Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: https://www.tandfonline.com/loi/tpsr20

Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data

Don-Min Kim , Bruce E. Wampold & Daniel M. Bolt

To cite this article: Don-Min Kim , Bruce E. Wampold & Daniel M. Bolt (2006) Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data, Psychotherapy Research, 16:02, 161-172, DOI: 10.1080/10503300500264911

To link to this article: https://doi.org/10.1080/10503300500264911

Published online: 22 Feb 2007.

Psychotherapy Research, March 2006; 16(2): 161–172

Routledge
Taylor & Francis Group

SPECIAL SECTION ARTICLES

Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data

DONG-MIN KIM, BRUCE E. WAMPOLD & DANIEL M. BOLT

University of Wisconsin—Madison

(Received 16 February 2004; revised 15 September 2004; accepted 23 September 2004)

Abstract
Data for completer and intent-to-treat samples from the two psychotherapy conditions of the National Institute of Mental Health Treatment of Depression Collaborative Research Program were analyzed to estimate the proportion of variability in outcomes resulting from therapists. Therapists, who were nested within treatments, were considered a random factor in multilevel analyses. These analyses, which modeled therapist variability in several different ways, indicated that about 8% of the variance in outcomes was attributable to therapists, whereas 0% was due to the particular treatment delivered. When therapist effects were appropriately modeled, previously detected differences in efficacy between the two psychotherapy conditions for more severely depressed patients disappeared, as predicted by methodological considerations.

The importance of history is often revealed not so much by what transpired but by what was left out. Asking the question, "What was absent to prevent the notion of probability occurring before the 17th century," provides fascinating insights into the philosophy of science that are unavailable by tracing the development of probability theory in and after this crucial century (see Hacking, 1975). Similarly, an examination of the history of the development of the randomized control group designs reveals that something important was left out (Wampold, 2001a; Wampold & Bhati, 2004). The major applications of the randomized control group design and the analysis of variance (ANOVA) were in education (e.g., McCall, 1923), agriculture (e.g., Fisher, 1935), and, later, medicine (see Gehan & Lamak, 1994; Shapiro & Shapiro, 1997). Research design was sold to educational administrators by psychologists, who understood that administrators possessed both money and power; the providers of the programs (viz., the teachers), predominantly low-paid women, who implemented the programs identified as effective, were considered interchangeable and unimportant (Danziger, 1990). The emphasis in agriculture was on scientific farming practices, which emphasized the agricultural methods that presumably could be applied uniformly by the farmer; consequently, variations among farmers were not considered. The double-blind design in medicine further reduced the role of the provider of services, because the physicians or physician proxies were unaware of whether they were providing an active medication or a placebo (Shapiro & Shapiro, 1997). From these applications grew a tradition of ignoring provider effects in the study of treatments, an inconspicuous but potentially critical omission. The omission was carried over to clinical trials of psychotherapy (using either crossed or nested designs), in which typically provider effects have not been modeled (Crits-Christoph et al., 1991; Crits-Christoph & Mintz, 1991; Wampold, 2001b; Wampold & Bhati, 2004).

There are two problems with ignoring provider effects. First, ignoring provider effects makes the assumption that provider effects are nonexistent and unimportant and has been the case in education, agriculture, and medicine as well as psychotherapy (Wampold, 1997, 2001a, 2001b). The research that

Correspondence: Bruce E. Wampold, Department of Counseling Psychology, 321 Education Building, 1000 Bascom Mall, University of Wisconsin–Madison, Madison, WI 53706. E-mail: wampold@education.wisc.edu

ISSN 1050-3307 print/ISSN 1468-4381 online
DOI: 10.1080/10503300500264911

has been conducted to estimate therapist effects has shown that a modest to large proportion of the variability in outcome is due to the therapist (Blatt, Sanislow, Zuroff, & Pilkonis, 1996; Crits-Christoph et al., 1991; Luborsky, McLellan, Diguer, Woody, & Seligman, 1997; Project MATCH Research Group, 1998). Estimates of the proportion of variability resulting from therapists are in the range of 6% to 10%. Given the fact that whether or not a person receives any treatment only accounts for about 13% of variability in outcomes, therapist variability, therefore, is an important factor (Wampold, 2001b).

Second, unless therapist effects are nonexistent, observations within therapists are not independent. Ignoring dependent observations in the ANOVA has consequences for the results, inflating Type I error rates as well as estimates of effect sizes produced by the examination of treatment differences (Barcikowsky, 1981; Crits-Christoph & Mintz, 1991; Kenny & Judd, 1986; Kirk, 1995; Walsh, 1947; Wampold & Serlin, 2000). The greater the effect that is due to therapists, the greater the inflation, because the observed differences among treatments are due to true differences as well as random variation among therapists (see, e.g., Wampold & Serlin, 2000). That is, unexamined therapist effects will yield liberal tests of treatment differences (accepting treatments effects when the true difference is zero) and inflated estimates of the size of the effect. The magnitude of these consequences for psychotherapy research has been studied and modeled and found to be a source of concern (Crits-Christoph & Mintz, 1991; Wampold & Serlin, 2000).

Choosing to appropriately consider therapists in the design and analysis of comparative clinical trials depends on considerations related to the therapist factor. The first decision is whether to cross therapists and treatments or to nest therapists within treatments (for a more complete discussion see Crits-Christoph & Mintz, 1991; Elkin, 1999; Wampold, 2001b). In the crossed design, all therapists provide each of the treatments, which has the advantage that the general skill level of therapists is controlled. However, the therapists may, and often do, have allegiance to and particular skill in delivering one of the treatments (Elkin, 1999; Serlin, Wampold, & Levin, 2003; Wampold, 2001b), which is problematic given the sizable allegiance effects detected in the psychotherapy literature (Luborsky et al., 1999; Wampold, 2001b). In the nested design, therapists are assigned or chosen to deliver one and only one of the treatments, thus raising the possibility that treatment and therapist skill are confounded. However, the nested design is able to control for allegiance.

The second decision relative to therapists is whether to treat therapists as a fixed factor or a random factor (for a discussion of fixed and random models in psychotherapy research see Crits-Christoph & Mintz, 1991; Siemer & Joormann, 2003; Serlin et al., 2003; Wampold & Serlin, 2000). If therapists are treated as a fixed factor, the results are conditioned on the particular therapists included in the clinical trial, thus restricting the conclusions to only those particular therapists in the trial. Although restricting the generality of the results yields an increase in power to test main effects, conclusions restricted only to a particular small set of therapists typically are unreasonable (see Serlin et al., 2003; Siemer & Joormann, 2003). More informative results are obtained from considering therapists as randomly selected from a population of therapists so that conclusions can be made about therapists in general (or at least therapists similar to those used in the study; see Serlin et al., 2003). The differences in the models are summarized by Siemer and Joormann (2003): "The crucial question is whether it is justified to treat providers as a random effect thereby seeking to generalize to a population of providers or whether one should treat providers as a fixed effect thereby restricting the inference to the providers included in that particular study, that is, to make statistical inference conditional on the set of providers included in the study" (p. 500). Although therapists are rarely randomly selected or assigned to treatment, the random-factors model is appropriate provided conclusions are limited to therapists similar to those delivering the treatment (see Serlin et al., 2003).

As Elkin (1999) has noted, interest in therapist effects has existed at least since "Kiesler's classic (1966) article, [in which] he characterizes the assumption of therapist uniformity as one of the 'myths' of psychotherapy research" (p. 11). There have been laudable attempts over the years to estimate variability among therapists, but typically those studies have examined this source of variability after the treatment effects have been published (e.g., Blatt et al., 1996; Crits-Christoph et al., 1991; Huppert et al., 2001; Luborsky et al., 1986, 1997). Although such post hoc analyses have the potential to provide viable estimates of therapist variance, the analyses have typically treated therapists as fixed effects (thus restricting the conclusions to the particular therapists in the trials), have used a variety of strategies insufficiently described to ascertain how the variability resulting from therapist was calculated, or have failed to reexamine the treatment effect taking therapist variability into consideration. The purpose of the current study was to use analytic strategies that (a) take into account the nesting of patients within therapists, (b) are able to consider

therapists as a random factor in various ways, (c) use estimation procedures that are robust for unequal sample sizes within therapists and between treatments, and (d) model treatment and therapist effects simultaneously. These methods, known as multilevel models or hierarchical linear models, are well suited for the task at hand (see, e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). These methods can also account for various ways in which the therapists can vary, as described in the Method section.

An exemplary study to which to apply multilevel modeling is the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP). In this study, therapists were nested within treatments and were selected in such a way as to minimize allegiance and skill confounds:

> All [therapists] had to meet specific background and experience criteria: at least two years of full-time clinical work following completion of professional training (i.e., following the Ph.D. and clinical internship for clinical psychologists and following the MD and psychiatric residency for psychiatrists); treatment of at least ten depressed patients; and a special interest in and commitment to the therapeutic approach in which they were trained. In addition, IPT therapists had to have previous training in a psychodynamic oriented framework, CB therapists were to have had some cognitive and/or behavioral background ... *Thus, the treatment conditions being compared in this study are, in actuality, "packages" of particular therapeutic approaches and the therapists who both choose to and are chosen to administer them*. (italics added; Elkin, Parloff, Hadley, & Autry, 1985, p. 308)

Briefly, in this study, cognitive–behavioral treatment (CBT) and interpersonal therapy (IPT) were found to produce comparable benefits to the depressed patients treated (Elkin et al., 1989), although there was a tendency for IPT to be superior for the treatment of patients with more severe depression (Elkin et al., 1995). Left unexamined is the issue of how the results reported are affected by therapist variability, as noted by the principle investigator Irene Elkin:

> The central question to be asked in regard to the TDCRP is whether the outcome findings for each of the treatments, and especially for differences between them, might be partly attributable to the particular therapists participating in the study. We have not yet dealt, in a forthright and comprehensive fashion, with this possibility (1999, p. 11).

Those results are reexamined in the current study in the context of determining the variability in outcomes that are due to therapists so that therapist and treatment effects can be compared directly. Three samples were used: (a) completers, (b) intent-to-treat sample, and (c) completers entering the trial with severe depression. In this study, no attempt was made to examine therapist characteristics or actions that might account for the variability among therapists.

There has been a previous attempt to estimate therapist effects in the NIMH data. Blatt et al (1996) analyzed the data from this study by attempting to find characteristics of therapists who were classified into three categories based on their outcomes ("less effective," "moderately effective," and "more effective"). This analysis was limited in terms of estimating therapist effects because therapists were considered as a fixed factor and because the classification strategy ensured outcome differences among the groups. Nevertheless, Blatt et al. found few therapist differences. The current study used recently developed methods that should detect therapist effects should they exist.

Method

Detailed descriptions of the procedures for the NIMH TDCRP are found elsewhere (Elkin et al., 1985, 1989). Patients in the trial were nonbipolar, nonpsychotic outpatients who met various research diagnostic criteria. For the completer sample, the outcomes of 86 patients completing IPT ($n = 46$, with nine therapists) and CBT ($n = 40$, with eight therapists) were analyzed.[1] The intent-to-treat sample consisted of 119 patients assigned to IPT ($n = 60$, with nine therapists) and CBT ($n = 59$, with eight therapists). The severe group, defined similarly to Elkin et al. (1995; i.e., Hamilton Rating Scale for Depression scores [HRSD] ≥ 20 at pretest), consisted of 33 patients (for IPT, $n = 15$, with seven therapists; for CBT $n = 18$, with eight therapists).

Patient status before treatment, during treatment, and at termination, as described by Elkin et al. (1985), was measured by the HRSD, Beck Depression Inventory (BDI), Hopkins Symptom Checklist-90, and Global Assessment Scale (GAS).

Analysis and Results

The design of the NIMH TDCRP is an unbalanced mixed model, in which therapists are considered to be a random factor and nested within treatments, which is a fixed factor (IPT vs. CBT). To accommodate the unbalanced nature of the design,

Table I. Multilevel Analysis of Treatment (Fixed) and Therapists (Random Intercept, Fixed Slopes) Effects for Completers.

	Treatment					Therapists			Error variance σ^2
Variance	Coefficient γ_{01}	*SE*	*t*	*p*	$\hat{\omega}^2$	Variance component τ_o^2	*p*	$\hat{\rho}t$	
HRSD	−0.903	1.432	−0.631	.537	.000	2.44	.211	.069	33.03
GAS	1.50	2.93	0.513	.615	.000	16.67	.062	.097	114.01
BDI	−3.197	1.99	−1.606	.129	.000	4.10	.293	.050	78.56
HSCL	−.048	0.107	−0.449	.659	.000	0.018	.143	.090	0.181

Note. HRSD = Hamilton Rating Scale for Depression; GAS = Global Assessment Scale; BDI = Beck Depression Inventory; HSCL = Hopkins Symptom Checklist-90; ω^2 = estimate of the variability in outcomes resulting from the treatment (a fixed effect, i.e., CBT vs. IPT); ρ_I = estimate of the proportion of variability resulting from therapists (a random effect; see Wampold & Serlin, 2000).

variance components and multilevel analyses were used to estimate treatment and therapist effects.

Treatments Fixed, Therapist Random Intercepts (Fixed Slopes)

The objective of the analyses is to model outcomes on each of the variables by considering the treatment provided (CBT vs. IPT) as a fixed factor and the therapists within treatment as a random factor. Because much of the variability in outcomes is due to the initial severity of the patient, we also want to model the pretest scores for each variable. There are two ways to model variability among therapists. In the first, and simpler, model, examined in this section, the intercepts for the various therapists are allowed to vary randomly, but the relationship between pre-test and post-test for each therapist is stipulated to be constant across therapists, in much the same way that residualized gain scores assume a common regression coefficient (see Figure 1, left panel). This model was set up as a bilevel model: patient level (often referred to as Level 1) and therapist level (often referred to as Level 2; Snijders & Bosker, 1999; Raudenbush & Bryk, 2002). The formulas for the model and model estimation are described in Appendix A.

The proportion of variance resulting from therapist is the intraclass correlation coefficient ρ_I, defined in this context (Wampold & Serlin, 2000) as the ratio of the variance attributable to therapist τ_o^2 to the total variance, which is the sum of the therapist variance and error variance σ^2 (see Appendix A for estimation procedures) Thus,

$$\rho_I = \frac{\tau_o^2}{\tau_0^2 + \sigma^2}$$

The coefficient γ_{01} in this model was used to test the treatment comparison (CBT vs. IPT; see Appendix A). The variation in outcomes resulting from treatments was assessed with ω^2, which is the best estimate of the population value of the proportion of variance attributable to the fixed effect (see Wampold & Serlin, 2000). A comparison of the estimates of ρ_I and ω^2 provides the relative importance of therapists and treatments, respectively, vis-à-vis outcomes.

The estimates for the mixed model (treatment fixed, therapist intercept random) described previously for the completers and the intent-to-treat samples are found in Tables I and II, respectively. Consistent with past analyses of these NIMH data, treatment effects were nil in all cases; that is, the proportion of variability resulting from treatment in all cases was zero. For the completers, the proportion of variance resulting from therapist (viz., $\hat{\rho}_I$) ranged approximately from 5% to 10%, although, because of the statistical problems with estimating

Table II. Multilevel Analysis of Treatment (Fixed) and Therapists (Random Intercept, Fixed Slopes) Effects for Intent-to-Treat Sample.

	Treatment					Therapists			Error variance σ^2
Variable	Coefficient γ_{01}	*SE*	*t*	*p*	$\hat{\omega}^2$	Variance component τ_o^2	*p*	$\hat{\rho}t$	
HRSD	−0.901	1.288	−0.705	.492	.000	0.463	.394	.008	54.17
GAS	2.09	2.765	0.757	.431	.000	14.11	.074	.082	157.61
BDI	−1.94	2.36	−0.820	.425	.000	10.77	.061	.087	113.44
HSCL-90	−0.039	0.110	−0.533	.729	.000	0.0078	.022	.032	0.3535

Note. HRSD = Hamilton Rating Scale for Depression; GAS = Global Assessment Scale; BDI = Beck Depression Inventory; HSCL = Hopkins Symptom Checklist-90; ω^2 = estimate of the variability in outcomes resulting from treatment (a fixed effect, i.e., CBT v. IPT) and ρ_I = estimate of the proportion of variability resulting from therapists (a random effect; see Wampold & Serlin, 2000).

variances and the relatively low power of such tests, the null hypothesis that the true parameter was zero could not be rejected at the .05 level (the results for the GAS approached significance, $p = .062$). For the intent-to-treat sample, the proportion of variance attributable to therapists (viz., $\hat{\rho}_I$) ranged approximately from 1% to 9%; the null hypothesis was rejected for the HSCL ($p = .022$) and approached significance for the GAS ($p = .074$) and the BDI ($p = .061$).

Treatments Fixed, Therapist Random Intercepts and Random Slopes

In the second model examined, the slopes of the regression of the posttest on the pretest were allowed to vary among the therapists. In practical terms, allowing random slopes across therapists permits therapists to vary not only in their overall effectiveness with patients (i.e., a therapist main effect) but also in terms of the types of patients (high depression, low depression) with which they are most effective (i.e., a therapist interaction with pretest score; Figure 1, right panel). The formulas for the model with random slopes and the estimation procedures are found in Appendix B.

The results for the completers and intent-to-treat samples are found in Tables III and IV, respectively. Note that three variance components are estimated: therapist intercept, therapist slope, and the covariance of intercept and slope. With some notable exceptions (BDI slope, $p = .02$), the estimates of variability in therapist intercept and slopes were not statistically significant. For the intent-to-treat sample, because of increased power (more patients per therapist), the variance estimates approached significance and in the case of BDI achieved significance for both intercept and slope ($p = .025$ and $p = .000$, respectively). Treatment effect, again tested with the coefficient γ_{01} (see Appendix B), did not approach statistical significance for any variable and

Table III. Multilevel Analysis of Treatment (Fixed) and Therapists (Random Intercept, Random Slopes) Effects for Completers.

	Treatment (Fixed)					Therapists (Random Effects)	
Variable	$\hat{\gamma}_{01}$	SE	t	p	$\hat{\omega}^2$	Variance component	p
HRSD							
Treatment	−.178	1.507	−.119	.908	.000		
Therapist intercept						4.155	.133
Therapist slope						.148	>.500
Intercept–slope Cov						.782 (correlation = .998)	
Residual						29.573	
GAS							
Treatment	.235	2.329	1.011	.328	.000		
Therapist intercept						14.11	.362
Therapist slope						.484	.417
Intercept–slope Cov						−1.810 (correlation = .996)	
Error						102.01	
BDI							
Treatment	−1.894	1.67	−1.131	.276	.000		
Therapist intercept						5.72	.211
Therapist slope						.188	.020
Intercept–slope Cov						1.03 (correlation = .994)	
Error						65.20	
HSCL-90							
Treatment	−.032	.089	−.354	.728	.000		
Therapist intercept						.019	>.500
Therapist slope						.025	>.500
Intercept–slope Cov						.021 (correlation = .998)	
Error						.172	

Note. HRSD = Hamilton Rating Scale for Depression; GAS = Global Assessment Scale; BDI = Beck Depression Inventory; HSCL = Hopkins Symptom Checklist-90; ω^2 = estimate of the variability in outcomes due to the treatment (a fixed effect, i.e., CBT v. IPT). $\hat{\rho}_t$, the estimate of the proportion of variability due to therapists, is dependent on the value of the independent variable (see text).

Table IV. Multilevel Analysis of Treatment (Fixed) and Therapists (Random Intercept, Random Slopes) Effects for Intent-to-Treat Sample.

	Treatment					Therapists	
Variable	γ_{01}	SE	t	p	$\hat{\omega}^2$	Variance component	p
HRSD							
Treatment	−.860	1.299	−.662	.517	.000		
Therapist intercept						.796	.298
Therapist slope						.052	.230
Intercept–slope Cov						.104[a]	
Residual						53.054	
GAS							
Treatment	.241	2.752	.779	.448	.000		
Therapist intercept						14.05	.120
Therapist slope						.037	>.500
Intercept–slope Cov						−.359[b]	
Error						155.80	
BDI							
Treatment	−1.903	1.903	1.00	.334	.000		
Therapist intercept						11.64	.025
Therapist slope						.338	.000
Intercept–slope Cov						1.92[c]	
Error						92.07	
HSCL-900							
Treatment	−.066	.113	−.583	.568	.000		
Therapist intercept						.014	.126
Therapist slope						.048	.128
Intercept–slope Cov						.019[d]	
Error						.338	

Note. HRSD = Hamilton Rating Scale for Depression; GAS = Global Assessment Scale; BDI = Beck Depression Inventory; HSCL-90 = Hopkins Symptom Checklist-90; ω^2 = estimate of the variability in outcomes due to the treatment (a fixed effect, i.e., CBT vs. IPT). [a]r = .513. [b]r = −.500. [c]r = .966. [d]r = .732.

in all cases the estimate of the proportion of variance resulting from treatments (i.e., $\hat{\omega}^2$) was equal to zero.

Because, as shown in Figure 1 (right panel), the variance among therapists depends on the level of the pretest score, estimates of the proportion of variance resulting from therapists are conditional on values of the pretest. Two descriptive statistics assist in interpretation of this model. First, to the extent that the random slopes are important sources of therapist variation, the residual variance should be decreased. Table V shows the relative decrease in the estimate of σ^2 from the first model (slopes fixed) to the second model (slopes allowed to vary; Singer, 1998). For the completers, allowing the slopes to vary reduced significantly the unexplained variance in the model; for the intent-to-treat sample, allowing slopes to vary resulted in a large reduction of unexplained variance for the BDI, suggesting that initial severity on the BDI was related to the variability in outcomes on the BDI among therapists.

A second way to examine therapist effects is to calculate the proportion of variance resulting from therapists at the mean of the pretest scores. Because the pretest scores were centered about the grand mean, the proportion of variance resulting from therapist when the pretest score is at the mean (i.e., the mean score of the centralized variable is zero) is given by $\rho_I = \tau_o^2/(\tau_0^2 + \sigma^2)$ (see Snijders & Bosker's Equations 5.5 and 5.6, where the x_{ij}s are zero; see also Figure 1, right panel). These intraclass correlations are shown in Table V. For the completers, the proportion of variance resulting from therapists ranged from approximately 8% to 12% (i.e., for the patient entering treatment with average severity, therapists accounted for between 8% and 12% of the variance in outcomes).

Initially Severe Sample

Elkin et al. (1995) demonstrated through random regression models that, for patients with initially

Table V. Additional Statistics for Interpreting Random Slope Analyses.

Variable	$\hat{\sigma}^2$ Fixed slope	$\hat{\sigma}^2$ Random slope	Proportion reduction in error	Grand mean $\hat{\rho}_I$
Completers				
HRSD	33.03	29.57	.10	.123
GAS	114.01	102.01	.11	.122
BDI	78.56	65.20	.17	.080
HSCL-90	.181	.172	.05	.099
Intent-to-treat				
HRSD	54.17	53.05	.02	.015
GAS	157.61	155.80	.01	.083
BDI	113.44	92.07	.19	.112
HSCL-90	.3535	.338	.04	.040

Note. HRSD = Hamilton Rating Scale for Depression; GAS = Global Assessment Scale; BDI = Beck Depression Inventory; HSCL = Hopkins Symptom Checklist-90; ρ_I = estimate of the proportion of variability due to therapists (a random effect; see Wampold & Serlin, 2000).

severe symptoms of depression (i.e., HRSD > 19), IPT tended to produce superior benefits to CBT, although the results for the HRSD failed to meet conventional alpha levels to reject the null hypothesis (viz., $p = .08$). This result is examined here in light of therapist effects. First, the cases meeting the severity criteria were subjected to a simple ANOVA, using residualized gain scores formed in the traditional manner. The results of this ANOVA, which are presented in Table VI, show that indeed the difference between the treatments approached statistical significance ($p = .09$), a result comparable to Elkin's. Table VI also contains the results of the ANOVAs in which the therapists are considered a nested random factor. Although the estimate of the proportion of variability resulting from therapist is relatively large (viz., $\rho_I = .12$), because of the low power it is not close to being statistically significant ($p = .37$). Nevertheless, the estimated 12% of the variance resulting from therapists accounts for the differences observed between IPT and CPT when therapists are ignored; treatment effects now are nonexistent (i.e., estimated to be zero). As explained by Wampold and Serlin (2000), treatment effects are overestimated when therapist variability is ignored because the expected variance among treatments also contains the expected variance among therapists (see Kenny & Judd, 1986; Kirk, 1995; Serlin et al., 2003). In this particular example, it appears that the trend for the superiority of IPT to CPT for patients with initial severity is explained entirely by the variability among therapists.

Discussion

Several multilevel analyses of the NIMH TDCRP data revealed sizable therapist effects, ranging from 1% to 12% depending on the outcome variable and the model adopted. Overall, a simple mean of all the estimates was about 8%. Given that the estimate of the population treatment effect (i.e., the difference between CBT and IPT) in this analysis was zero, it seems clear that therapists were an important source of variability in these data. Although many caveats are discussed next, the results suggest that, with regard to outcomes, therapists are more important than treatments.

Several considerations with regard to the size of therapist effects found in this analysis need to be noted. First, therapist effects found in the NIMH sample were similar to estimates of 6% to 10% found in previous analyses of clinical trials (cf. Crits-Christoph et al., 1991; Crits-Christoph & Mintz, 1991, Wampold, 2001b). Crits-Christoph et al.

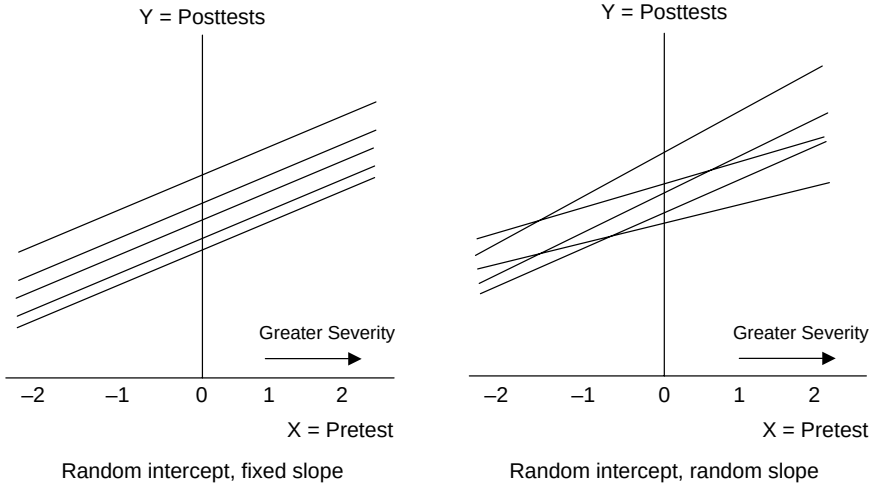

Figure 1. Hypothetical regression lines for five therapists.

Table VI. Treatment and Therapist Effects for Severe Sample Using Residualized Scores.

	Treatment (fixed)			Therapist (random)	
Model	*F*	*p*	ω^2	ρ_I	*p*
Treatment, no therapists	2.94	.09	.06	—	—
Treatment and therapists	1.48	.22	.00	.12	.35

Note. ω^2 = estimate of the variability in outcomes resulting from treatment (a fixed effect, i.e., CBT vs. IPT); $\rho_{I=}$ estimate of the proportion of variability resulting from therapists (a random effect; see Wampold & Serlin, 2000).

(1991) found, in a reanalysis of 16 studies, that the mean estimate of the proportion of variability resulting from therapists was in the neighborhood of 9%, a result that is consistent with the mean therapist effect found in this study. Second, the therapist effects generally were larger for the completer sample than the intent-to-treat sample, which makes sense because the therapists purportedly had less influence on those who dropped out of treatment than those who continued, a result that is tentative, however, given the small differences between the two samples. Third, the therapist effects were greater when the slopes as well as the intercepts (i.e., mean level) were allowed to vary, implying that therapist effects were dependent on level of severity. Fourth, although the therapist effects were fairly large, because of low power to detect such effects in this study, with some exceptions, the estimates of variability resulting from therapists in the various instances were not statistically significant. However, in all cases, the estimates of the proportion of variability in outcomes resulting from therapists was larger that the estimates of the proportion of variance resulting from treatment, which were nil. Fifth, the variability in outcomes produced by therapists in the NIMH TDCRP occurred in the context in which the therapists were selected, trained, and supervised with the goal of minimizing therapist variability (Elkin, 1999).

It appears from the various analyses that variability among therapists is greater as the initial severity of the patients increases. When slopes are allowed to vary, the error variance is reduced (see Table V), supporting the improved fit of a random slope and intercept model that introduces an interaction between therapists and pretest score. This interaction requires that therapist variance be interpreted conditional on pretest score and makes it a function of various characteristics of the model, including the correlation between therapist intercepts and slopes, and therapist slope variance. In the current analysis, therapist was studied at the mean pretest score but can be expected to vary across different levels of the pretest (see Snijders & Bosker, 1999, pp. 68–72, for a cogent discussion of the residual intraclass correlation and its relationship to the covariance of slopes and intercepts and slope variance).[2] Finally, the analysis of the patients with higher initial severity produced larger therapist variability on the HRSD than did the same analysis with all completers (12% vis-à-vis 7%; see Tables I and VI). The importance of severity is not surprising because it could be argued that patients with more severe depression present greater challenges to therapists and that better therapists are better able to meet that challenge.

Several methodological considerations that have important implications for interpreting the results and for future research should be noted. Because therapists were assumed to constitute a random effect in this study, the estimates apply to therapists similar to the those conducting the NIMH TDCRP psychotherapies (Wampold & Serlin, 2000; Serlin et al., 2003; i.e., to therapists with experience in CBT and IPT, respectively, who receive extensive training and supervision and who provided manualized version of the two therapies). Moreover, the estimates of the proportion of variability resulting from therapists and from treatments in this analysis have been corrected for sample bias and thus are not optimized on this particular sample.[3]

Another methodological issue that has importance for interpretation is that treatment effects are confounded with therapist effects in randomized designs. That is, observed differences among treatments are due, in part, to variability among therapists, as noted earlier (Barcikowsky, 1981; Elkin, 1999; Kenny & Judd, 1986; Kirk, 1995; Walsh, 1947; Wampold & Serlin, 2000). The consequence is that if therapist effects are ignored, treatment effects will be overestimated. This is demonstrated for the high initial severity sample, in which the proportion of variance resulting from treatment was estimated to be 6% when therapists were ignored; however, this variability was entirely accounted for by therapist variance, even though the therapist variance did not approach conventional levels of significance. As noted by Crits-Christoph and Mintz (1991), "setting the *p* value for ruling out therapist effects at .05 is inappropriate, because the test for treatment effects will still be significantly affected even if the test for therapist differences does not reach the .05 level" (p. 21).

The results of the current study are consistent with previous research that has produced estimates of moderately large therapist effects. Ironically, as the evidence mounts that therapist variability is important, the focus on therapist characteristics has

declined, a phenomenon noted by Beutler et al. (2004) in their review of therapist variables:

> The strongest impression with which we are left at the conclusion of this review is that over the last two decades, there has been a precipitous decline of interest in researching areas that *are not associated with specific effects of treatment and its implementation* [italics added]. Observable and inferred traits of the therapist have seen the greatest decline in research interest, even though several factors within these clusters of variables have, over the years, been viewed as being very promising predictors of treatment outcome. (pp. 289–290)

The focus on manualized therapies has increased attention to therapist actions (e.g., adherence and competence) and their relationship to differences among therapists (see, e.g., Huppert et al., 2001). Multilevel models have the potential to disentangle therapy and therapist effects and to identify the characteristics and actions of therapists that account for therapist differences.

Notes

[1] A therapist assigned only one client was eliminated because the methods used estimate covariation of scores within therapists.

[2] The correlations are consistently in the same direction, after the scaling of the instruments is taken into account.

[3] Care must be taken when comparing findings across studies to ensure that population proportions rather than sample proportions are used. Some studies (e.g., Huppert et al., 2001), which consider therapists as a fixed factor in an ANOVA or regression, tend to report sample values. For example, if therapist was considered an independent variable in an ordinary least squares analysis, then the R^2 attained would tend to overestimate the true proportion of variance resulting from therapists. Moreover, because therapists would be a fixed factor, the results would be conditioned on the therapists in the particular study (Serlin, Wampold, & Levin, 2003).

References

Barcikowski, R. S. (1981). Statistical power with group means as the unit of analysis. *Journal of Educational Statistics*, *6*, 267–285.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 227–306). New York: Wiley.

Blatt, S. J., Sanislow, C. A., Zuroff, D. C., & Pilkonis, P. A. (1996). Characteristics of effective therapists: Further analyses of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, *64*, 1276–1284.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Carroll, K., Luborsky, L., McLellan, T., et al. (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 comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, *59*, 20–26.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge, UK: Cambridge University Press.

Elkin, I. (1999). A major dilemma in psychotherapy outcome research: Disentangling therapists from therapies. *Clinical Psychology: Science and Practice*, *6*, 10–32.

Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, *63*, 841–847.

Elkin, I., Parloff, M. B., Hadley, S. W., & Autry, J. H. (1985). NIMH Treatment of Depression Collaborative Research Program: Background and research plan. *Archives of General Psychiatry*, *42*, 305–316.

Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*, *46*, 971–982.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Gehan, E., & Lemak, N. A. (1994). *Statistics in medical research: Developments in clinical trials*. New York: Plenum Press.

Hacking, I. (1975). *The emergence of probability*. Cambridge, UK: Cambridge University Press.

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcomes in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology*, *69*, 747–755.

Kenny, D. A., & Judd, C. M. (1986). Consequence of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*, 422–431.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed). Pacific Grove, CA: Brooks/Cole.

Luborsky, L., Crits-Christoph, P., McLellan, A. T., Woody, G., Piper, W., Liberman, B., Imber, S., & Pilkorus, P. (1997). Do therapists vary much in their success? Findings from four outcome studies. *American Journal of Orthopsychiatry*, *4*, 501–512.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S., & Schweizer, E. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, *6*, 95–106.

Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice*, *4*, 53–65.

McCall, W. A. (1923). *How to experiment in education*. New York: Macmillan.

Project MATCH Research Group. (1998). Therapist effects in three treatments for alcohol problems. *Psychotherapy Research*, *8*, 455–474.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., & Congdon, R. (2001). *HLM 5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Serlin, R. C., Wampold, B. E., & Levin, J. R. (2003). Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: A comment on Siemer and Joorman (2003). *Psychological Methods*, *8*, 524–534.

Shapiro, A. K., & Shapiro, E. S. (1997). *The powerful placebo: From ancient priest to modern medicine*. Baltimore, MD: Johns Hopkins University Press.

Siemer, M., & Joormann, J. (2003). Power and measures of effect size in analysis variance with fixed versus random nested factors. *Psychological Methods*, *8*, 497–517.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational & Behavioral Statistics*, *23*, 323–355.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Walsh, J. E. (1947). Concerning the effect of intraclass correlation on certain significance tests. *Annals of Mathematical Statistics*, *18*, 88–96.

Wampold, B. E. (1997). Methodological problems with identifying efficacious psychotherapies. *Psychotherapy Research*, *7*, 21–43.

Wampold, B. E. (2001a). Contextualizing psychotherapy as a healing practice: Culture, history, and methods. *Applied and Preventive Psychology*, *10*, 69–86.

Wampold, B. E. (2001b). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.

Wampold, B. E., & Bhati, K. S. (2004). Attending to the omissions: A historical examination of the evidenced-based practice movement. *Professional Psychology: Research and Practice, 36,* 563–570.

Wampold, B. E., & Serlin, R. C. (2000). The consequences of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, *5*, 425–433.

Appendix A: Multilevel Model: Treatments Fixed, Therapist Random Intercepts (Fixed Slopes)

The multilevel mixed model, when treatments are considered a fixed factor and therapists are considered as a random factor (with fixed slopes), is given by

$$\mathrm{Y_{ij}} = \beta_{0\mathrm{j}} + \beta_{\mathrm{ij}} + \mathrm{r_{i,}}, \tag{1}$$

where Y_{ij} are the last scores for the variable examined for participant i and therapist j, β_{0j} is the intercept for therapist j, β_{1j} is the slope for regression for group j predicted from x_{ij}; some patient variable (here the pretest score) and r_{ij} are the residuals. The variance of the residuals, here called patient variance or error variance, is given by σ^2. The intercepts can be decomposed as follows:

$$\beta_{0\mathrm{j}} = \gamma_{00} + \gamma_{01}\mathrm{z}_1 + \mathrm{u_{oj}}, \tag{2}$$

where γ_{00} is the grand mean (i.e., mean of the intercepts for the therapists), γ_{01} is the regression coefficient related to the therapist variable z_1, which in our case is the treatment administered (coded CBT = 1 and IPT = 2), and u_{oj} are the random deviations about the mean (i.e., therapist deviations from the average intercept). For notational consistency, needed in the next model, the regression coefficient β_1 is written as γ_{10}.

Substituting Equation 2 into Equation 1, we have

$$\mathrm{Y_{ij}} = \gamma_{00} + \gamma_{01}\mathrm{z}_1 + \mathrm{u_{oj}} + \gamma_{10}\mathrm{x_{ij}} + \mathrm{r_{i,}}, \tag{3}$$

Rearranging,

$$\mathrm{Y_{ij}} = [\gamma_{00} + \gamma_{01}\mathrm{z}_1 + \gamma_{10}\mathrm{x_{ij}}] + [\mathrm{u_{oj}} + \mathrm{r_{i,}}]. \tag{4}$$

The terms in the first bracket are the fixed components and the terms in the second bracket are the random components. With regard to fixed effects, γ_{01} is the important parameter because it reflects treatment effects. The variance of the u_{oj}s, which are the therapist deviations from the mean intercept, is the therapist variance in the model and is denoted by τ_0^2. Because the r_{ij}s are the residuals, their variance, denoted by σ^2, is the error variance.

Equation 4 was used to model treatment and therapist effects in the completer and intent-to-treat sample using HLM5 (Raudenbush, Bryk, Cheong, & Congdon, 2001), with pretests centered at the grand mean, by using the method of restricted maximum likelihood. All models converged without problems. SAS Proc MIXED, GLM, and VARCOMP were used to determine estimates needed to calculate ω^2 and to verify results obtained with HLM5.

Appendix B: Multilevel Model: Treatments Fixed, Therapist Random Intercepts, and Random Slopes

The model for case with random slopes begins similarly to that of the previous case:

$$\mathrm{Y_{ij}} = \beta_{0\mathrm{j}} + \beta_1\mathrm{x_{ij}} = \mathrm{r_{ij}}, \tag{5}$$

However, because the slopes are being allowed to vary,

$$\beta_{1\mathrm{j}} = \gamma_{10} + \mathrm{u_{1j}}, \tag{6}$$

where γ_{10} is the average slope for the j therapists and u_{1j} are the random deviations in slopes (i.e., they represent the degree to which the regressions of the posttest onto the pretest differ). The variance of the deviations in slope is denoted by τ_1^2.

Substituting Equation 6 and Equation 2 from the previous model into Equation 5, we now have

$$\mathrm{Y_{ij}} = \gamma_{00} + \gamma_{01}\mathrm{z}_1 + \mathrm{u_{oj}} + (\gamma_{10} + \mathrm{u_{1j}})\mathrm{x_{ij}} + \mathrm{r_{ij}}. \tag{7}$$

Rearranging,

$$\mathrm{Y_{ij}} = [\gamma_{00} + \gamma_{01}\mathrm{z}_1 + \gamma_{10}\mathrm{x_{ij}}] + [\mathrm{u_{oj}} + \mathrm{u_{1j}}\mathrm{x_{ij}} + \mathrm{r_{ij}}]. \tag{8}$$

Within the first brackets are the fixed part of the model; γ_{01} is the critical parameter to estimate because it reflects treatment effects. Within the second bracket are the random parts of the model, and the variances of each of these deviations will be estimated (i.e., intercept τ_0^2, slope τ_1^2, and error σ^2).

Because it could be that slope and intercept are correlated, the covariance of slope and intercept, denoted by τ_{01}, is also estimated (this parameter is interpreted in the Discussion section).

As was the case previously, HLM5 and SAS were used to generate the statistics needed for model identification. The restricted maximum likelihood method produced results that converged without any problems.

Zusammenfassung

Therapeuteneffekte in der Psychotherapie: Ein Modellansatz mit Zufallseffekten mit den Daten des gemeinsamen Forschungsprogramms des Nationalen Instituts für psychische Gesundheit zur Behandlung von Depression

Daten von vollständigen Stichproben mit einer Behandlungsabsicht, von den beiden Psychotherapiebedingungen des gemeinsamen Forschungsprogramms des Nationalen Instituts für psychische Gesundheit zur Behandlung von Depression, wurden analysiert, um den Anteil der Variabilität im Ergebnis zu schätzen, der auf Therapeuteneinfluss beruht. Die Therapeuten, innerhalb ihrer Behandlungen vernetzt, wurden als Zufallsfaktor in Multiebenen-Analysen betrachtet. Diese Analysen, die Therapeutenvariabilität, auf verschiedene Art modelliert, gab zu erkennen, dass ca. 8 Prozent der Varianz im Therapieergebnis den Therapeuten zuzuschreiben war, während 0% auf die spezifische Behandlung zurückzuführen war. Wenn die Therapeuteneffekte in geeigneter Weise modelliert wurden, verschwanden, wie durch methodische Überlegungen vorhergesagt, die früher gefundenen Effektivitätsunterschiede zwischen den beiden Psychotherapiebedingungen für schwerer depressive Patienten.

Résumé

L'effet du thérapeute en psychothérapie : un modele a effets aléatoires des données du programme collaboratif sur la dépression de l'institut national de la santé mentale

Les données pour les échantillons de patients avec intention de se soigner et qui ont terminé le traitement et pour les deux conditions du Programme Collaboratif sur la Dépression de l'Institut National de Santé Mentale (NIMH) ont été analysées pour estimer la proportion de variabilité des résultats provenant des thérapeutes. Les thérapeutes, qui étaient emboîtés avec les traitements, ont été considérés comme un facteur aléatoire dans des analyses multi-niveaux. Ces analyses, qui modélisent la variabilité du thérapeute de plusieurs manière différentes, indiquent que 8% environ de la variance des résultats est attribuable aux thérapeutes, alors que 0% est attribuable au traitement particulier délivré. Lorsque les thérapeutes sont modélisés de manière appropriée, les différences détectées précédemment entre les deux conditions thérapeutiques pour les patients les plus sévèrement déprimés disparaissent, comme prédit par des considérations méthodologiques.

Resumen

Efectos del terapeuta en la psicoterapia: una modelización randomizada de efectos en el Programa de investigación colaborativa para el tratamiento de la depression del Instituto Nacional de Salud Mental

Se analizó la variabilidad de los resultados debidos al terapeuta en muestras de tratamientos completados y prospectivos (*completer and intent-to-treat samples*) de las dos psicoterapias para el tratamiento de la depresión del Programa de investigación colaborativa del Instituto Nacional de Salud Mental. Se consideró que los terapistas, ubicados en los tratamientos (*nested within treatments*), eran un factor para randomizar en análisis multinivel (*multilevel analyses*). Estos análisis, que configuran la variabilidad del terapeuta en diferentes formas, indicaron que aproximadamente el 8% de la varianza en los resultados es atribuible a los terapeutas, mientras que no hay varianza debida al tratamiento particular administrado. Como se predijo por consideraciones metodológicas, cuando los efectos del terapeuta fueron modelizados apropiadamente, desaparecieron las diferencias en eficacia previamente detectadas entre las dos psicoterapias para pacientes más severamente deprimidos.

Resumo

OS EFEITOS DO TERAPEUTA EM PSICOTERAPIA: uma aná lise de modelagem dos efeitos aleatórios dos dados do Programa Colaborativo de Investigação do Tratamento da Depressão do Instituto Nacional de Saúde Mental dos EUA

Foram analisados dados de amostras que completaram o tratamento e que iniciaram tratamento das duas condições psicoterapêuticas do Programa Colaborativo de Investigação do Tratamento da Depressão do Instituto Nacional de Saúde Mental dos EUA para determinar a proporção da variabilidade de resultados terapêuticos resultante dos terapeutas. Os terapeutas, que no desenho forneciam apenas um tipo de tratamento, foram considerados como um factor aleatório nas análises de múltiplos níveis. Estas análises, reflectem a variabilidade dos terapeutas de diversas formas, que modelaram cerca de 8% da variância dos resultados terapêuticos era atribuível aos terapeutas, enquanto que 0% se devia ao tratamento específico prestado. Quando os efeitos dos terapeutas estavam adequadamente adaptados, desapareciam as diferenças modeladas anteriormente na eficácia entre as duas condições psicoterapêuticas para os pacientes mais deprimidos, tal como o prediziam os pressupostos metodológicas.

Sommario

Gli effetti dei terapeuti nella psicoterapia: un *modelling* a risultati casuali dell'istituto nazionale di trattamento della salute mentale nei dati del programma di ricerca in *equipe* sulla depressione

I dati per i campioni più trattati e da trattare con le due condizioni psicoterapeutiche dell'Istituto Nazionale di Trattamento della Salute Mentale del programma di

ricerca in *equipe* sulla depressione sono stati analizzati al fine di valutare la proporzione di variabilità negli esiti che risultano dai terapeuti.

I terapeuti che erano inseriti nei trattamenti sono stati considerati un fattore casuale nelle analisi multilivello. Queste analisi, che hanno modellato la variabilità del terapeuta in parecchi modi diversi, hanno indicato che circa l'8% della varianza degli esiti era attribuibile ai terapeuti, mentre lo 0% era dovuto al particolare trattamento somministrato.

Quando gli effetti dei terapeuti erano appropriatamente modellati, le differenze nell'efficacia precedentemente rilevate tra le due condizioni psicoterapeutiche per i pazienti più severamente depressi scomparvero, come predetto dalle considerazioni metodologiche.

摘要

本研究之資料係針對全國憂鬱症心理健康治療的合作研究計畫中的兩組樣本，分別是完成治療與即將意圖接受治療者，藉以瞭解其療效中有多少的變異量是來自於治療者的因素。治療者是以巢狀分派的方式在不同處遇中，也被視為是多重水準的分析中的一個隨機因子。這些分析將治療者的變異性以不同方式加以模式化，結果發現約有 8%的療效變異量是來自於治療者因素，而特殊處遇策略的影響力是 0%。若將治療者的效應適當地模式化，則如同研究方法的考慮，先前所發現的兩種心理治療情境對於重度憂鬱症患者不同療效的看法將無法成立。