

Waiting for Supershrink: An Empirical Analysis of Therapist Effects

John Okiishi,¹ Michael J. Lambert,^{1*} Stevan L. Nielsen¹
and Benjamin M. Ogles²

¹Brigham Young University, UT, USA

²Ohio University, OH, USA

Improving the effects of psychotherapy has been accomplished through a variety of methods. One infrequently used method involves profiling patient outcomes within therapist in order to find the empirically supported psychotherapist. This study examined data collected on 1841 clients seen by 91 therapists over a 2.5-year period in a University Counseling Center. Clients were given the Outcome Questionnaire-45 (OQ-45) on a weekly basis. After analysing data to see if general therapist traits (i.e. theoretical orientation, type of training) accounted for differences in clients' rate of improvement, data were then analysed again using Hierarchical Linear Modeling (HLM), to compare individual therapists to see if there were significant differences in the overall outcome and speed of client improvement. There was a significant amount of variation among therapists' clients' rates of improvement. The therapists whose clients showed the fastest rate of improvement had an average rate of change 10 times greater than the mean for the sample. The therapists whose clients showed the slowest rate of improvement actually showed an average increase in symptoms among their clients. Use of this information for improving quality of patient outcomes is discussed. Copyright © 2003 John Wiley & Sons, Ltd.

INTRODUCTION

In 1974 D.F. Ricks described an exceptional therapist whom the author called 'supershrink'. This therapist, who worked with highly disturbed adolescents, demonstrated exceptional outcomes when the boys were later examined as adults. In contrast to 'supershrink' another therapist, later labelled 'pseudoshrink' by Bergin and Suinn (1975), was the antithesis of supershrink. The boys treated by this therapist had adult adjustment that was alarmingly

poor. In the decades since this report there has been little research carried out on the effects of the individual therapist (Lambert & Okiishi, 1997). Instead, researchers turned their attention to the search for effective psychotherapies rather than effective therapy providers (Task Force, 1995).

A focus on therapies makes good theoretical sense and typifies the approach of academics who are interested in identifying effective treatments and developing theories of change. In the long run such studies can be highly useful by helping to identify uniquely effective treatments. In the applied world, however, such studies make less sense and are inadequate for improving the quality of patient care as the treatment is being offered. Research about effective treatments generally

* Correspondence to: Professor Michael J. Lambert, Department of Psychology, Brigham Young University, Provo UT 84602, USA.
E-mail: mike_lambert@byu.edu

assumes that the individual therapist is a relatively unimportant part of the outcome equation, rather than the central figure that facilitates patient improvement. Clinical trials research attempts to minimize the contribution of individual therapists (through standardizing treatment with treatment manuals, adherence assessments, and supervision) in order to discover the impact of specific techniques on outcome. Interpretation of such research thus overemphasizes the importance of theory-based techniques in the clinical setting, and essentially ignores the contribution of the individual therapists.

Unfortunately, the central place of theory-based interventions in clinical training and in clinical practice overshadows possibly more important therapist-based contributions often subsumed under the general rubric of common factors (Steering Committee, 2002). Within the general domain of quality management in routine practice it may make more sense to study the effects of particular therapists and then develop interventions to enhance patient outcomes. Referring patients to the most effective therapists and offering feedback and training to therapists with less than adequate patient outcome could further enhance positive client change.

Although it is clinically assumed that some therapists are better than others at facilitating change (Albert, 1997; Jennings & Skovholt, 1999), very little is known about the actual outcomes of individual therapists. Orlinsky and Howard (1980), in a retrospective study, examined the outcome of 143 female clients seen by 23 therapists. Six of the 23 had clients whose outcome was outstanding, with none deteriorating. On the other hand, five of the 23 therapists were labelled 'X-rated', that is, their clients showed low improvement rates and more than 10% were worse at termination. Additional attempts to measure the outcome of individual therapists have been reported in the literature (e.g. Luborsky, McClellan, Woody, O'Brien, & Auerbach, 1985). In a summary of this literature Crits-Christoph and Mintz (1991) conducted a meta-analysis of 15 studies on the contribution of the individual therapist and concluded that the effect of individual therapists was negligible in some studies and so large in others that it accounted for a significant amount of the variance. Such findings argue for the importance of examining individual therapist outcome, but the use of research to enhance the outcome of clients while being seen by a therapist has rarely been reported (Lutz, 2002).

The following study was conducted in order to clarify the extent of variability in outcome as a function of individual therapists and to explore the feasibility of improving patient outcome within a clinic through quality management efforts. It took place within a clinic that eventually developed an experimental outcome management system. The results of experimental investigations into the effects of this feedback system have been reported on elsewhere (Lambert *et al.*, 2001, 2002; Whipple *et al.*, 2003). The studies involved collecting weekly ratings of client treatment response and providing therapists with progress graphs and messages about the extent to which clients were recovering as expected. Results indicated that feedback to therapists about likely treatment failure reduced premature termination and enhanced outcomes for patients who were predicted to be treatment failures. The studies showed that deterioration rates could be reduced and that success rates could be increased if feedback on progress was provided to therapists. Although this quality management practice proved to be consistently effective, it was apparent that the kind of data generated from such studies could also be used to assess the effects of individual providers and then be used to make treatment assignment decisions and programmatic interventions for therapist training. Data for the current study was collected prior to implementation of the feedback studies but examined similar weekly outcome ratings (that were not shared with the therapists).

The purpose of this study was to use this archival data to examine client outcome within therapists in a treatment as usual circumstance. The study addressed two questions: Do clients show similar outcomes across therapists? And, are therapists equally efficient? Obviously the outcome for clients attending a treatment clinic can be enhanced if the most effective therapists see the majority of clients. In addition, the sooner a client's symptoms abate the better for the client; and within contexts where treatment costs are considered, faster improvement has economic implications for those who are paying for the treatment.

METHOD

Participants

Clients

The client sample for this study consisted of college students seen for individual psychotherapy at a large university counselling centre. Treatment

was available to full-time students of the University. Clients at the centre presented with a wide range of problems from simple homesickness to personality disorders. The most common diagnoses in the final dataset were mood disorders ($n = 614$, 34.5%), anxiety disorders ($n = 372$, 20.89%) and adjustment disorders ($n = 305$, 17.17%).

Students were initially seen in a 30-min intake interview and then assigned to a particular therapist based on the client's needs and therapist availability. No experimental control was exercised over this routine assignment procedure. Although there was a 14-session 'limit' at the centre, clients were allowed to be seen for more sessions if their therapist felt that it was necessary. The range of sessions in this sample was 1 to 69, with a mean of 5.16 ($SD = 7.2$).

Although 3266 students were seen at the centre over the 2-year period of data collection, the analysis being used for this study included only clients with at least two data points. This selection criterion yielded a data set of 1841 clients. Demographic information was collected on clients and is shown in Table 1.

Therapists

Ninety-one therapists contributed data to the entire data pool of 3266 clients. Therapists who had data for fewer than 15 clients were excluded from the study. Using this selection criterion as well as

the three data-point minimum described above, the final sample consisted of 56 therapists who had seen a total of 1779 clients. Data were also collected on a variety of therapist variables: level of training (pre-internship, internship and post-internship), type of training (clinical psychology, counselling psychology and social work), sex and primary theoretical orientation (cognitive-behavioural, humanistic, and psychodynamic). A summary of these therapist variables is shown in Table 2. The modal therapist was a male, licensed, counselling psychology PhD, who identified their primary theoretical orientation as cognitive-behavioural/behavioural.

Measure

Client progress in this study was tracked using the Outcome Questionnaire (OQ), a 45-item self-report measure developed specifically for the purpose of tracking and assessing client outcomes in a therapeutic setting (Lambert *et al.*, 1996). The OQ is a well-established instrument that has been validated across the USA and in Germany and across a broad range of normal and client populations. Lambert *et al.* (1996) reported an internal consistency for the OQ of 0.93 and a 3-week test-retest value of 0.84, both of which are considered adequate. Concurrent validity figures were calculated by comparing the OQ total score with total scores from other measures including the SCL-90 (Derogatis, 1983), BDI (Beck, Steer, & Garbin, 1988), Zung Depression Scale (Zung, 1965) and the STAI (Spielberger, 1983). All of the concurrent validity

Table 1. Client demographics

	N	Percentage
Sex		
Male	614	34.49
Female	1165	65.51
Race		
Caucasian	1548	87.01
Hispanic	77	4.31
Asian-American	58	3.24
Native American	16	0.88
Other ethnicity	80	4.56
Diagnosis		
Mood Disorder	614	34.50
Anxiety	372	20.89
Adjustment	305	17.17
Psychotic	27	1.54
Eating Disorder	109	6.12
Axis II	31	1.76
Sexual	56	3.13
Other	113	6.33
Deferred	152	8.56
Initial OQ Score	Mean 73.00	SD 18.57 Range 0-176

Table 2. Therapist characteristics

	N	Percentage
Experience level		
Pre-Internship	20	35.72
Intern	10	17.85
Post-Internship	26	46.43
Type of training		
Clinical Psychology	16	28.57
Counselling Psychology	38	67.86
MSW	2	3.57
Primary theoretical orientation		
Cognitive Behavioural/Behavioural	23	41.07
Humanistic/Existential	19	33.93
Psychodynamic/Interpersonal	14	25.00
Sex		
Male	34	60.71
Female	22	39.29

figures with the OQ and these instruments were significant at the 0.01 level with a range of *r*s from 0.50 to 0.85. The OQ has been shown to be sensitive to change over treatment and relatively stable over time when symptomatic individuals are not treated (Vermeersch, Lambert, & Burlingame, 2000).

The OQ is scored using a 5-point scale (0 = never, 1 = rarely, 2 = sometimes, 3 = frequently, 4 = almost always), which yields a possible range of scores from 0 to 180. High scores on the OQ indicate more distress and as clients improve scores decrease. Although not used in this study, the OQ has three subscales that measure quality of interpersonal relations, social role functioning and symptom distress. The total score, which provides a global assessment of functioning, was used in this study. As often as possible, the OQ was administered to clients before each session. However, the statistical methods being used in this study do allow for missing values and collection of data at variable times. These methods will be discussed in more detail in the following section.

Procedures

In addition to providing information about the progress of the clients, each OQ contained information making it possible to identify the client, the therapist and the date of the session. The OQ, administered before each psychotherapy session, is used as a routine part of accountability and quality improvement procedures at the counselling centre. After these data were collected, they were entered into a data base that also contained information gathered at intake including the clients' ages, gender, and ethnicity. Therapist identities were protected by randomly assigning each therapist a number so that identifying individual therapists would be impossible by viewing the data set. Client identities were similarly protected.

Data Analysis

Hierarchical Linear Modelling (HLM)

Data were analysed using a statistical technique called multi-level modelling or hierarchical linear modelling (HLM). The computer program used for this analysis was HLM for Windows (Bryk, Raudenbush, & Congdon, 1996). HLM has been demonstrated to have a number of advantages over other multivariate repeated measures methods (Bryk & Raudenbush, 1992). HLM

actually represents individuals' growth rather than being based on the interaction of repeated observations for a group of subjects. It is more flexible in its data requirements and the data are nested, which allows for the study of organizational effects on growth. Multivariate repeated measures techniques do not offer these advantages. HLM is ideal for a naturalistic study, such as this, in that it accounts for missing and erratic data.

Before the main analyses were performed, an initial HLM was carried out taking into account therapist variables (i.e. level of training, primary theoretical orientation). This was done in order to determine if some therapist variable other than the individual therapists themselves might be responsible for differences in clients' outcome. It was not anticipated that any of these variables would contribute significantly to the outcome of a client, however in the case that they did, this possibility needed to be considered before drawing conclusions about variance between therapist outcomes.

Following this initial check of therapist variables, HLM information was used in two ways. First, OQ scores for clients were used to generate improvement curves for each client and then for all clients within each therapist. Once these combined recovery curves were computed it was possible to see if there was significant variation among therapists on their clients' initial OQ scores to answer the question 'do some therapists see patients whose average initial disturbance is greater than other therapists?' Using these same curves, it was also possible to examine the rate at which clients' OQ scores decrease in order to answer the question 'are there some therapists whose clients improve significantly more or less rapidly than others?'

The second way HLM information was used was to examine how much change clients experienced on average when seeing a particular therapist. This was accomplished by taking the average rate of change for a therapist's clients and multiplying it by the average number of sessions for that therapist. Although HLM information about clients' rate of change is important, this method provided an analysis to answer the question: 'when average treatment duration is considered, do some therapists' clients improve more than others?'

RESULTS

As seen in Table 3, clients' initial OQ scores had an intercept of 73.79 ($p < 0.0001$). Similarly, the

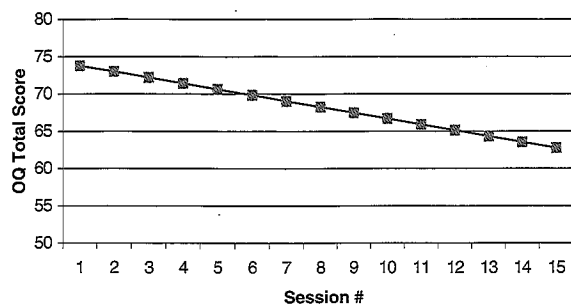


Figure 1. HLM growth curve for the entire client sample

slopes, or rates of improvement for clients showed significant variation ($p < 0.001$), with a mean slope of -0.79 . The negative slope indicates a decrease in OQ points (i.e. a lessening of endorsed client symptomatology). These findings suggest that both the slope and intercept are necessary to describe the growth trajectory for this sample (Bryk & Raudenbush, 1992). Figure 1 represents this growth curve graphically.

As can be seen in the Table 3, the HLM analysis of this sample indicated that clients within therapists differed significantly on both their initial

OQ scores (HLM intercept; $p < 0.001$) and in their rates of improvement (HLM slope; $p < 0.001$). This suggests that clients showed a broad range of initial endorsed symptomatology and varying rates of improvement.

Therapist Variables

The initial analysis of therapist data were aimed at seeing if therapists differed from one another on outcomes based on four therapist variables: type of training, amount of training, theoretical orientation and gender. An HLM was performed using these four variables as predictors. This analysis indicated that none of these therapist variables differentially contributed to the outcomes of patients. Results of this analysis are shown in Table 4. These results indicate that the type of training (counselling psychology, clinical psychology, social work), the years of training (pre-internship, internship, post-internship), theoretical orientation (cognitive behavioural, humanistic, psychodynamic) and gender did not change the slope of improvement across therapists. This suggests that any differences found between individual therapists are the result of other therapist variables.

Table 3. HLM modelling of change in OQ scores for the entire sample

	Coefficient	SE	T-ratio	SD	VAR	df	Chi-square	p-value
Fixed Effect: Client Recovery Line								
Intercept	73.787	0.295	249.94					0.001
Slope	-0.79	0.060	-13.10					0.001
Random Effect: Variation of clients within therapists								
Intercept				14.389	207.04	1778	9629.09	0.001
Slope				1.156	1.338	1778	3504.65	0.001
Random Effect: Variation among therapists								
Intercept				0.481	0.231	55	48.095	>0.50
Slope				0.240	0.057	55	80.567	0.014

Table 4. HLM with experience level, sex, level of training and theoretical orientation as predictors

Fixed effect	Coefficient	Standard Error	T-ratio	p-value
Experience	-0.108273	0.061340	-1.765	0.083
Sex	-0.038107	0.118145	-0.323	0.748
Training	0.012136	0.110959	0.109	0.914
Orientation	0.052142	0.070472	0.740	0.463

Differences Between Therapists

After the general HLM growth curve was generated for all the participants in the study and it had been determined that the therapist characteristics examined did not account for variation in outcomes, growth curves were generated based on all of the clients in each therapist's caseload. Using this client growth curve data, curves were generated for each of the therapists in the study in order to compare therapists' outcomes to the general growth curve and to each other.

In the HLM analysis for the entire centre, clients showed significant variation in initial intercept. As seen in the 'variation among therapists' portion of Table 3, when clients were grouped by therapist, there was no significant variation among therapists' clients on their average initial OQ score ($p > 0.50$). This indicates that although clients differed significantly from one another initially and represented a wide range of initial symptomatology when looked at as a whole, therapists could not be shown to have unequivocal caseloads at intake.

Again referencing the 'variation among therapists' section of Table 3, the HLM analysis indicated that therapists' clients differed significantly on their rate of change. The therapists' growth curve slopes (i.e. rate at which their clients' growth curves moved in a negative direction, indicating less endorsed symptomatology), showed a wide range of variability ($p < 0.001$). This finding suggests differential recovery rates among clients depending on which therapist they were assigned to see. Slopes, intercepts and average number of sessions for all the therapists in the sample are shown in Table 5. (It should be noted that in this table '0' session refers to the intake session. This is why each therapist has a minimum number of sessions at zero.) This is also why it is possible for a therapist to have an average session length of less than one.

For example, the therapist (no. 1) whose clients demonstrated the steepest HLM improvement slopes had an average drop of 7.97 OQ points per session. This is 10 times the average rate of change for the entire centre. The therapist whose clients showed the least improvement (no. 56) actually had a positive slope, indicating that on average clients were endorsing more pathology at the end of treatment than when they entered treatment. This therapist's clients endorsed a higher degree of pathology at a rate of 0.68 points per session. A graph of these two therapists' growth curves and

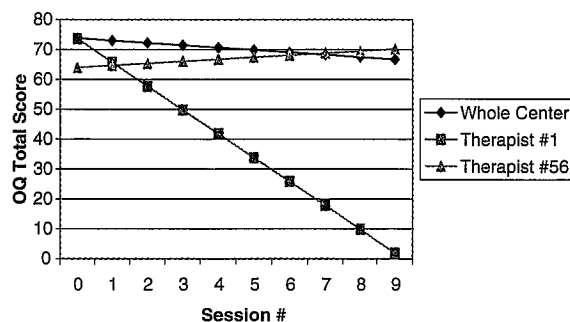


Figure 2. HLM growth curves for best, worst and whole centre therapist samples

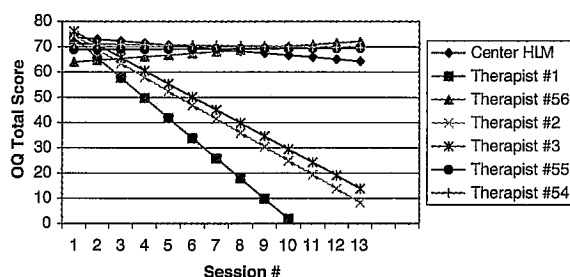


Figure 3. HLM growth curves for top three versus bottom three therapists

the growth curve for the entire centre is shown in Figure 2. As can be seen by examining this graph, the difference in rate of change between these two therapists is dramatic.

The graph of therapist 1 and therapist 56 provides a dramatic example of differences between individual therapists. However, it is not necessary to go to these extremes in order to demonstrate the clear differences in therapist outcomes. Figure 3 is a graph of the top three therapists versus the bottom three therapists. Again, there is a stark contrast between the two groups.

Even though these therapists had dramatic differences in the rate at which their clients changed, it is important to note the difference in the average number of sessions for each of the therapists. As shown in Table 5 and Figure 2, the average number of sessions for therapist 1 was 0.8 (SD = 1.88) with a maximum number of 10 sessions (note: the session length of 0.8 is possible because intake sessions are considered session '0'). This differs greatly from the 8.46 sessions (SD = 9.28) and 36 maximum sessions of therapist 56. Therapist 1

Table 5. Therapist mean sessions, HLM slopes and HLM intercepts for whole sample

Therapist	Mean no. of Sessions	Session SD	Minimum sessions	Maximum sessions	HLM slope	HLM intercept
Whole Center	4.47	6.89	0	69	-0.79	73.79
1	0.80	1.88	0	10	-7.97	73.67
2	3.77	3.72	0	16	-5.51	74.46
3	2.85	2.72	0	13	-5.19	76.12
4	3.44	2.58	0	10	-5.07	74.93
5	2.78	2.63	0	12	-4.85	76.68
6	2.50	3.34	0	17	-4.21	73.48
7	2.88	4.42	0	22	-4.06	77.09
8	4.55	6.28	0	28	-3.96	70.88
9	3.64	4.49	0	17	-3.57	74.71
10	2.60	2.85	0	14	-3.43	74.59
11	3.15	3.84	0	12	-3.18	77.28
12	4.59	3.12	0	12	-3.14	67.65
13	2.67	3.22	0	15	-3.10	72.30
14	4.03	5.28	0	32	-3.08	77.52
15	4.43	3.79	0	18	-2.98	75.74
16	2.34	2.07	0	9	-2.93	74.09
17	3.87	3.15	0	13	-2.90	78.09
18	4.87	5.21	0	26	-2.87	71.35
19	7.31	11.95	0	49	-2.65	75.67
20	3.85	5.26	0	27	-2.56	70.87
21	3.01	4.92	0	23	-2.53	73.79
22	3.36	3.04	0	14	-2.52	74.66
23	3.25	2.39	0	11	-2.49	72.61
24	3.14	2.47	0	9	-2.45	70.32
25	4.76	5.53	0	26	-2.40	72.22
26	4.49	4.73	0	22	-2.36	69.86
27	4.31	5.69	0	34	-2.02	73.41
28	4.31	4.15	0	24	-1.74	69.40
29	6.81	12.38	0	58	-1.51	68.25
30	4.04	4.24	0	20	-1.49	70.21
31	4.11	6.01	0	43	-1.48	75.04
32	5.38	6.69	0	30	-1.44	72.54
33	9.09	11.80	0	57	-1.43	76.91
34	3.46	5.33	0	31	-1.33	74.09
35	5.09	4.65	0	19	-1.33	71.50
36	9.15	7.97	0	35	-1.31	71.41
37	6.46	6.27	0	34	-1.18	69.69
38	5.81	7.00	0	49	-1.13	73.90
39	6.01	7.48	0	35	-1.11	70.81
40	4.03	4.72	0	33	-1.07	74.67
41	3.03	3.25	0	15	-1.01	70.50
42	6.81	5.92	0	25	-0.99	73.35
43	6.68	5.41	0	21	-0.99	71.58
44	4.30	3.52	0	14	-0.92	68.12
45	5.57	4.70	0	19	-0.92	68.77
46	4.83	6.04	0	28	-0.86	70.32
47	7.05	8.91	0	37	-0.83	73.55
48	6.95	8.52	0	50	-0.83	68.16
49	3.82	4.94	0	23	-0.75	74.31
50	6.51	11.04	0	69	-0.68	70.62
51	3.83	6.46	0	40	-0.67	73.94
52	5.03	8.04	0	57	-0.45	70.54
53	5.10	5.75	0	24	-0.32	67.07
54	6.17	7.76	0	44	-0.13	71.21
55	9.06	11.08	0	54	0.05	68.90
56	8.46	9.28	0	36	0.68	64.01

Table 6. One-way ANOVA examining rank ordered therapists and number of sessions for the entire sample

	Sum of squares	df	Mean square	F	p-value
Between groups	1162.011	2	581.006	11.148	0.001
Within groups	574045.877	11014	52.120		
Total	575207.888	11016			

Post-hoc comparison: Tukey HSD

Ranking group	Comparison group	Mean difference	SE	p-value	95% confidence interval Lower bound Upper bound
Top one-third	Middle	-0.82*	0.179	0.001	-1.24 -0.40
	Bottom	-0.56*	0.166	0.002	-0.95 -0.17
Middle one-third	Top	0.82*	0.179	0.001	0.40 1.24
	Bottom	0.27	0.167	0.249	-0.13 0.66
Bottom one-third	Top	0.56*	0.166	0.002	0.17 0.95

*The mean difference is significant at the 0.05 level.

clearly saw clients for a much shorter time than did therapist 56.

Examining average session length helps further clarify the picture of how these therapists practiced. The average number of sessions for the top three therapists was 2.4 (SD = 3.05) with a maximum of 16 sessions. The bottom three therapists had a mean session length of 7.05 (SD = 8.75) with a maximum of 54 sessions. As with the top and bottom individual therapists, the top three and bottom three therapists showed significantly different lengths of treatment ($t = 9.42$, $p < 0.01$).

Due to the large difference in session length between the two groups the sample was split into three groups based on the ranking of therapists by HLM slope. A one-way ANOVA was calculated in order to determine the relationship between the ranking of the therapists and their average session length based on these three groups. Results are presented in Table 6.

As can be seen in this table, there were significant differences in length of session between the top one-third and the bottom two-thirds of therapists. This finding suggests that the clients with the greatest rate of change were seen by therapists who saw them for shorter periods of time on average. Figure 4 presents therapist rankings and average session length graphically to better illustrate this finding.

Average Session Length

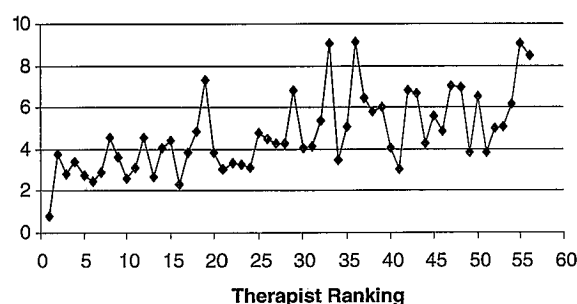


Figure 4. Scatterplot of therapist rankings and average session length

'Exceptional' Therapists: An Alternate View

In order to synthesize our findings and to clarify which therapists might be considered 'super-shrinks' and which might be 'pseudoshrinks' the average session length of individual therapists was multiplied by the slope of their HLM line. This provided an index of therapist 'exceptionality' that takes into account both the rate of change and the average amount of time spent in therapy. Results of this procedure are shown in Table 7.

Comparing Table 7 to Table 5 provides insight into who might actually be considered the 'super-shrinks' and 'pseudoshrinks' of the sample using

Table 7. Therapist average session length \times average treatment gain per session for whole sample

Therapist	Mean no. of sessions	Session SD	Maximum sessions	HLM slope	HLM intercept	Mean session \times HLM slope
Whole centre	4.47	6.89	69	-0.79	73.79	-3.5313
2	3.77	3.72	16	-5.51	74.46	-20.7727
19	7.31	11.95	49	-2.65	75.67	-19.3715
8	4.55	6.28	28	-3.96	70.88	-18.018
4	3.44	2.58	10	-5.07	74.93	-17.4408
3	2.85	2.72	13	-5.19	76.12	-14.7915
12	4.59	3.12	12	-3.14	67.65	-14.4126
18	4.87	5.21	26	-2.87	71.35	-13.9769
5	2.78	2.63	12	-4.85	76.68	-13.483
15	4.43	3.79	18	-2.98	75.74	-13.2014
33	9.09	11.8	57	-1.43	76.91	-12.9987
9	3.64	4.49	17	-3.57	74.71	-12.9948
14	4.03	5.28	32	-3.08	77.52	-12.4124
36	9.15	7.97	35	-1.31	71.41	-11.9865
7	2.88	4.42	22	-4.06	77.09	-11.6928
25	4.76	5.53	26	-2.4	72.22	-11.424
17	3.87	3.15	13	-2.9	78.09	-11.223
26	4.49	4.73	22	-2.36	69.86	-10.5964
6	2.5	3.34	17	-4.21	73.48	-10.525
29	6.81	12.38	58	-1.51	68.25	-10.2831
11	3.15	3.84	12	-3.18	77.28	-10.017
20	3.85	5.26	27	-2.56	70.87	-9.856
10	2.6	2.85	14	-3.43	74.59	-8.918
27	4.31	5.69	34	-2.02	73.41	-8.7062
22	3.36	3.04	14	-2.52	74.66	-8.4672
13	2.67	3.22	15	-3.1	72.3	-8.277
23	3.25	2.39	11	-2.49	72.61	-8.0925
32	5.38	6.69	30	-1.44	72.54	-7.7472
24	3.14	2.47	9	-2.45	70.32	-7.693
37	6.46	6.27	34	-1.18	69.69	-7.6228
21	3.01	4.92	23	-2.53	73.79	-7.6153
28	4.31	4.15	24	-1.74	69.4	-7.4994
16	2.34	2.07	9	-2.93	74.09	-6.8562
35	5.09	4.65	19	-1.33	71.5	-6.7697
42	6.81	5.92	25	-0.99	73.35	-6.7419
39	6.01	7.48	35	-1.11	70.81	-6.6711
43	6.68	5.41	21	-0.99	71.58	-6.6132
38	5.81	7	49	-1.13	73.9	-6.5653
1	0.8	1.88	10	-7.97	73.67	-6.376
31	4.11	6.01	43	-1.48	75.04	-6.0828
30	4.04	4.24	20	-1.49	70.21	-6.0196
47	7.05	8.91	37	-0.83	73.55	-5.8515
48	6.95	8.52	50	-0.83	68.16	-5.7685
45	5.57	4.7	19	-0.92	68.77	-5.1244
34	3.46	5.33	31	-1.33	74.09	-4.6018
50	6.51	11.04	69	-0.68	70.62	-4.4268
40	4.03	4.72	33	-1.07	74.67	-4.3121
46	4.83	6.04	28	-0.86	70.32	-4.1538
44	4.3	3.52	14	-0.92	68.12	-3.956
41	3.03	3.25	15	-1.01	70.5	-3.0603
49	3.82	4.94	23	-0.75	74.31	-2.865
51	3.83	6.46	40	-0.67	73.94	-2.5661
52	5.03	8.04	57	-0.45	70.54	-2.2635
53	5.1	5.75	24	-0.32	67.07	-1.632
54	6.17	7.76	44	-0.13	71.21	-0.8021
55	9.06	11.08	54	0.05	68.9	0.453
56	8.46	9.28	36	0.68	64.01	5.7528

the criteria of total OQ-45 point change. For example, therapist 2 saw clients an average of 3.77 times and the average gain for these clients was -20.77 points. Therapist 19 was a close second (although remarkably less efficient), seeing patients for 7.31 sessions while they improved an average of -19.37 points. Clients of therapists 8 and 4 also had remarkably good outcomes with average treatment gains of -18.02 and -17.44 points. Interestingly, therapist 1, who was considered to be the supershrink therapist in terms of HLM slope, fell to 38 in the rankings when the data were examined in this way (due to this therapist's low average session rate).

At the bottom end of the spectrum, therapist 56 is a clear standout as a 'pseudoshrink'. Therapist 56 saw patients for an average of 8.46 sessions (the fourth highest average) while they worsened by an average of +5.75 points. This is more than 5 points on average worse than the penultimate lowest-ranked therapist.

DISCUSSION

The purpose of this study was to systematically examine a large sample of therapists who had seen a large number of clients in order to answer the question: 'are there some therapists whose clients have better or worse outcomes than others?' This is a question that has rarely been investigated in a sample the size of this study (Lambert & Okiishi, 1997). The HLM analysis of the entire sample provided the clearest evidence that there are indeed significant differences among therapists in the outcomes of their psychotherapy clients. The HLM analysis indicated that overall, clients seen at the centre showed a broad range of initial symptomatology and varying rates of improvement. Clients had a similar level of symptomatology at intake across therapists. In fact, the similarity in degree of disturbance across therapist caseloads called into question the necessity of case-mix adjustments for the purpose of fairly contrasting therapists in this sample.

As expected based on prior research (e.g. Beutler, Machado, & Neufeldt, 1994), there were no significant differences in client outcome between therapists' based on the four therapist demographic variables of sex, level of training, type of training or theoretical orientation. This supports the idea that even though graduate school training and managed health care tend to focus on training in specific techniques, something else, perhaps the

individual therapists themselves are responsible for variation in client outcomes (Lambert & Okiishi, 1997).

Although therapists appeared to have roughly equivalent caseloads at intake, there was a significant, sometimes dramatic difference when improvement curves of clients (using session-by-session outcome ratings) within therapist caseloads were examined and compared to each other. There were also large differences between therapists when outcome was assessed based on pre-post differences. Both of these methods of defining outcomes, provided very powerful evidence that in fact some therapists' clients do have much better and worse outcomes than others.

A client seeing one of the top three therapists for their average session length of 2.47 sessions can expect to improve by about 15 points on the OQ-45. This is a reliable change according to Jacobson and Truax's (1991) criteria and amounts to a client losing about four full symptoms. A client seeing one of the bottom three therapists for their average session length of 7.89 sessions could expect to get worse by about five OQ-45 points. The difference between these best and worst outcomes is 20 OQ-45 points, which is well over a standard deviation in difference. Additionally, a person being seen by one of the best three could expect to be done with therapy at a much more rapid rate. A client seen by one of the 'best' therapists can expect to be feeling significantly better after a few weeks of treatment. A client seen by one of the 'worst' therapists can expect to feel about the same, if not worse, than when they started treatment, and this after almost three times as much treatment as those clients seeing the more efficient therapists.

The significant correlation between therapist ranking and session length provided interesting insight into the best and worst therapists. The therapists who saw clients that had the most rapid rates of improvement had significantly shorter treatment times. This makes logical sense. If someone seeking treatment improves at a more rapid rate, then they will stay in treatment for a shorter period of time. Conversely and in relation to the worst therapists, clients who experience no change or an increase in symptomatology will likely stay in treatment longer because they are not feeling better. These findings are consistent with another large data set where more symptomatic clients were shown to stay in therapy for significantly longer periods of time (Lueger *et al.*, 2001).

So does this mean that shorter treatment is better? We would argue 'no'. These data demonstrate that therapists whose clients show superior rates of change see clients on average for a shorter period of time, but short treatment time is not the determinate so much as rapid improvement of the clients. It appears likely that something about these therapists and the way they work, independent of the amount of time spent with clients, has a significant impact.

When average session length and HLM slope were looked at together a different picture of the super- and pseudoshinks emerged. With total average change for a course of therapy as the determinate of the effectiveness of a therapist, some of the more efficient therapists (those with the steepest HLM slopes) go down in the rankings and therapists who showed slower rates of improvement but kept clients in therapy longer are higher in the rankings.

Findings on the worst therapists when examined this way are more consistent with the findings using HLM slope only. The worst therapists seem to spend a lot of time with clients even though no measurable gains can be found. It could be inferred from this data that these therapists are encouraging dependency from their clients. Their clients stick around for much longer despite no changes in symptomatology.

These data could be used in several ways to enhance outcome for clients in this clinic. First, results of the study could be provided to each therapist and the implications for practice and training discussed. Some analysis of the causes for differences could be explored and lead to qualitative and process research. Intervention strategies, such as supervision by the supershrinks could be encouraged. Continued monitoring of client progress and then updating of the database are essential and could be used to test the effects of interventions with the therapists and revise judgments about who are most effective.

Second, clients who come to the centre could be referred to the most effective therapists from a rank ordered listing of therapists (based on past outcomes). Clients would be referred to the most effective therapists if available. This would mean that the least effective therapists would be given more support tasks rather than treatment tasks, thereby freeing up the most effective therapists for clinical work. Less effective therapists could spend more time in training activities, administration, or centre-related tasks for which they may be better suited.

Limitations

There are a number of limitations of this study including: (1) lack of random assignment of clients to therapists, (2) not being able to identify therapists and their traits specifically, and (3) the OQ-45 being the only measure used in the study. In this section we will discuss these limitations in detail and describe how they may have impacted the results of this study.

As stated previously, the clients in this study were not randomly assigned. At the centre where data were collected, case assignments were made by individuals (full-time staff and interns) doing intakes and were routinely assigned presumably based on the clients perceived level of pathology, difficulty, gender and goodness of personality fit with the therapist. Most importantly, therapists' availability of times to meet with new clients determined assignment. These assignments were made entirely on the basis of clinical judgment and without regard to OQ-45 scores at intake. This method of deliberate case assignment could pose a threat to the validity of the findings. Without random assignment, it was possible for a particular therapist to be given a disproportionate number of 'easy' or 'difficult' cases, thus inflating their level of effectiveness. Case-mixing was originally proposed to counter-balance this problem, but it was found that in the analysis of the sample, there were no significant differences overall in the average score among therapists and the possible benefits of case-mixing were nullified. Nevertheless, some therapists may have seen clients because of a specific specialty interest (e.g. eating disorders) which would presumably influence the outcome rates. Further examination of other client variables may be needed in order to rule out case-mix factors that are not captured by the OQ scores. At the same time the present analysis was based on archival data that were not collected with the intent of examining outcomes by therapist, a factor which would limit any tendency for experimental bias to determine assignment of cases.

Outcome was measured with the use of a single self-report measure. This does have its methodological limitations. A self-report measure only taps into what the client is willing to show to the clinician and researchers. If a client wanted to misrepresent what they were experiencing it would be as easy as filling out a few bubbles differently. It would be naïve to assume that this measure fully captured the 'reality' about a person's psychological functioning and their pattern of change.

OQ scores are the product, as are all self-report measures, of the trade-off between depth and breadth. It was possible to collect a large amount of repeated measures data across a wide range of clients, but at the cost of a fuller picture of each client's mental state. This is a trade-off that is common in quality assurance research undertaken in routine practice.

Possibly the biggest limitation of this study was our inability to examine identifying information about therapists more closely. This specific information was masked in order to protect the confidentiality of therapists and ensure that a large number of therapists were included in the study. We were able to identify that there were exceptional therapists on both ends of the spectrum, as well as how exceptional they were, but were unable to dig deeper into who they actually were or use the data for quality assurance purposes. If possible it would have been valuable to know more about what therapists actually did and how they thought and felt about their work. Knowing how these individuals were perceived by themselves, their colleagues and their clients would also provide valuable information. Client satisfaction as well as interviews with co-workers and the therapists themselves could easily yield this data if confidentiality of therapists was not an issue.

Investigation of therapist development could also be important to examine. How did these therapists become significantly better or worse than their peers? The possibilities for exploration into this topic are limitless. Although the questions we did answer were important, a closer look at the exceptional therapists we were able to identify will be important to examine in the future. Unfortunately, what these therapists did to be 'super-shrinks' and 'pseudoshrinks' remains a mystery.

Finally, examining follow-up data on the clients in this study could also be very useful. Although it is clear that clients benefited from therapy and benefited more so from receiving treatment from particular therapists, it is important to determine if these gains are maintained after therapy. In light of the finding that the most effective therapists saw clients for a shorter period of time, this is a particularly important question to explore. It could be argued that since clients were seen for a relatively short amount of time by the most effective therapists that they had a 'flight into health' rather than having true therapeutic gains. Although there has been some research that indicates that rapid treatment gains are stable at follow-up, this has never been examined while taking into account

individual therapists (Haas, Hill, Lambert, & Morrell, 2002). Follow-up on these clients post-treatment, particularly the clients of the 'super' and 'pseudo' shrinks could help us ascertain whether these gains (and losses) were stable after termination of treatment.

In the spirit of patient-focused research we encourage researchers and clinicians to enhance patient outcomes through studies that examine the treatment response of clients as a function of the individual provider. Bohart (2000) suggested that the quality of clinical services would benefit more from a research paradigm that emphasizes 'empirically supported psychotherapy practice' than one focused on 'empirically supported treatments'. We take an even stronger stance and believe that emphasizing 'empirically supported therapists' may prove even more beneficial to client outcomes. There is an urgent need to take account of the effectiveness of the individual therapist and it is time for clinicians to welcome such research.

REFERENCES

- Albert, G. (1997). What are the characteristics of effective psychotherapists? The experts speak. *Journal of Practical Psychology and Behavioral Health*, 3, 36–44.
- Beck, A.T., Steer, R.A., & Garbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: twenty-five years later. *Clinical Psychology Review*, 8, 77–100.
- Bergin, A.E., & Suinn, R.M. (1975). Individual psychotherapy and behavior therapy. *Annual Review of Psychology*, 26, 509–556.
- Beutler, L.E., Machado, P.P.P., & Neufeldt, S.A. (1994). Therapist variables. In A.E. Bergin, & S.L. Garfield (Eds), *Handbook of psychotherapy and behavior change* (4th ed.). New York: Wiley.
- Bohart, A.C. (2000). Paradigm clash: Empirically supported treatments versus empirically supported psychotherapy practice. *Psychotherapy Research*, 10, 488–493.
- Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1996). HLM for Windows (version 4.01.01) (Computer Program) SSI, Inc.
- Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, 59, 20–26.
- Derogatis, L.R. (1977). *The SCL-90 manual: Scoring, administration and procedures for the SCL-90*. Baltimore: Johns-Hopkins University School of Medicine, Clinical Psychometrics Unit.

- Haas, E., Hill R., Lambert, M.J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology, 58*, 1157-1172.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.
- Jennings, L., & Skovholt, T.M. (1999). The cognitive, emotional and relational characteristics of master therapists. *Journal of Counseling Psychology, 46*, 3-11.
- Lambert, M.J., & Okiishi, J.C. (1997). The effects of the individual psychotherapist and implications for future research. *Clinical Psychology: Research and Practice, 4*, 66-75.
- Lambert, M.J., Hansen, N.B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G.M., Heufner, J.C., & Riesinger, C.R. (1996). *Administration and scoring manual for the Outcome Questionnaire*. Wilmington, Delaware: APCS-LLC.
- Lambert, M.J., Whipple, J.L., Smart, D.W., Vermeersch, D.A., Nielsen, S.L., & Hawkins, E.J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: are outcomes enhanced? *Psychotherapy Research, 11*, 49-68.
- Lambert, M.J., Whipple, J.L., Vermeersch, D.A., Smart, D.W., Hawkins, E.J., Nielsen, S.L., & Goates, M. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: a replication. *Clinical Psychology and Psychotherapy, 9*, 91-103.
- Luborsky, L., McClellan, A.T., Woody, G.E., O'Brien, C.P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42*, 602-611.
- Lueger, R.J., Howard, K.I., Martinovitch, Z., Lutz, W., Andersen, E.E., & Grissom, G. (2001). Assessing treatment progress with individualized models of expected response. *Journal of Consulting and Clinical Psychology, 69*, 150-158.
- Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychotherapy Research, 12*, 251-272.
- Orlinsky, D.E., & Howard, K.I. (1980). Gender and psychotherapeutic outcome. In A.M. Brodsky, & R.T. Hare-Mustin (Eds), *Women and psychotherapy* (pp. 3-34). New York: Guilford Press.
- Ricks, D.F. (1974). Supershrink: methods of a therapist judged successful on the basis of adult outcomes of adolescent patients. In D.F. Ricks, M. Roff, & A. Thomas (Eds), *Life history research in psychopathology*. Minneapolis: University of Minnesota Press.
- Spielberger, C.D. (1983). *Manual for the State-Trait Anxiety Inventory: STAI (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.
- Steering Committee. (2002). Empirically supported therapy relationships: Conclusions and recommendations of the Division 29 Task Force. *Psychotherapy: Theory, Research, Practice, & Training, 38*, 495-497.
- Task Force. (1995). Training in and dissemination of empirically-validated psychological treatments. *The Clinical Psychologist, 48*, 3-23.
- Vermeersch, D.A., Lambert, M.J., & Burlingame, G.M. (2000). Outcome Questionnaire: item sensitivity to change. *Journal of Personality Assessment, 74*(2), 242-261.
- Whipple, J.L., Lambert, M.J., Vermeersch, D.A., Smart, D.W., Nielsen, S.L., & Hawkins, E.J. (2003). Improving the effects of psychotherapy: the use of early identification of treatment failure and problem solving strategies in routine practice. *Journal of Counseling Psychology, 50*, 59-68.
- Zung, W.W. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12*, 63-70.

Practice-Based Evidence: Benchmarking NHS Primary Care Counselling Services at National and Local Levels

Chris Evans,^{1*} Janice Connell,² Michael Barkham,²
Chris Marshall² and John Mellor-Clark²

¹Rampton Hospital, Nottinghamshire Healthcare NHS Trust and Tavistock & Portman NHS Trust, UK

²Psychological Therapies Research Centre, University of Leeds, Leeds, UK

There are a number of problems for evidence-based practice (EBP) including limited generalizability of efficacy research results, the consequent lack of confidence in the relevance of such research, and the conceptual distance of most practitioners from the research process. The result is that EBP, although sound in principle, often fails to achieve its aim of improving practice. Practice-based evidence (PBE) provides a complementary bridge for the gap between research and practice to offset some of these problems, promoting collaboration between mental health services and academic institutions. This paper presents the initial results of such a collaboration via three phases: (1) the development of a referential database for primary care counselling services, (2) 'practitioner-friendly' feedback on grouped data to services, and (3) the combination of the two to build an evidence base for work with ethnic minorities—an area in which research trials are not well adapted to provide much evidence. Copyright © 2003 John Wiley & Sons, Ltd.

INTRODUCTION

The evidence-based practice (EBP) paradigm has become increasingly influential in all health care. In this paradigm, primacy is generally given to evidence derived from systematic reviews, meta-

analytic reviews and randomized controlled trials (RCTs). In the UK, the output of this paradigm for psychological therapies in primary care has been considerable including high quality RCTs (e.g. Ward *et al.*, 2000), systematic reviews (e.g. Bower, Richards, & Lovell, 2001; Rowland *et al.*, 2000) and the development of practice guidelines for use in primary care (e.g. Department of Health, 2001). These products are central components in the overall plan for an appropriate knowledge base for the psychological therapies as set out in the strategic review of psychotherapy (Department of Health, 1996).

In principle the evidence-based practice paradigm should empower practitioners to answer locally pertinent questions, clinical or managerial, through critical appraisal of the best available

*Correspondence to: Dr C. Evans, Rampton Hospital, Nottinghamshire Healthcare NHS Trust, Retford, Notts. DN22 0PD, UK.

E-mail: chris@psyctc.org

Contract/grant sponsor: Artemis Trust.

Contract/grant sponsor: Mental Health Foundation.

Contract/grant sponsor: Leeds Community and Mental Health Trust.

Contract/grant sponsor: Nottinghamshire Healthcare NHS Trust.

Contract/grant sponsor: Tavistock & Portman NHS Trust.