Journal of Consulting and Clinical Psychology
1991, Vol. 59, No. 1, 20–26

0022-006X/91/$3.00

Implications of Therapist Effects for the Design and Analysis of Comparative Studies of Psychotherapies

Paul Crits-Christoph
Department of Psychiatry
University of Pennsylvania

Jim Mintz
Department of Psychiatry
University of California, Los Angeles

Technical reasons are presented as to why therapist should be included as a random design factor in the nested analysis of (co)variance (AN[C]OVA) design commonly used in psychotherapy research. Incorrect specification of the ANOVA design can, under some circumstances, result in incorrect estimation of the error term, overly liberal F ratios, and an unacceptably high risk of Type I errors. Review of studies indicates that the great majority of investigators continue to ignore this issue. Computer simulation studies revealed that considerable bias can be introduced by not specifying therapist as a random term. Finally, a reanalysis is presented of data from 10 psychotherapy outcome studies that indicated that therapist effects vary considerably and at times can be large. More recent studies that implement better quality controls appear to demonstrate less variance due to therapist. The implications of these results for the design of future studies are discussed.

In most psychotherapy research studies, each participating psychotherapist sees more than one patient. Over 10 years ago, Martindale (1978) presented evidence that this apparently minor methodological detail was an important statistical consideration. On the basis of reanalyses of several examples drawn from the published literature, he concluded that the failure to include the therapist as a blocking or stratification factor in the statistical design could seriously distort tests of statistical significance.

The implications of Martindale's paper were sobering. His methodological opinion was that, strictly speaking, almost all of the published findings in the psychotherapy research literature applied only to the specific samples of therapists used in each study. If researchers want to generalize their results to a larger population of therapists, he argued, then they must treat therapists as a random factor in the statistical analyses. His own review of the literature, however, indicated that this practice was very much the exception rather than the rule. Unhappily, adoption of his methodological counsel would result in serious reductions in statistical power in most studies. Furthermore, Martindale's statistical reanalyses suggested strongly that the erroneous statistical designs in common use tended to overestimate the statistical significance of differences.

Our article reconsiders the issue of therapist effects and their statistical implications for the analysis of psychotherapy outcome studies. First, the statistical issue is briefly reviewed. The psychotherapy research literature in the 10-year period since Martindale's critique (1978) is then surveyed with respect to the correctness of the statistical designs and attention to therapist effects. A series of computer simulation studies of the relation between the size of the therapist effect, the patient to therapist ratio, and Type I error rates are reported. Finally, 10 psychotherapy outcome studies are reanalyzed to illustrate concretely the size of therapist effects that are actually typical in the field.

Therapist as a Random Factor

A factor in an analysis of (co)variance design is said to be random when it is sampled from a large population to which we wish to generalize our results, even though this sampling is often opportunistic rather than strictly random. Fixed factors, on the other hand, are specifically chosen by the experimenter for study rather than sampled. The most obvious fixed factor involves the choice of specific modalities used in a study. The most obvious random factor in psychotherapy research designs involves the sampling of patients.

Should we consider therapists to be sampled as well? The answer depends on whether or not they differ. It obviously does not matter which therapists we happen to use if they are all equivalent. However, if the patients of some therapists tend to have better (or worse) outcomes than others, the particular sample of therapists obviously does make a difference. When we ignore the therapist in the statistical analysis, we are making the assumption that sampling of therapists has nothing to do with the observed differences between treatments. If that assumption is incorrect, the differences between treatments that we observe in a particular study will be influenced by the particular therapists used, and generalization to other therapists

Preparation of this article was supported in part by National Institute of Mental Health (NIMH) Grant MH-40472 and NIMH Career Development Award MH-00756 to Paul Crits-Christoph.

We wish to acknowledge the generosity of the investigators who lent us their data, including Aaron T. Beck, Tom Borkovec, Robert DeRubeis, Dolores Gallagher, Steve Hollon, Stan Imber, Bernard Liberman, Lester Luborsky, Tom McLellan, Paul Pilkonis, William Piper, Larry Thompson, George Woody, and Charlotte Zitrin. We also thank David Kenny for his helpful suggestions and comments on a draft of this article.

Correspondence concerning this article should be addressed to Paul Crits-Christoph, Hospital of the University of Pennsylvania, 308 Piersol Building, 3400 Spruce Street, Philadelphia, Pennsylvania 19104-4283

might not be warranted. In that case, it is critical to specify therapist as a random factor in the analysis of variance (ANOVA) design.

It should be emphasized that specifying therapists as a random factor certainly does not imply that all therapists are interchangeable and therefore the same (see Paul & Licht, 1978). On the contrary, it is precisely because therapists are different and potentially have a different impact on patients that they must be considered a random factor so that we can take these differences into account in drawing our conclusions.

When different samples of therapists are used within each treatment modality, therapists are said to be "nested within treatments." In this design, when there are differences between therapists, the correct F ratio with which to evaluate treatment differences uses the mean square for therapists in the denominator, not the between-patients (within cell) error, as would be the case if the therapist factor was ignored or specified as fixed. It can easily be shown that ignoring or specifying the therapist factor incorrectly in the analysis of nested designs will consistently overestimate the between-treatments effect and lead to increased risk of Type I error. Note also that the effective sample size for significance testing with therapist as a random effect is related to the number of therapists, not to the number of patients. This then becomes an important consideration for statistical power calculations in designing a study.

In designs where each therapist provides all modalities, therapists are said to be "crossed with treatments." This kind of design is particularly common in studies of pharmacotherapies, in which the same doctors provide both active drug and placebo. In this design, the correct error term for the F ratio for the treatment factor is the Treatment × Therapist interaction. In the crossed design, the effects on Type I error rates of ignoring or specifying the therapist factor incorrectly depend on whether there are therapist main effects (across treatments), Therapist × Treatment interactions, or both. Ignoring a Therapist × Treatment interaction will yield F ratios that are too large. Ignoring a sizable main effect for therapists, on the other hand, will lead to an inflated estimate of error and F ratios for treatment that are too small.

The therapist factor should, then, be treated as a random factor in studies of the comparative efficacy of psychotherapies whenever there is reason to believe that there are systematic outcome differences among therapists. Under those circumstances, psychotherapy outcome studies should be analyzed with mixed model ANOVAs in which treatment modality is specified as a fixed factor and therapist is specified as a random factor. Failure to specify therapist as a random factor is a potential violation of the ANOVA assumption of nonindependence. Kenny and Judd (1986) have described how serious errors can result from violation of this assumption.

How Have Researchers Treated the Therapist Factor?

How does one know whether or not therapist effects can safely be ignored? There is, of course, a straightforward statistical test for this factor in the analyses of (co)variance (AN[C]OVA) design, but failure of the therapist effect to reach statistical significance may not be a sufficient basis for ignoring it. Psychotherapy research is often based on small patient samples, and the statistical power to detect therapist differences may be quite low (Kazdin & Bass, 1989). Methodologists (Winer, 1971) warn against ignoring (i.e., pooling) components of variance simply because a statistical test in a small sample fails to reach conventional levels of statistical significance.

Martindale (1978) reviewed 33 treatment outcome studies and concluded that researchers were rarely implementing the appropriate analysis. We were interested in the extent to which researchers have become aware of this issue since the 1978 Martindale article. All treatment studies published in the *Journal of Consulting and Clinical Psychology* from 1980 through February 1990 were examined. We used 1980 as the starting point in our review to allow for the lag time in publication following the appearance of the Martindale paper.

We examined all studies that involved the comparison of two or more psychosocial treatments. A total of 140 studies were located. In 26 of these 140 studies, there was only one therapist (or pair of therapists working together) in the study, or one therapist per treatment, making it impossible to analyze therapist as a separate factor and thereby leaving treatment effects and therapist effects inherently confounded.

Most often, the therapist factor was completely ignored in the remaining studies (77 of the 114 studies with more than one therapist). In 32 studies, a preliminary one-way ANOVA was performed to rule out therapist effects. In these studies, however, the p value for deciding whether or not there was evidence of a therapist effect was set at .05 (or in one case, .10) or was not specified. As Martindale (1978) points out, 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. Kirk (1968) has recommended that p level be set at .25, and Winer (1971) suggests .20 or .30 for ruling out a factor in a preliminary analysis. In one study, a significant therapist effect was obtained in a separate analysis, but the investigators continued to analyze treatment effects without incorporating therapist as a random factor. The therapist was treated as a random factor or therapist effects were ruled out in a preliminary analysis with a p level specified as $>.20$ in only four studies.

Bias in Significance Testing

The extent to which significance testing in the ANOVA is distorted by not including a factor as random when it should be, has not, to our knowledge, been evaluated in the literature on ANOVA. Kenny and Judd (1986) give formulas for calculating the effects of nonindependence on expected mean squares. Nonindependence, however, also affects the F ratio by producing more variable estimates of the mean squares and by producing a correlation between mean square treatment and mean square error (Kenny & Judd, 1986). Our task was to investigate the effects of all of these sources of distortion.

Computer simulation studies were designed to evaluate the relationships among size of therapist effects, patient to therapist ratio, and rate of Type I error. Artificial data were created by computer random number procedures to analyze both nested and crossed ANOVA designs. The basic design for these simulations included three treatment groups with therapists either nested within or crossed with treatments. We then varied the number of therapists, the number of patients per therapist, and the percentage of variance in the "outcome scores" resulting from the therapist main effect (nested design) or the Treatment × Therapist interaction (crossed design).

For the nested design, we used 2, 5, 10, and 15 therapists per treatment, with values of 2, 4, 8, and 15 patients per therapist. Therapist effects accounting for 5%, 15%, and 25% of the outcome variance were examined. The values used to evaluate the crossed design were identical, except that we built in Treatment × Therapist interactions of 5%, 15%, and 25% rather than therapist main effects. In addition, in the crossed design the number of therapists (2, 5, 10, and 15) indicates the total number of therapists in the study, rather than the number of therapists per treatment group.

Although we built in significant differences between therapists, we did not introduce any treatment effects. Once a particular design was specified (e.g., two therapists per treatment modality, four patients per therapist, 15% of the variance due to therapist), we performed an ANOVA, ignoring the therapist factor, and tabulated whether a significant (at $p < .05$) treatment effect was found or not. This procedure was repeated 2,000 times for each design using new random numbers with each trial. Because no treatment effect existed in our simulated data, we would of course expect only 5% of the trials to yield a p value of .05 or less. The actual percentage of "significant" findings over the 2,000 trials represents the probability of Type I errors or false positive treatment effects.

In the nested design, the number of patients per therapist and the percentage of variance due to therapist showed systematic relationships to the probability of false positive treatment effects. Those who are familiar with the expected values of the ANOVA mean squares will not find this result surprising. The number of therapists per treatment, however, did not appreciably affect Type I error rates.

Figure 1 presents data for 5%, 15%, and 25% of within-treatment variance due to therapist and 2, 4, 8, and 15 patients per therapist (results are pooled across findings for varying numbers of therapists per treatment). The probability of Type I error (spurious differences between experimental treatments) can be read from Figure 1 (or interpolated) on the basis of the number of patients per therapist and percentage of variance due to therapist. The analysis of the simulated crossed design yielded essentially (within error) the identical curve seen in Figure 1. Thus, this same curve also illustrates the effect of Treatment × Therapist interactions of various magnitudes.

As can be seen in Figure 1, the probability of false positive treatment effects rises linearly with the number of patients per therapist, and this increase becomes more pronounced as the size of therapist effects increases. At the extreme, a design with 15 patients per therapist and 25% of the variance due to therapist yields a probability of Type I error of .52. A design using 8 patients per therapist, with 15% of the outcome variance attributable to therapist, yields a Type I error rate of .23 at the nominal .05 level. These false positive rates are clearly unacceptable.

Do the types of designs used in our simulations actually occur in practice? In other words, are we kicking a straw man? In order to determine how often investigators actually conduct studies using eight or more patients per therapist, we reexamined our sample of *Journal of Consulting and Clinical Psychology* studies. Not counting those studies with only one therapist and a number of studies where we could not determine or estimate the number of patients per therapist, 30% of the sample of studies used at least eight patients per therapist within a treatment condition. Thus, a substantial number of studies use designs that could lead to highly inflated p values if therapist effects are present yet not analyzed correctly.

Therapist Effects in 10 Treatment Outcome Studies

The computer simulations demonstrated that ignoring large therapist effects can lead to seriously inflated Type I error rates, particularly if each therapist sees a relatively large number of patients. Do such large effects actually occur in practice? To address this question, we obtained the raw data from several psychotherapy outcome studies for reanalysis. These studies included a total of 55 therapists treating a total of 652 patients. The studies obtained are summarized in Table 1.

For each study, ANOVAs were performed on each outcome measure using treatment group (where appropriate) and therapist as factors. The therapist factor was specified as a random term, in some studies crossed with treatment and in others nested. Three studies consisted of only a single treatment modality. In these, only therapist was specified as a design factor. Percentage of variance due to therapist and Treatment × Therapist interaction were calculated from the equations for expected mean squares using the approach described by Dwyer (1974) for fixed and random factors, using the complete least squares approach (Scheffé, 1959) to unbalanced ANOVA.

In order to obtain an overall sense of the size of therapist effects within each study, the percentages of outcome variance due to therapist were averaged across all of the measures used within each study. The number of outcome measures used within each study varied from 1 to 18 (*Mdn* = 7.5). The results for these average effect sizes for each study are presented in Table 2. The percentage of outcome variance due to therapist is presented for the nested and single treatment group studies. For the crossed designs, the therapist main effect and Treatment × Therapist interaction are both shown.

Note the large variability in therapist effects across studies. The Beck, Hollon, Young, Bedrosian, and Budenz (1985) and Rush, Beck, Kovacs, and Hollon (1977) studies (data from these two studies were combined), the Luborsky and Crits-Christoph (1988) and the Thompson, Gallagher, and Breckenridge (1987) studies all yielded estimates of therapist effects equal to zero. Most of the other studies had modest therapist main effects (in the 5–10% of variance range). One study (Nash et al., 1965) had a therapist main effect of 13.5%, averaged across six outcome measures. Therapist × Treatment interactions were small with the exception of one study (Piper, Debbane, Bienvenu, & Garant, 1984) in which the average Therapist × Treatment interaction variance across outcome measures was 10.5%.

The average therapist effects in these studies tended to be relatively small, but this averaging across measures may obscure large effects on certain measures within a study. Table 2 therefore also presents results for the largest therapist main effect and Treatment × Therapist interaction found within each study. As can be seen, many of the studies evidenced quite large effects on at least one of the outcome measures used. For exam-

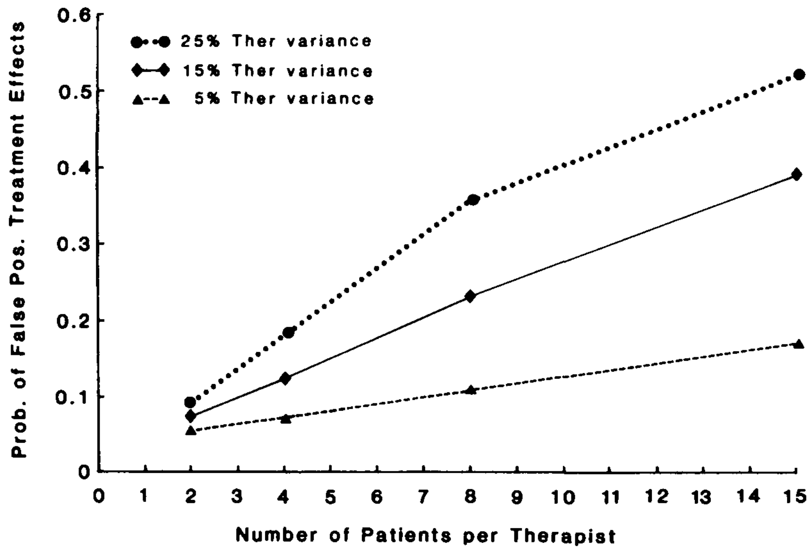

Figure 1. Probability of false treatment effects as a function of number of patients per therapist and percentage of variance due to therapist.

ple, in the Pilkonis, Imber, Lewis, and Rubinsky (1984) study, 23.4% of the outcome variance was due to therapist on one of the measures. We found a 39.0% effect on one of the measures in the Piper et al. (1984) study. For Nash et al. (1965) a 36.4% effect was found, and for Woody, McLellan, O'Brien, and Luborsky (1989) a 20.1% effect was found. The Nash et al. (1965) study also showed a large (37.7%) Treatment × Therapist interaction on one measure. These kinds of effects would generally be viewed as large effects for the behavioral sciences (Cohen, 1969).

The reasons for the variability in therapist effects within and across these studies are not entirely clear. We can speculate that some modalities of treatment may be more prone to therapist effects (e.g., unstructured treatments like psychodynamic therapy may give more leeway for aspects of the therapist to come into play, in contrast to highly structured treatments such as

Table 1
Descriptive Characteristics of Studies Examined for Therapist Effects

Author	Year	Treatment	Patient population
1a. Beck, Hollon, Young, Bedrosian, & Budenz	1985	Cognitive	Depression
1b. Rush, Beck, Kovacs, & Hollon	1977	Cognitive	Depression
2. Zitrin, Klein, & Woerner	1978	Behavior therapy, plus imipramine Behavior therapy, plus placebo Supportive therapy, plus imipramine	Phobias
3. Thompson, Gallagher, & Breckenridge	1987	Behavior therapy Cognitive therapy Dynamic therapy	Elderly depressed
4. Pilkonis, Imber, Lewis, & Rubinsky	1984	Individual Conjoint Group	Mixed diagnoses
5. Piper, Debbane, Bienvenu, & Garant	1984	Short and long term individual dynamic Short and long term group dynamic	Mixed diagnoses
6. Nash et al.	1965	Dynamic with and without role induction interview	Mixed neurotic
7. Hollon et al.	1983	Cognitive Cognitive plus imipramine	Depressed
8. Woody, McLellan, O'Brien, & Luborsky	1989	Dynamic	Opiate addicts
9. Luborsky & Crits-Chrostoph	1988	Dynamic	Depressed
10. Borkovec & Mathews	1988	Nondirective Cognitive Coping desensitization	Generalized anxiety disorder Panic disorder

Note. Data from Studies 1a and 1b were combined for analyses.

Table 2
Therapist Effects in 10 Studies

		% of variance			
		Therapist		Treatment × Therapist	
Study	No. of outcome measures	Average effect	Largest effect	Average effect	Largest effect
Nested design (or single treatment study)					
Woody, McLellan, O'Brien, & Luborsky (1989)	9	4.3	20.1		
Luborsky & Crits-Christoph (1988)	2	0	0		
Pilkonis, Imber, Lewis, & Rubinsky (1984)	10	7.6	23.4		
Thompson, Gallagher, & Breckenridge (1987)	3	0	0		
Beck, Hollon, Young, Bedrosian, & Budenz (1985)/Rush, Beck, Kovacs, & Hollon (1977)	1	0	0		
Crossed design					
Zitrin, Klein, & Woerner (1978)	18	3.8	15.1	1.6	6.4
Hollon et al. (1983)	1	5.2	5.2	0	0
Nash et al. (1965)	6	13.5	36.4	10.5	37.7
Piper, Debbane, Bienvenu, & Garant (1984)	10	5.0	39.0	1.6	10.7
Borkovec & Mathews (1988)	11	2.5	11.1	0.6	4.0

Note. Average effect is the mean effect over outcome measures within each study. For the largest effect, the outcome measure with the greatest percentage of variance due to therapist or Treatment × Therapist interaction was selected for each study.

systematic desensitization or cognitive therapy). A recent report (Perry & Howard, 1989) suggests that the level of therapist experience may relate to the size of therapist effects, with less experienced therapists showing high therapist effects (near 50% of outcome variance) and highly experienced therapists showing no significant variability.

Therapist variables such as competence or skill (cf. Crits-Christoph, Cooper, & Luborsky, 1988; O'Malley et al., 1988) have been shown to relate to outcome, and variability on these dimensions may explain the presence of therapist effects to some extent. Additionally, the processes used to select, train, and supervise therapists in a given study may be responsible for differences in the quality of therapists within a study. Treatment manuals have been implemented in psychotherapy research precisely to try to control this variability among therapists.

We calculated Spearman rank correlations between the date of publication and the size of the therapist effect for the studies in Table 1. For the crossed designs, the variance due to main effect and interaction were totaled to obtain an overall index of the impact of the therapist. With the measure of average percentage of variance, the Spearman correlation was −.60; and for the largest effects selected from each study, the correlation was −.54. These correlations indicate clearly that more recent studies have been more successful at standardizing the delivery of treatment across therapists. A more extensive analysis of the factors that may be related to the size of therapist effects across studies is reported elsewhere (Crits-Christoph et al., 1990).

Most notable in the data presented in Table 2 is that although therapist main effects and Treatment × Therapist interactions are generally small when all outcome measures are averaged, most studies showed large effects on at least one outcome measure. Certain types of outcome measures may be more influenced or responsive to individual differences among therapists. An obvious consideration is whether the outcome measure is derived from therapist ratings. In this case, the "therapist effect" would actually be a judge effect. This explanation, however, is not likely to account for Treatment × Therapist interactions.

Implications

There are several important ramifications of these findings. Obviously, the results of some published studies are called into question by our results. Studies in the literature with modest to large therapist effects may have drawn misleading conclusions on the basis of overly liberal tests of statistical significance. That is, the presence of therapist effects may have led to conclusions that treatments differ when in fact they do not.

Although an increasing number of psychotherapy researchers appear to be attentive to the possibility of therapist effects since the publication of Martindale's (1978) article, the preliminary tests for the presence of therapist effects used in most studies involve inappropriate statistical criteria ($p = .05$ rather than .20 or more) for judging whether or not those effects can safely be ignored. Clearly, greater attention to this issue is warranted.

We found evidence that the better controlled outcome studies performed in recent years have smaller therapist effects, often to the vanishing point. This finding implies that the quality control procedures commonly implemented in contemporary outcome trials (e.g., careful selection, training, and supervision of therapists and the use of treatment manuals) to control for differences among therapists may have been quite successful. If this interpretation of the data is correct, concern about therapist differences and inflated Type I errors may apply to fewer studies performed today than to those performed 10 years ago.

It should be noted that the nonindependence issue presented here may also have implications for other types of treatment studies besides psychotherapy outcome trials. For example, to the extent that nonspecific factors in the doctor–patient relationship are important, "doctor" effects may be present in trials involving pharmacotherapy. It remains to be seen whether the use of double-blind designs and crossing of therapists with treatments effectively controls for this potential problem.

Although we have focused on psychotherapy outcome, the problem of therapist effects is also germane to studies of the process of psychotherapy, studies of group psychotherapies, and studies of some special psychiatric populations. In fact, the importance of specifying factors as random is not limited to psychotherapy research. For example, Clark (1973) has discussed the designation of stimulus words as random effects in language studies. Myers, Di Cecco, and Larch (1981) describe a similar nonindependence problem in studies that compare group dynamics with individual performances. This latter article is particularly relevant to psychotherapy studies that compare group and individual psychotherapy.

Of course, therapist effects are not just a nuisance to deal with statistically en route to analyzing treatment differences. Differences between therapists may be important in their own right. Studies of effective versus ineffective therapists (cf. Lafferty, Beutler, & Crago, 1989) may be a useful way to pursue an understanding of how psychotherapy works and how to best train therapists to be successful.

Recommendations

We can offer several recommendations to researchers conducting comparative studies of psychotherapy. Perhaps the most important recommendation we can make is in terms of the planning of future treatment outcome studies. If there is reason to believe that therapist differences may be found, a study should be designed using as many therapists as possible. By maximizing the number of therapists, the researcher will have the greatest possible degrees of freedom for testing treatment effects. Smaller patient to therapist ratios will minimize the inflation of Type I error rates, but use of only one patient per therapist produces complete experimental confounding that precludes testing for therapist effects at all.

Once data are collected, investigators should routinely test for therapist differences in preliminary analyses. A lack of information about therapist differences only leaves the study vulnerable to potentially erroneous conclusions, and it prevents researchers from understanding a potentially interesting and meaningful source of outcome variation. It is encouraging that investigators are increasingly performing preliminary analyses to check for the presence of therapist effects. For ruling out therapist effects in a preliminary analysis, however, p values greater than .2 or .3 should be employed, not the standard .05.

If, despite everything, an investigator chooses to simply ignore therapist effects and proceed with conventional least squares methods, awareness that the F ratios are not correct should provide a basis for real caution in interpretation of results. At the least, this means including a sentence in the discussion section of the research report stating that the results may not hold up if another sample of therapists were used in a replication attempt.

Our study demonstrates clearly that large therapist effects can seriously inflate nominal significance levels. Even in the most controlled research setting, then, there seems to be no good reason not to systematically evaluate therapist outcome effects in a preliminary data analytic step. Heterogeneity of therapists with regard to outcome presents the researcher with very difficult data analytic choices. Technically correct analyses may sacrifice power to unacceptable levels, but incorrect analyses can result in an unacceptably high frequency of incorrect inferences.

There are, however, some encouraging notes. The fact that attention to this issue appears to be increasing is salutary. Our reanalysis of a small sample of psychotherapy studies gathered over a 20-year period suggests that research therapists appear to be more homogeneous today than they used to be. Although we can only draw the conclusion indirectly, it is tempting to infer that this increased homogeneity of therapist effectiveness results directly from today's increased emphasis on careful selection of therapists, special training in research modalities, ongoing supervision of therapists, and the use of detailed treatment manuals. This is not to say that these kinds of research controls should be implemented in all forms of psychotherapy research. Studies of individual differences in therapists' skill levels, for example, might best be designed without these controls so that *less* homogeneity among therapists results and, consequently, there is a greater basis for generalizing to the larger group of practicing therapists in the community. For comparative studies of psychotherapies, however, homogeneity among therapists is an asset.

References

Beck, A. T., Hollon, S. D., Young, J. E., Bedrosian, R. C., & Budenz, D. (1985). Treatment of depression with cognitive therapy and amitriptyline. *Archives of General Psychiatry, 42,* 142–148.

Borkovec, T. D., & Mathews, A. M. (1988). Treatment of nonphobic anxiety disorders: A comparison of nondirective, cognitive, and coping desensitization therapy. *Journal of Consulting and Clinical Psychology, 56,* 877–884.

Clark, A. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A. T., Carroll, K., Perry, K., Luborsky, L., McLellan, A. T., Woody, G., Thompson, L., Gallagher, D., & Zitrin, C. (1990, June). *Meta-analysis of therapist effects in psychotherapy outcome studies.* Paper presented at the Society for Psychotherapy Research, Wintergreen, VA.

Crits-Christoph, P., Cooper, A., & Luborsky, L. (1988). The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *Journal of Consulting and Clinical Psychology, 56,* 490–495.

Dwyer, J. H. (1974). Analysis of variance and the magnitude of effects: A general approach. *Psychological Bulletin, 81,* 731–737.

Hollon, S. D., Tuason, V. B., Wiemer, M. J., DeRubeis, R. J., Evans, M. D., & Garvey, M. J. (1983). *Cognitive therapy, pharmacotherapy, and combined cognitive-pharmacotherapy in the treatment of depression. I. Differential outcome in the CPT project.* Unpublished manuscript.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57,* 138–147.

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

Kirk, R. E. (1968). *Experimental design procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole.

Lafferty, P., Beutler, L. E., & Crago, M. (1989). Differences between more and less effective psychotherapists: A study of select therapist variables. *Journal of Consulting and Clinical Psychology, 57,* 76–80.

Luborsky, L., & Crits-Christoph, P. (1988, May). *A pilot study of psychodynamic psychotherapy for major depressive disorder* Paper presented at a National Institute of Mental Health workshop on planning of clinical trials with psychodynamic psychotherapy, Rockville, MD.

Martindale, C. (1978). The therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology, 46,* 1526–1530.

Myers, J. L., Di Cecco, J. V., & Larch, R. F. (1981). Group dynamics and individual performances: Pseudogroup and quasi-F analyses. *Journal of Personality and Social Psychology, 40,* 86–98.

Nash, E., Hoehn-Sacric, R., Battle, C., Stone, A., Imber, S. & Frank, J. (1965). Systematic preparation of patients for short-term psychotherapy: 2. Relation to characteristics of patient, therapist and the psychotherapeutic process. *Journal of Nervous and Mental Disorders, 140,* 374–383.

O'Malley, S. S., Foley, S. H., Rounsaville, B. J., Watkins, J. T., Sotsky, S. M., Imber, S. D., & Elkin, I. (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology, 56,* 496–501.

Paul, G. L., & Licht, M. H. (1978). Resurrection of uniformity assumption myths and the fallacy of statistical absolutes in psychotherapy research. *Journal of Consulting and Clinical Psychology, 46,* 1531–1534.

Perry, K., & Howard, K. I. (1989, June). *Therapist effects. In search of the therapists' contribution to psychotherapy outcome.* Paper presented at the Society for Psychotherapy Research, Toronto, Ontario, Canada.

Pilkonis, P. A., Imber, S. D., Lewis, P., & Rubinsky, P. (1984). A comparative outcome study of individual, group, and conjoint psychotherapy. *Archives of General Psychiatry, 41,* 431–437.

Piper, W. E., Debbane, E. G., Bienvenu, J. P., & Garant, J. (1984). A comparative study of four forms of psychotherapy. *Journal of Consulting and Clinical Psychology, 52,* 268–279.

Rush, A. J., Beck, A. T., Kovacs, M., & Hollon, S. D. (1977). Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research, 1,* 17–37.

Scheffé, H. (1959). *The analysis of variance.* New York: Wiley.

Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology, 55,* 385–390.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Woody, G., McLellan, A. T., O'Brien, C., & Luborsky, L. (1989). *Supportive expressive psychotherapy in the treatment of opiate dependence.* Unpublished manuscript.

Zitrin, C. M., Klein, D. F., & Woerner, M. G. (1978). Behavior therapy, supportive psychotherapy, imipramine, and phobias. *Archives of General Psychiatry, 35,* 307–316.

Received May 17, 1990
Revision received June 28, 1990
Accepted June 29, 1990 ■