The Psychotherapist Matters: Comparison of Outcomes Across Twenty-Two Therapists and Seven Patient Samples

Lester Luborsky, University of Pennsylvania
A. Thomas McLellan, University of Pennsylvania and Philadelphia VA Medical Center
Louis Diguer, Laval University, Québec, Canada
George Woody and David A. Seligman, University of Pennsylvania

This research relies on a useful focus for finding the contribution of the psychotherapist to the outcome of the treatment: the outcomes of each therapist's caseload. Our data consist of 22 therapists' caseloads within seven samples of drug-addicted and depressed patients. We concluded that (a) there were important differences in the improvement levels and post-treatment outcomes of patient caseloads among the therapists sampled, and (b) these differences in improvement could not be explained by differences in patient background or severity. Some of the differences appeared to reflect the therapists' efficacy with their patients because (a) a unique feature of the design was that three therapists took part in more than one study and therefore more than one caseload; these three therapists showed a similar efficacy in each new caseload. (b) At this time, it may be that the safest basis for choosing therapists for research studies or for clinical purposes is their "work sample" record of efficacy with their previous caseloads.

Key words: **psychotherapy, psychotherapist, outcome, drug addiction, depression.** ***[Clin Psychol Sci Prac 4:53–65, 1997]***

Those who wish advice about choosing a therapist for themselves or their patients often call with this question: "Would you please tell me the names of some good therapists?" I always offer them some names based on my impressions about a few therapists' past performance with their patients. But, I often wish I would get more help in answering such questions from psychotherapy research.

After decades of studies of therapist requisite qualities, it is amazing that the term "neglected" still fits this topic (Bergin, 1997). We hope this article will explain the aspect of this research that needs most expansion—a shift from a research strategy based on therapist characteristics and process measures to one that evaluates the performance of therapist caseload as the main basis for evaluating therapist efficacy.

In the years from 1940 to 1960 studies listed the qualities of good therapists through the views of supervisors and other experts (as in Holt & Luborsky, 1958). The experts clearly viewed the most effective therapists as near-paragons, having psychological mindedness, intelligence, supportiveness, and other obviously and generally helpful qualities (Luborsky, 1952). An awakening slowly occurred over the next decades when it was found that even such generally desirable qualities of therapists as those listed by the experts did not correlate well with the outcomes of their treatments (e.g., see the review by Parloff, Waskow, & Wolfe, 1978).

Even a recent review of studies correlating therapist qualities with outcomes showed that only a few qualities showed even moderate effect sizes (Table 7.2 in Beutler, Machado, & Neufeldt, 1994). Although correlations are usually small, the highest of these that appear in different studies are overall mean alliance of the patient and therapist (.26; Horvath & Symonds, 1991), the therapist's short-term therapy training (.43; Berman & Norton, 1985), professional affiliation as a psychologist (.28; Smith, Glass, & Miller, 1980), the use of manuals (.38; Crits-Christoph et al., 1991), and the therapist's use of the technique of directiveness (−.45; Svartberg & Stiles, 1992). With such mixed success of prediction based on

Address correspondence to Dr. Lester Luborsky, 514 Spruce St., Philadelphia, PA, 19106.

ideas of what the therapist should be and should do, it is no wonder that the therapist's contribution to outcome has been a neglected variable.

THE NEW DIRECTION: THE OUTCOMES OF EACH THERAPIST'S CASELOAD

The beginning of a new direction in the design of studies of therapist performance started with an impressive study by Ricks (1974), who compared two therapists, one with a high success rate and one with a low success rate, in the treatment of severely disturbed adolescent boys. Rather than confine evaluations of therapist efficacy to ratings of qualities that were assumed to be related to good performance of the therapy, Ricks's study focused directly on measures of patient outcome as evidence of therapist performance. The newer strategy takes each therapist caseload as the unit to be studied; an attempt is then made to compare the outcomes of each therapist's caseload, after taking into account differences in pretreatment severity of the caseloads. This new direction of research designs was continued in several articles, including one by Orlinsky and Howard (1980). These design principles were (a) a focus on each therapist's performance as measured by outcomes of their caseloads and (b) an examination of the therapist and therapy qualities that were associated with that performance. Research using these design principles has yielded indications that therapist variables are at least moderately related to treatment outcomes. For example, in the last 15 years a set of studies (Crits- Christoph et al., 1991; Crits-Christoph & Mintz, 1991; Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985; Luborsky et al., 1986; McLellan, Woody, Luborsky, & Goehl, 1988) has demonstrated differences in each therapist's *general* level of help to the patients in their caseloads. A similar but expanded framework of analyses was applied in Najavits and Strupp (1994) to establish differences in caseload outcome performance among 16 therapists in the Vanderbilt II study and then to relate different therapist behaviors with these differences in performance. It was also discovered that differences among therapists in percentage of explained variance were less among those therapists who followed a manual than for those therapists who did not (Crits-Christoph et al., 1991).

CURRENT RESULTS WITH THE NEWER RESEARCH MODEL

With this newer research model we explored (a) the range of differences in patient improvement among therapist caseloads and (b) the correlations between the therapists' personal qualities and the outcome performance of their caseloads. In the studies we used well-validated measures of outcomes of patient engagement, such as of patient retention in treatment, and of patient outcome including the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Global Assessment scale (GAS) or Global Assessment of Functioning (GAF; Luborsky et al., 1993), and the Hamilton Depression Rating scale (HDRS; Hamilton, 1960). For each of these outcome measures we evaluated not only the scores at the end of treatment but also the changes in these measures (improvement) from start of treatment to follow-up.

These analyses of benefits are reported here for each therapist's caseload in each of seven psychotherapy samples using the BDI, GAS, Addiction Severity Index (ASI), or HDRS: (a) three samples of opiate-addicted VA patients (Luborsky et al., 1985, 1986), (b) two samples of opiate-addicted community clinic patients (Luborsky, Woody, Hole, & Velleeo, 1995; Woody, McLellan, Luborsky, & O'Brien, 1995), (c) the Supportive-Expressive (SE) depression outpatient sample (Luborsky, Mark, et al., 1995), and (d) the SE-chronic depression outpatient sample (Luborsky et al., 1996). None of the therapists in the seven samples were beginners; all were experienced in the particular form of psychotherapy that they performed in the course of the study. They were either psychologists with a Ph.D. or psychiatrists with an M.D.

As a whole, the present analyses continue and expand the framework of analytic principles in Luborsky et al. (1985), and especially, they add more on the issue of the stability of therapist performance over different samples.

SAMPLES 1–3: OPIATE-ADDICTED VA SAMPLES

Samples 1–3 came from an early study of two forms of professional psychotherapy with opiate-addicted male, veteran, methadone- maintained patients (as more fully described in Luborsky et al., 1985; Woody et al., 1983). They were randomly assigned to one of three 6-month treatment conditions: Sample 1, SE dynamic psychotherapy plus Drug Counseling (DC; $n = 32$); Sample 2, Cognitive-Behavioral (CB) psychotherapy plus DC ($n = 39$); or Sample 3, Drug Counseling alone ($n = 39$). The attendance for the patients in psychotherapy ranged from 3 to 24 sessions with a mean of 12 sessions; patients assigned to DC alone averaged 16 sessions. Each treat-

ment was guided by a special treatment manual and each therapist's adherence to the manual was rated from tape-recorded sessions by two independent, experienced judges.

Therapists

Eighteen drug counselors and nine psychotherapists (five SE and four CB) participated. We selected for detailed analyses the three therapists from each treatment group with the largest patient caseloads. The subjects were between 18 and 55 years of age, were nonpsychotic, and were beginning a new drug abuse treatment episode in a methadone maintenance program for opiate dependence. Although opiate dependent, these patients had diagnoses that were similar to those for patients typically given psychotherapy; most common diagnoses were affective disorder, especially depressive disorder (43%), and anxiety disorders (9%). Antisocial personality was 15% of the samples and, as such, relatively more frequent than in other patient groups.

Measures of Treatment Outcome

The following measures were administered to each patient at the start of treatment and at 7-month follow-up (i.e., 1 month after the 6 months of treatment): the BDI (Beck et al., 1961), the Maudsley Personality Inventory (MPI; Eysenck, 1959), the Hopkins Symptom Check List-90 (SCL-90; Derogatis, 1983), the Shipley Institute of Living scale (SILS; Shipley, 1940), and the ASI (McLellan et al., 1985).

Treatment Outcomes

In all three treatments there was a significant reduction in drug use and criminal activity. Psychiatric symptoms generally improved, and there was a beneficial reduction in the use of medications both prescribed and nonprescribed. More complete results are given in Woody et al. (1983) and Woody, Luborsky, O'Brien, and McLellan (1989).

Variability in Each Therapist's Level of Outcome With His/Her Caseload

We used two objective measures of therapist effectiveness: (a) mean percentage improvement in the patients in each therapist's caseload from admission to the 7- month follow-up and (b) adjusted (by pretreatment score) post-treatment outcome at the 7-month follow-up (1 month after termination).

Percentage Improvement From Admission to 7 Months. The improvement was based on the mean percentage change of each therapist's caseload from the start of the study to the 7-month evaluation point [(Baseline − Follow-up)/ Baseline] for all patients in each therapist's caseload and for each of the seven outcome measures within the ASI. In Table 1 we can see obvious differences in the level of caseload improvements among therapists. There was a significant, $p < .05$, between-therapist effect on all measures as shown through a one-way analysis of variance on the percentage change scores.

The wide range of outcomes among the therapists in effectiveness with their randomly assigned caseloads appears to be shown, for example, for SE psychotherapy (Table 1) by Therapist $T1^{1,6,7}$ (superscript numbers indicate which samples the therapist participated in, if more than one), with a mean percentage improvement in ASI Psychiatric Status of 82% from the start to the 7-month point; Therapist T2's patients showed a mean change of only 41%, while Therapist T3's patients showed −1% (worsening)! The three therapists differed widely on the percentage improvement in the BDIs of the patients in the caseloads.

Change, in Effect Size Terms. The magnitude of the change in each therapist's caseload was also given as an effect size score for each of the seven measures, and then these individual effect sizes were averaged to produce a single, general score (Luborsky et al., 1985). These average effect sizes for the nine therapists paralleled the *percentage* change measures.

Adjusted Post-treatment Outcomes for Each Therapist's Caseload. Post-treatment outcome status of the patients in each therapist's caseload was also important. This was shown in the same earlier study by analyses of covariance on the 7-month follow-up scores with treatment admission score on the criterion variable as the covariate. This analysis offered a second basis for comparing main outcome status among each therapist's caseload. The nine adjusted scores (for Samples 1–3) for each therapist's caseload on the seven outcome measures are compared (analysis of covariance, Duncan's multiple range test) on each of the 7-month outcome measures. These are ASI composite scores, summing standardized combinations of several items measuring status during the 30 days prior to follow-up. The composite scores range from 0 to 1.0, with larger scores showing worse status.

Table 1. Sample 1: Percentage change and effect size (ES) for each therapist's caseload of opiate-addicted patients in supportive-expressive (SE) psychotherapy

		BDI					ASI Psychiatric Status				
Therapist	*N*	Mean Initial	Mean Term.	Mean % Change	ES	R	Mean Initial	Mean Term.	Mean % Change	ES	R
T1[1,6,7]	10	17	7	58	1.14	1	.206	.081	61	1.01	1
T2	8	13	8	38	.58	2	.176	.103	41	.42	2
T3	8	13	12	7	.11	3	.173	.175	−1	−.01	3
Mean		14	9	34.3			.185	.120	33.7		

Note: R = rank.

Table 2. Sample 2: Percentage change and effect size (ES) for each therapist's caseload of opiate-addicted patients in cognitive-behavioral (CB) psychotherapy

		BDI					ASI Psychiatric Status				
Therapist	*N*	Mean Initial	Mean Term.	Mean % Change	ES	R	Mean Initial	Mean Term.	Mean % Change	ES	R
T4	11	19	12	36	.94	1	.237	.152	36	.49	1
T5	10	16	12	24	.57	2	.224	.186	17	.21	2
T6	9	14	12	14	.29	3	.202	.175	13	.14	3
Mean		16	12	24.7			.221	.171	22.0		

Note: R = rank.

Table 3. Sample 3: Percentage change and effect size (ES) for each therapist's caseload of opiate-addicted patients in drug counseling (DC)

		BDI					ASI Psychiatric Status				
Therapist	*N*	Mean Initial	Mean Term.	Mean % Change	ES	R	Mean Initial	Mean Term.	Mean % Change	ES	R
T7	9	13	12	5	.08	2	.221	.208	6	.04	3
T8	6	13	13	−3	−.07	3	.331	.302	9	.13	1
T9	7	12	10	14	.20	1	.188	.165	12	.10	2
Mean		13	12	5.3			.247	.225	9.0		

Note: R = rank.

Two findings appeared: (a) The therapists differed significantly on all of the adjusted 7-month composite scores of their patient samples, $p < .05$. (b) Therapists whose caseloads had shown the greatest percentage improvement also showed the best post-treatment outcomes.

Why Therapists Differ in Benefits to the Patients in Their Caseloads (for Samples 1–3)

Patient Qualities. We first examined whether the caseload of each therapist differed from the other caseloads in patient background and pretreatment status. We especially checked patient psychiatric severity because if such

differences were present they could have affected the outcomes (Diguer, Barber, & Luborsky, 1993). But severity and other measures were mostly not significantly different, probably because the cases had been randomly assigned; such differences in patient pretreatment status seemed unlikely to account for much of the therapists' performance differences.

Therapist Qualities. We then examined therapist qualities as judged by their peers (Luborsky et al., 1985). The Therapist's Personality Qualities scales consist of 11 five-point scales to measure three main qualities: personal adjustment, therapeutic skill, and interest in helping patients. The scales were rated by three independent judges who knew the therapists and their work with patients. A factor analysis of the 11 items yielded two main factors: (a) interest in helping patients and (b) psychological health and skill. We then correlated the mean adjusted 7-month outcome scores from the caseloads of the same therapists, with the scores on these two factors. None of them reached the .05 level of significance, probably because there were only nine therapists. However, interest in helping patients correlated .44 with the patients' psychological functioning at termination; therapist's psychological health and skill correlated .41 with the patients' psychological functioning at termination.

Patient-Therapist Relationship Qualities. We then examined patient-therapist relationship qualities measured during the third session by the Helping Alliance questionnaire (HAq- I). This questionnaire estimates the degree to which the patients experienced the therapist and the therapy as being helpful. Correlations with outcomes were generally high. For example, the HAq-I total score correlated .58 with patients' psychological functioning at the 7-month point, across the nine therapists.

Therapy Qualities. We then examined therapy qualities with a focus on how much the therapists had adhered to their treatment manual. The adherence measure is based on McLellan's measure of "purity" of the therapist's technique (Luborsky et al., 1985), which is the ratio of the quality ratings for the intended therapy to the total of all quality ratings (SE, CB, and DC qualities as described in the manual). As might be expected, therapists whose patients benefited most had greater adherence (purity) to the therapeutic techniques described in their manual. This was especially true for the SE therapists in their sample of 41 patients.

SAMPLES 4 AND 5: OPIATE-ADDICTED COMMUNITY CLINIC PATIENTS

The study of Samples 4 and 5 was aimed at testing the efficacy of individual psychotherapy of psychiatrically symptomatic, opiate-dependent patients during methadone maintenance treatment in three community programs (Woody et al., 1995). Like the first study, this one also consisted of 24 weeks of (a) DC plus SE psychotherapy (Sample 4) and (b) DC plus supplemental DC (Sample 5). Follow-ups were done 1 month and at 6 months after treatment. Eighty-four subjects were evaluated at all follow-up points. Five SE therapists took part. Fifty-three drug counselors worked in the three treatment programs of the community clinic and participated in the project; 26 served as additional counselors.

Measures of Treatment Outcome

Psychological tests administered at the start of treatment and at 1- and 6-month follow-up were the BDI, the MPI, the SCL-90, and the SILS along with SADS-L given at the study intake and the ASI given at intake and at both follow-up points.

Treatment Outcomes

Patients in the two treatments, SE and DC, had similar proportions of *opiate*-positive urines; however, the SE patients had fewer *cocaine*-positive urines and required lower doses of methadone (Woody et al., 1995). By 1 month after treatment both groups had made significant gains but with no significant differences between them. By the 6-month follow-up many of the gains by DC patients had diminished but most gains by SE patients were maintained or improved.

Results

All of the following results are from Tables 4 and 5 on the performance of five therapists in the opiate abuse community clinic study (Samples 4 and 5).

There was a wide range in the benefits shown by each therapist's caseload. For the 11 patients using baseline to follow-up improvements on the BDI, for example, $T10^{4,6}$'s caseload averaged 37% improvement, as compared with T11 with 27%, $T12^{4,6}$ with 45%, T13 with 11%, and T14 with 7.4%.

Table 4. Sample 4: Percentage change and effect size (ES) in each therapist's caseload of opiate-addicted patients in the study of community clinics in supportive-expressive (SE) psychotherapy (12-month data)

		BDI					ASI Psychiatric Status				
Therapist	*N*	Mean Initial	Mean Term.	Mean % Change	ES	R	Mean Initial	Mean Term.	Mean % Change	ES	R
T10[4,6]	11	18.93	12.00	37	.94	2	.441	.208	53	.69	1
T11	14	20.07	14.63	27	.76	3	.423	.256	39	.49	3
T12[4,6]	11	21.8	12.09	45	1.35	1	.444	.224	50	.65	2
T13	11	25.18	22.36	11	.39	4	.379	.392	−3	−.02	5
T14	7	17.78	16.00	10	.25	5	.388	.388	0	.01	4
Mean		21	15	26.0			.415	.294	27.8		

Note: R = rank.

Table 5. Sample 5: Percentage change and effect size (ES) for each therapist's caseload of opiate-addicted patients in the study of community clinics in drug counseling (DC)

		BDI					ASI Psychiatric Status				
Therapist	*N*	Mean Initial	Mean Term.	Mean % Change	ES	R	Mean Initial	Mean Term.	Mean % Change	ES	R
T15	9	12	12	0	.07	2	.224	.208	7	.05	3
T16	6	13	13	0	−.06	3	.335	.302	10	.10	1
T17	7	12	10	17	.22	1	.187	.165	12	.07	2
Mean		12	12	5.7			.249	.225	9.7		

Note: R = rank.

A noteworthy finding is that T12[4,6] had the highest percentage improvement on the BDI in this sample of opiate abuse patients and also had the highest percentage of improvement on the BDI in Sample 6 of depressed patients.

The trends in the results with GAS were similar to those from the BDI.

In Samples 4 and 5 the percentage of improvement shown by patients in the therapists' caseloads tended to be slightly less than in Samples 1–3.

SAMPLE 6: SUPPORTIVE-EXPRESSIVE DYNAMIC PSYCHOTHERAPY FOR DEPRESSION

Sample 6 included 33 patients who fit the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed. rev.; DSM-III-R; American Psychiatric Association, 1987) criteria for depression. One of the aims of this study was to examine the effectiveness of SE dynamic psychotherapy for depression, with the treatment as described in a special manual (Luborsky, Mark, et al., 1995). The length of the therapy was 16 sessions, at one session per week. Assessments were at the initial point, at the end of 4 months, and 6 months later.

Measures of Treatment Outcome

The measures consisted of initial and termination follow-up for the GAF (Spitzer, Gibbon, Williams, & Endicott, 1995); the GAF is a slight rewording of the earlier Health-Sickness Rating scale (Luborsky, 1962). The third measure was the HRSD (Hamilton, 1960).

Variability in Therapists' Caseloads

These patients were treated by one of four SE dynamically trained therapists: T10[4,6], T1[1,6,7], T12[4,6], and T18. Note that three of the four therapists had also taken part in the other treatment samples. Results for Sample 6 are described more extensively in Luborsky et al. (1996), where they were combined with the chronic depression sample to make a total sample of 49 patients, but they

Table 6. Sample 6: Percentage of change in the patients in each of four therapists' caseloads in the SE depression sample

		BDI				GAS				HDRS		
Therapist	*N*	Mean Initial	Mean Term.	Mean % Change	R	Mean Initial	Mean Term.	Mean % Change	R	Mean Initial	Mean Term.	Mean % Change
T10[4,6]	9	29	21	28	4	51	63	24	4	19	11	42
T1[1,6,7]	13	30	15	50	2	48	65	36	2	20	9	55
T12[4,6]	11	29	13	55	1	48	66	38	1	18	8	56
T18	4	24	13	46	3	51	65	28	3	17	9	47
Mean		28	15.5	44.8		49.5	64.75	31.5		18.0	9.25	50.0

Note: R = rank; GAS = Global Assessment Scale.

Table 7. Sample 7: Percentage change in the patients in each of five therapists' caseloads in the chronic depression sample

		BDI				GAS/GAF				HDRS			
Therapist	*N*	Initial	Term.	Mean % Change	R	Initial	Term.	Mean % Change	R	Initial	Term.	Mean % Change	R
T19	3	25	8	68	3	58	73	26	2	15	4	73	1
T20	3	20	7	65	4	60	72	20	3	10	3	70	2
T1[1,6,7]	2	28	8	71	2	58	64	10	5	13	4	69	3
T21	4	27	5	81	1	56	72	29	1	17	6	65	4
T22	4	27	28	−4	5	60	67	12	4	14	12	14	5
Mean		25	11	56.2		58	70	19.4		14	6	58.2	

Note: R = rank; GAF = Global Assessment of Functioning.

are presented here separately because the set of therapists differed from the chronic depression set.

Findings From Sample 6, SE Depression

Three of the four therapists had moderately large samples of patients: T10[4,6] had 9, T1[1,6,7] had 13, T12[4,6] had 11, but T18 had 4 (Table 6). It is helpful to this comparison of change that all four therapists' caseloads started with initial BDI means that were very similar to each other, probably reflecting the fact that the cases were randomly assigned.

The percentages of improvement in each of the caseloads were based as usual on the absolute change divided by the initial level. These mean outcome percentages for each caseload differed considerably from therapist to therapist. For the BDI the improvement of each of the four therapists was 28%, 50%, 55%, and 46%. For the GAS, the improvement was 24%, 36%, 38%, and 28%. The percentage changes on the HRSD were fairly similar to those on the BDI: 42%, 56%, 57%, 51%. In fact, the rankings of therapist performance across the three measures were identical.

Sample 6 can be especially informative about another type of finding: the differences in outcome *within* each caseload (Table 9), because at least three of the therapists had nine or more patients. We can see that the range of improvements, as well as their standard deviation, differs from therapist to therapist. Within the caseloads of T10[4,6], T1[1,6,7], and T12[4,6] we can see very large variation from patient to patient in the percentage change in the BDI and GAS. These variations from case to case within the caseload must be a product of the usual conditions that influence outcome, both from the patients' and therapists' side, as well as from their interaction; they suggest that even among the therapists whose patients have the highest mean levels of improvement, there is a large range of outcomes.

The trustworthiness of the inferences from the results, of course, depends on the comparability of the cases for each caseload. Yet we have reasons to trust some of the

Table 8. Summary of mean performance for therapists who were in more than one sample

	BDI			ASI or GAS		
	Mean % Change	Sample	Rank	Mean % Change	Sample	Rank
Mean of T1[1,6,7]	58	1	1	61	1 ASI	1
	50	6	2	36	6 GAS	2
	71	7	2	10	7 GAS	5[a]
Mean of T12[4,6]	45	4	1	50	4 ASI	2
	55	6	1	38	6 GAS	1
Mean of T10[4,6]	37	4	2	53	4 ASI	1
	28	6	4	24	6 GAS	4

[a]Based on only two patients.

results: (a) The initial levels of the caseloads on the main measures were very similar at baseline due to random assignment. (b) More importantly, T1[1,6,7] and T12[4,6,] who had the best performance here, were also the outstanding therapists in other samples.

SAMPLE 7: CHRONIC DEPRESSION SAMPLE

Sample 7 was primarily diagnosed with major depression chronic (DSM-III-R) and provided with 20 sessions of SE dynamic psychotherapy. Although we will draw conclusions from this sample, the conclusions are tentative because the number of patients in each therapist's caseload was small; the three largest are T19, $n = 3$; T20, $n = 3$; and T22, $n = 4$.

Variability in Therapist Effectiveness

Even though the sample size of each therapist's caseload was small, it was still of interest to examine the three largest in terms of percentage on each measure of change from initial to termination. For T20 the BDI percentage change was 65, for T19 it was 68, and for T22 it was only 4. Therapist rankings on percentages of change shown by their caseloads on the BDI tend to be similar to the ranks of percentages on the GAS and the HRSD.

As in the other samples, the BDI mean scores of the three caseloads at the initial point were not markedly different from each other, and these do not seem sufficient to explain the observed differences in percentage change. For example, for T22 the initial caseload mean BDI was 27, which implied a depressed state, but the termination BDI was almost identical at 28. This stands in contrast to the change from 27 to 5 by the caseload of T21.

CONCLUSIONS AND DISCUSSION

The results suggest two main conclusions.

(a) The therapists in all seven samples differed widely in the mean level of improvement shown by the patients in their caseloads. Our studies are not the first to show large differences in patient outcome across therapists. Yet, these are somewhat surprising because (a) patients within each sample were similar in terms of diagnosis; (b) they were randomly assigned; (c) the therapists had been selected for their competence in their particular form of psychotherapy; and (d) the therapists were regularly supervised and were further guided by treatment manuals. Despite these steps that should have maximized skill and minimized differences, the range of percentages of improvement for the 22 therapists in the seven samples was from slightly negative change, to slightly more than 80% improvement. Ever since the impressive study by Ricks (1974), in which single therapist caseloads were the unit of analysis, such variation has been occasionally reported. A wide range of mean improvement was also reported for the 28 therapists in the recent reanalysis of the National Institute of Mental Health (NIMH) study by Blatt, Sanislow, Zuroff, and Pilkonis (1996).

(b) Therapists who excelled in one sample excelled in other samples. A unique asset of these samples is that three of the 22 therapists in these seven samples took part in more than one of the studies; we could therefore compare their performance rankings in the different samples as is summarized in Table 8. If we look only at the results on the BDI, we see that the caseload of the therapist T1[1,6,7] in Sample 1 for opiate- addicted patients showed a 57% improvement from initial to termination. The same therapist participated in Sample 6 with depressed patients, where her patients' BDIs showed a mean of 50% improvement. Although T1[1,6,7] also saw only two patients in the major depression chronic, Sample 7, each also showed impressive improvement. In all three different samples, therefore, this therapist's patients were among the most improved. A similar consistency across samples was shown in Samples 4 and 6 by T12[4,6,] whose patients

Table 9. Sample 6: Performance of individual patients *within* each of four psychotherapists' caseloads in the SE depression sample

		BDI			GAS			HDRS			
Therapist	Patient ID	Initial	Term.	Mean % Change	Initial	Term.	Mean % Change	Initial	Term.	Mean % Change	# Sessions Attended
T10[4,6]	6(X)	16	(19)[a]	−18.8	51	—[b]		17	—		1
(*N* = 9)	13	22	27	−22.7	58	63	8.6	16	9	43.8	16
	18	33	6	81.8	48	76	58.3	18	4	77.8	16
	25	29	15	48.3	59	66	11.9	14	8	42.9	16
	33	24	27	−12.5	42	60	42.9	17	12	29.4	16
	50	38	44	−15.8	40	39	−2.5	24	28	−16.7	16
	56	32	19	40.6	56	59	5.4	15	15	.0	16
	81	27	4	85.2	54	79	46.3	15	2	86.7	16
	84	21	11	47.6	56	80	42.9	15	7	53.3	17
	97	37	36	2.7	46	48	4.3	20	16	20.0	4?
	104(X)	27	(7)	74.1	50	—		19	—		2
Means		29	21	28.4	51	63	24.2	17	11	37.5	15?
T1[1,6,7]	21	28	28	.0	44	54	22.7	20	13	35.0	16
(*N* = 13)	23	29	36	−24.1	51	55	7.8	17	17	.0	16
	31	33	0	100.0	50	80	60.0	19	0	100.0	16
	42	37	16	56.8	44	61	38.6	18	8	55.6	16
	47	19	4	78.9	59	80	35.6	14	0	100.0	16
	48	30	24	20.0	39	55	41.0	27	13	51.9	16
	66	31	4	87.1	54	79	46.3	15	2	86.7	16
	80	32	21	34.4	50	68	36.0	20	12	40.0	16
	89	28	26	7.1	40	40	0.0	20	19	5.0	8?
	90	30	2	93.3	51	81	58.8	19	1	94.7	16
	102	39	1	97.4	51	78	52.9	28	3	89.3	16
	109	34	27	20.6	39	46	17.9	17	17	.0	16
	113	22	4	81.8	55	68	23.6	20	5	75.0	16
Means		30	15	50.3	48	65	33.9	20	8	56.4	16?
T12[4,6]	9	43	16	62.8	43	49	14.0	19	9	52.6	16
(*N* = 13)	11	35	18	48.6	45	60	33.3	22	15	31.8	16
	14	15	2	86.7	54	81	50.0	14	1	92.9	16
	36	19	13	31.6	45	62	37.8	18	6	66.7	16
	58	33	32	3.0	45	65	44.4	19	7	63.2	16
	70	21	8	61.9	54	69	27.8	14	5	64.3	18
	72	22	5	77.3	56	74	32.1	14	4	71.4	18
	78	32	12	62.5	41	61	48.8	24	10	58.3	17
	85	40	1	97.5	50	80	60.0	16	0	100.0	14?
	93	47	47	.0	36	—		27	—		9
	96	27	31	−14.8	40	38	−5.0	19	25	−31.6	5?
	103	33	2	93.9	59	85	44.1	14	2	85.7	16
Means		29	13	55.5	48	66	35.2	18	8	59.6	15?
T18	5	23	14	39.1	47	60	27.7	19	11	42.1	16
(*N* = 4)	17	25	6	76.0	49	75	53.1	19	4	78.9	16
	28	29	21	27.6	54	54	.0	15	15	.0	16
	41	20	11	45.0	53	70	32.1	15	5	66.7	16
Means		24	13	46.9	51	65	28.2	17	8.8	46.9	16
Overall mean		28	16	45.3	50	65	30.4	18	9	50.1	16

Note: (X) = dropped out after only one or two sessions.
[a]Numbers in parentheses were taken from the last session which the patient attended.
[b]Dashes (—) indicate that this information was not available/does not exist. Patients with missing data were not included in any of the tabulations.

improved the most on the BDI in these samples (45% and 55%, respectively, in each sample). The high ranking of these therapists across the different samples and the fact that their caseloads were randomly assigned strongly suggested that it is the therapist's capacities that are responsible, not differences in the patient samples. The across-sample replication of top performance by these two therapists in the ordering of their performance is unique in reported studies and supports the conclusion that their performance reflects a therapist quality. A finding much like it has been included in a study by Blatt et al. (1996), who note that two therapists in the NIMH-sponsored Treatment of Depression Collaboration Research Program who were in the imiprimine-clinical management condition achieved a high level of therapeutic efficacy, and then these two therapists also achieved a high level of therapeutic efficacy in the placebo-clinical management condition.

Such differences in therapist effectiveness can have a significant impact on the findings in comparative treatment studies, as shown by Shapiro, Firth-Cozens, and Stiles (1989). They found that one among their three therapists was responsible for most of the reported advantage of prescriptive exploratory treatment.

A convenient way to estimate a therapist's efficacy is by ranking each therapist within the sample on patient mean percentage improvement and other outcome measures. These results show perfect agreement for most samples, and for some, only fair to moderately good agreement on outcomes. For example, for Sample 6 the four therapists had identical rankings for their mean percentage improvement in their caseloads on all three outcome measures: BDI, GAS, and HDRS. The same was true for Samples 1 and 2, which showed identical rankings for mean percentage improvement on BDI and ASI.

So far we have offered several bases that, taken together, imply that therapist differences in outcomes of their caseloads are meaningful: (a) the large range of percentage changes and effect sizes for the therapists in each sample; for example, for Sample 1 with BDIs the percentage changes are from 7% to 58% and the effect sizes are from .11 to 1.14; (b) the ranks for the outcomes tend to be similar for different outcome measures; (c) the three therapists who took part in different samples ranked similarly across samples; and (d) the therapists whose patients performed best and least show some meaningful therapist process differences consistent with the numerical differences; these analyses will be reported more completely in other reports.

It is to be noted that in the more recent treatment studies, therapists were selected to carry out the treatment according to a manual. Such a selection may serve to select the more competent and consistent therapists and reduce therapist differences. In contrast, it is likely that in ordinary clinical practice differences in competence may be even greater than in research studies. There already is a review of studies (Crits-Christoph et al., 1991) with evidence that differences in therapist performance tend to be only moderate and smaller in studies that use manuals than in the studies that do not—this surely extends the implications out of the academic and into the community arena.

Another value for these data on outcomes of therapists' caseloads is that they focus us on the task of trying to describe and to explain the differences. The observed differences in therapist efficacy may come from several bases: (a) some therapists may engage and retain a high proportion of assigned patients, while showing only average benefit in their caseload; and (b) some of the "best" therapists may do well with a small proportion of their patients, while failing to engage or retain many of them. The data available from Sample 6 (Table 9), showing that two therapists with the highest efficacy had a relatively low rate of drop out and negative response, suggests that neither (a) nor (b) fit as well as (c): the therapists whose caseloads showed the most improvement were also the therapists who were able to engage and retain the largest proportion of assigned patients. A set of related findings can also be thought of as supportive of (c): we have found repeated evidence from our work with the HAq-I (Luborsky, 1994; Luborsky et al., 1985) that the most effective therapists are rated by their patients, even after a few sessions, as being helpful and part of an alliance with them.

CURRENT LIMITS AND FUTURE DIRECTIONS

The research presented in this article has certain limits in its data and in its analyses: it would be desirable to have more information about the range of outcomes within each caseload and more correlations of therapists' outcomes of their caseloads with other variables; it would also be helpful to have more samples of therapists' perfor-

mance with their caseloads and larger caseloads for each therapist. Although we tried hard to match caseloads, there is room for more and better matching. Finally, the most novel method, of repeating the same therapist in different caseloads, should be expanded in further research. A review of these many limits suggests several recommendations for future studies.

The large range and variation in case outcomes within each therapist's caseload need to be better understood. So far we have examined this within-caseload variation closely only for one of the larger samples, Sample 6, where we could relatively easily evaluate the extent of the range within each caseload (Table 9). We see that just because a therapist is a good therapist does not mean that all the patients in the caseload do well; the range and variation are huge. But the better therapists do better with most of the patients they treat.

Therapist characteristics and behaviors that correlate with the therapist's effectiveness need more study. Only three types of variables were examined for Samples 1–3 and all three were found to be related to therapists' differences in effectiveness: (a) the therapist qualities as rated by their peers, (b) the helping alliance as judged by their patients, and (c) the therapists' adherence to their treatment manual. Clearly, there is still much to be learned. The next agenda should be to expand our knowledge of these and other correlates of differences in therapist success with the patients in their caseloads.

The basis for consistency in rankings of each therapist's mean effectiveness across samples needs to be studied. One direction in which to increase our knowledge is suggested by our observation of the striking difference in the treatment techniques of the two outstanding therapists: T1[1,6,7] used mostly supportive techniques, while T12[4,6] used mostly expressive techniques. This difference implies that there may be quite different routes to successful performance. Yet, even though the technical routes differed, the two most effective therapists showed one common trait: a substantial alliance with their patients.

The basis for choosing therapists needs more studies. Another implication for the future use of these findings is for the selection of therapists for research studies as well as for clinical practice. In selecting therapists, information should be collected on the outcomes of the therapist's treatments. It may not be sufficient to do what has been done in past research studies, that is, to only examine the therapist's behavior in the sessions in terms of adherence of the therapist to the treatment manual. It may well be more rewarding to choose therapists who also have a record of better outcomes with their caseloads, comprising six to eight cases at least.

ACKNOWLEDGMENTS

This work was partially supported by Research Scientist Award DA 00168 from the National Institute on Drug Abuse to Lester Luborsky, and NIMH Clinical Research Center Grant P50 MN 45178 to Paul Crits-Christoph.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Beck, A. T. Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory to measure depression. *Archives of General Psychiatry, 4,* 561–571.

Bergin, A. E. (1997). Neglect of the therapist and the human dimensions of change: A commentary. *Clinical Psychology: Science and Practice, 4,* 83–89

Berman, J., & Norton, N. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98,* 401–407.

Beutler, L., Machado, P., & Neufeldt, S. (1994). Therapist's variables. In A. Bergin & S. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 229–269). New York: Wiley.

Blatt, S. J., Sanislow, C. A., Zuroff, D., & Pilkonis, P. (1996). Characteristics of the effective therapist: further analysis of data from the NIMH TDCRP. Unpublished manuscript.

Crits-Christoph, P., Barnackie, K., Kurcias, J., Beck, A. T., Carroll, K., Perry, K., Luborsky, L., McLellan, A. T., Woody, G., Thompson, L., Gallagher, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91.

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparison studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59,* 20–26.

Derogatis, L. R. (1983). *SCL-R-90: Administration, scoring, and procedures manual–II.* Towson, MD: Clinical Psychometric Research.

Diguer, L., Barber, J. P., & Luborsky, L. (1993). Three concomitants: Personality disorder, psychiatric severity and outcome of dynamic psychotherapy of major depression. *The American Journal of Psychiatry, 150,* 1246–1248.

Eysenck, H. (1959). *The manual of the Maudsley Personality Inventory.* London: University of London Press.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56–62.

Holt, R. R., & Luborsky, L. B. (1958). *Personality patterns of psychiatrists: A study in selection techniques* (Vol. 2). Topeka: The Menninger Foundation.

Horvath, A., & Symonds, B. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38,* 139–149.

Luborsky, L. B. (1952). The personality of the psychotherapist. *Menninger Quarterly, 6,* 1–6.

Luborsky, L. (1962). The patient's personality and psychotherapeutic change. In H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy* (Vol. 2, pp. 115–133). Washington, DC: American Psychological Association.

Luborsky, L. (1994). Therapeutic alliances as predictors of psychotherapy outcomes: Factors explaining the predictive success. In A. Hovarth & L. Greenberg (Eds.), *The working alliance—Theory, research, and practice* (pp. 38–50). New York: Wiley.

Luborsky, L., Crits-Christoph, P., McLellan, T., Woody, G., Piper, W., Imber, S., & Liberman, B. (1986). Do therapists vary much in their success? Findings from four outcome studies. *American Journal of Orthopsychiatry, 56,* 501–512.

Luborsky, L., Diguer, L., Barber, J. P., Cacciola, J., Moras, K., Schmidt, K., & DeRubeis, R. (1996). Outcomes of short-term dynamic psychotherapy for chronic versus non-chronic major depression. *Journal of Psychotherapy Research and Practice, 5,* 152–159.

Luborsky, L., Diguer, L, Luborsky, E., McLellan, A. T., Woody, G., & Alexander, L. (1993). Psychological health-sickness (PHS) as a predictor of outcomes in dynamic and other psychotherapies. *Journal of Consulting and Clinical Psychology, 61,* 542–548.

Luborsky, L., Mark, D., Hole, A. V., Popp, C., Goldsmith, B., & Cacciola, J. (1995). Supportive-expressive dynamic psychotherapy of depression: A time- limited version. In J. P. Barber & P. Crits-Christoph (Eds.), *Psychodynamic psychotherapies for psychiatric disorders (Axis I)* (pp. 13–42). New York: Basic Books.

Luborsky, L, McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42,* 602–611.

Luborsky, L., Woody, G., Hole, A. V., & Velleeo, A. (1995). Supportive- expressive dynamic psychotherapy for treatment of opiate drug dependence. In J. P. Barber & P. Crits-Christoph (Eds.), *Psychodynamic psychotherapies for Axis I disorders* (pp. 131–160). New York: Basic Books.

McLellan, A. T., Luborsky, L., Cacciola, J., Griffith, J., McGahan, P., & O'Brien, C. (1985). *Guide to the Addiction Severity Index: Background, administration and field testing results* (NIDA Treatment Research Report, Monograph 30, PHHS Publication No. ADM 85–1419). Washington, DC: National Institute on Drug Abuse.

McLellan, A. T., Woody, G., Luborsky, L., & Goehl, L. (1988). Is the counselor an "active ingredient" in methadone treatment? An examination of treatment success among four counselors. *Journal of Nervous and Mental Disease, 176,* 423–430.

Najavits, L., & Strupp, H. (1994). Differences in the effectiveness of psychodynamic therapists: A process- outcome study. *Psychotherapy, 31,* 114–123.

Orlinsky, D., & Howard, K. (1980). Gender and psychotherapeutic outcome. In A. M. Brodsky & R. T. Hare-Muslin (Eds.), *Women and psychotherapy* (pp. 3–34). New York: Guilford.

Parloff, M., Waskow, I., & Wolfe, B. (1978). Research on therapist variables in relation to process and outcome. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (2nd ed., pp. 233–282). New York: Wiley.

Ricks, D. F. (1974). Supershrink: Methods of a therapist judged successful on the basis of adult outcome of adolescent patients. In D. Ricks, M. Roff, & A. Thomas (Eds.), *Life history research in psychopathology* (Vol. 3). Minneapolis: University of Minnesota Press.

Shapiro, D. A., Firth-Cozens, J., & Stiles, W. (1989). A question of therapist's differential effectiveness: A Sheffield psychotherapy project addendum. *British Journal of Psychiatry, 154,* 383–385.

Shipley, B. (1940). *The Shipley Institute of Living scale manual.* Hartford, CT: Neuropsychiatric Press.

Smith, M., Glass, G., & Miller, T. (1980). *The benefits of psychotherapy.* Baltimore: The Johns Hopkins University Press.

Spitzer, R. L., Gibbon, M., Williams, J. B., & Endicott, J. (1995). Global Assessment of Functioning (GAF) scale. In L. I. Sederer & B. Dickey (Eds.), *Outcomes assessment in clinical practice* (pp. 76–78). Baltimore: Williams & Wilkins.

Svartberg, M., & Stiles, T. (1992). Predicting patient change from therapist competence and patient-therapist complementarity in short-term anxiety- provoking psychotherapy. *Journal of Consulting and Clinical Psychology, 60,* 304–307.

Woody, G., Luborsky, L., McLellan, A. T., O'Brien, C., Beck, A. T., Blaine, J., Herman, I., & Hole, A. V. (1983). Psychotherapy for opiate addicts: Does it help? *Archives of General Psychiatry, 40,* 639–645.

Woody, G., Luborsky, L., O'Brien, C., & McLellan, A. (1989). Individual psychotherapy for substance abuse. In H. Kleber (Ed.), *American Psychiatric Association handbook on psychiatric treatments* (pp. 1417–1430). Washington, DC: American Psychiatric Press.

Woody, G., McLellan, A. T., Luborsky, L., & O'Brien, C. (1995). Psychotherapy in community methadone programs: A validation study. *American Journal of Psychiatry, 152,* 1302–1308.

Received December 11, 1995; revised February 29, 1996; accepted March 21, 1996.