Bull* 1987; 23:309-324
24. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; 25:56-62
25. Beck AT: *Depression Inventory*. Philadelphia, Center for Cognitive Therapy, 1978
26. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420-428
27. Fleiss JL: *Statistical Methods for Rates and Proportions*, 2nd edition. New York, Wiley, 1981
28. Kazdin AE: Comparative outcome studies of psychotherapy: methodological issues and strategies. *J Consult Clin Psychol* 1986; 54:95-105
29. Butler SF, Strupp HH: Specific and nonspecific factors in psychotherapy: a problematic paradigm for psychotherapy research. *Psychotherapy* 1986; 23:30-40
30. Bergin AE, Garfield SL (eds): *Handbook of Psychotherapy and Behavior Change*, 4th edition. New York, Wiley, 1994