Journal of Consulting and Clinical Psychology
1997, Vol. 65, No. 3, 355–365

0022-006X/97/$3.00

Evidence That Therapy Works in Clinically Representative Conditions

William R. Shadish
University of Memphis

Georg E. Matt
San Diego State University

Ana M. Navarro
University of California, San Diego

Greg Siegle
San Diego State University

Paul Crits-Christoph
University of Pennsylvania Medical Center

Mark D. Hazelrigg
Federal Correctional Institute, Butner, North Carolina

Anthony F. Jorm
Australian National University

Larry C. Lyons
Hollins Communication Research Institute

Michael T. Nietzel
University of Kentucky

H. Thompson Prout
Florida State University

Leslie Robinson
University of Memphis

Mary Lee Smith
Arizona State University

Martin Svartberg
University of Trondheim

Bahr Weiss
Vanderbilt University

This article reports a secondary analysis of past therapy outcome meta-analysis. Fifteen meta-analysts provided effect sizes from 56 studies in previous reviews that met 1 of 3 increasingly stringent levels of criteria for clinical representativeness. The effect sizes were synthesized and compared with results from the original meta-analyses. Effect sizes from more clinically representative studies are the same size at all 3 criteria levels as in past meta-analyses. Almost no studies exist that meet the most stringent level of criteria. Results are interpreted cautiously because of controversy about what criteria best capture the notion of clinical representativeness, because so few experiments have tested therapy in clinical conditions, and because other models for exploring the generalizability of therapy outcome research to clinical conditions might yield different results.

William R. Shadish and Leslie Robinson, Department of Psychology, University of Memphis; Georg E. Matt and Greg Siegle, Department of Psychology, San Diego State University; Ana M. Navarro, Department of Family and Preventive Medicine, University of California, San Diego; Paul Crits-Christoph, Center for Psychotherapy Research, Department of Psychiatry, University of Pennsylvania Medical Center; Mark D. Hazelrigg, Federal Correctional Institute, Butner, North Carolina; Anthony F. Jorm, National Health and Medical Research Council Social Psychiatry Research Unit, Australian National University, Canberra, Australia; Larry C. Lyons, Hollins Communication Research Institute, Roanoke, Virginia; Michael T. Nietzel, Department of Psychology, University of Kentucky; H. Thompson Prout, Psychological Services in Education Program, Department of Human Services and Studies, Florida State University; Mary Lee Smith, Department of Psychology, Arizona State University; Martin Svartberg, Department of Psychiatry and Behavioral Medicine, University of Trondheim, Trondheim, Norway; Bahr Weiss, Department of Psychology, Vanderbilt University. Larry C. Lyons is now in independent practice in Manassas, Virginia. H. Thompson Prout is now at the Department of Educational and Counseling Psychology, University of Kentucky.

Correspondence concerning this article should be addressed to William R. Shadish, Department of Psychology, University of Memphis, Memphis, Tennessee 38152.

Many quantitative reviews of therapy outcome studies have been reported (Lipsey & Wilson, 1993). To our knowledge, all have concluded that study participants who receive therapy have better outcomes (on the average) than those of participants who do not receive therapy. Such meta-analyses are now cited in policy debates about whether therapy is sufficiently effective to merit insurance reimbursement for patients (e.g., Depression Guideline Panel, 1993; Office of Technology Assessment, 1994). Implicit in such uses of meta-analytic results is the presumption that we may generalize from the therapy research reviewed in these meta-analyses to the therapy that occurs in actual clinical practice.

In a recent review, however, Weisz, Weiss, and Donenberg (1992; see also Weisz, Donenberg, Han, & Weiss, 1995) gave a very thought-provoking challenge to this assumption. They examined four meta-analyses that integrated more than 200 controlled outcome studies of the effects of child and adolescent therapy. As with other meta-analyses, they found consistent evidence for the beneficial effects of psychotherapeutic interventions with children and adolescents. However, Weisz et al.

pointed out that most studies included in these meta-analyses examined what they called *research therapy* rather than *clinic therapy,* the former involving "experimental procedures, nonreferred subjects, specially trained therapists with small caseloads, and other features that may not represent conventional clinic therapy" (1992, p. 1578). When Weisz et al. examined six clinically representative studies (we describe their criteria shortly), they found that "research focusing on more representative treatment of referred clients in clinics has shown more modest effects; in fact, most clinic studies have not shown significant effects" (1992, p. 1578). Of course, Weisz et al. (1992) intended their conclusion to hold only for the child and adolescent therapies they examined. Nonetheless, if this conclusion is true in the child and adolescent literature, it might be true more generally. The latter finding would challenge the generalizability of the therapy research literature to the clinical conditions under which most therapy occurs.

Debate about the generalizability of therapy research has been going on for decades, of course. For example, therapy researchers have debated the usefulness of therapy analogue research in which conditions approximate the clinical situation but in which, for example, target problems may be less severe, patient populations may be less disturbed, and therapists may be less experienced (Kazdin, 1978, 1986). Such research provides more opportunity for laboratory control of variables, but questions about its generalizability have been raised. Of course, the controlled efficacy studies on clinical samples that are the stuff of most meta-analyses are a far cry from analog research, but we cannot assume that either kind of study will necessarily generalize to the full clinical setting.

Perhaps the most recent contribution to this issue is the *Consumer Reports* (1995) therapy survey (see also Seligman, 1995). This study surveyed 180,000 of its readers in its annual subscriber survey, of which about 7,000 responded to questions about mental health problems they experienced and how they dealt with those problems. With some notable exceptions, the results of this survey converged with results from previous therapy meta-analyses and generally supported the effectiveness of therapy. The study's great strength is its thoroughgoing immersion in the actual field practice of therapy. Its results are widely viewed as supportive of the generalizability of therapy outcome research studies. Its most salient weakness is that the study was a survey rather than an experiment, leading to significant ambiguity in causal inference.

Therapy meta-analyses have not clearly addressed this issue in the past, studying bits and pieces of the distinction between clinic and research, without consistent results. For example, Smith, Glass, and Miller (1980) reported that effect sizes from therapy studies conducted in college facilities ($\bar{d}$ = 1.04) were much larger than those conducted in typical clinical settings (e.g., mental health centers, $\bar{d}$ = .47; other clinic or outpatient facilities, $\bar{d}$ = .79). On the other hand, Andrews and Harvey (1981) found statistically significant differences favoring stuttering treatments delivered in both outpatient ($\bar{d}$ = .76) and inpatient settings ($\bar{d}$ = 1.00), compared with those delivered in a university setting ($\bar{d}$ = .67). In between, Dole, Rockey, and DiTomasso (1983), Jorm (1989), Shapiro and Shapiro (1982), Steinbrueck, Maxwell, and Howard (1983), and Wittmann and Matt (1986) found no statistically significant differences between laboratory and clinical settings. Similarly inconclusive results have held in meta-analysis for potential features of clinic therapy, such as the use of experienced therapists (e.g., Stein & Lambert, 1995).

Weisz et al.'s (1992) contribution is more complex, and it is not about the effects of just one variable such as the use of experienced therapists. Their point concerns the joint effects of a constellation of variables that might pertain to clinical relevance. It is that joint constellation that has not been tested in the general meta-analytic literature. The constellation of variables they proposed for describing differences between clinic and research therapies were as follows: (a) In research therapy, participants are recruited by the researcher, but in clinic therapy they are referred by others or are self-referred; (b) in research therapy, patients are often homogenous in their personal characteristics, compared with the heterogenous kinds of clients treated in clinic therapy; (c) in research therapy, treatment is often provided for one focal problem, compared with the range of problems typically dealt with in clinic therapy; (d) in research therapy, therapists are often trained in the specific procedure under study, but this does not happen in clinic therapy; (e) in research therapy, the therapist is usually instructed to use only the protocol under study, but in clinic therapy the therapist does not usually follow a protocol; and (f) in research therapy, a treatment manual is often used and treatment implementation is monitored, but this does not happen in clinic therapy. Moreover, judging from the clinic therapy studies that Weisz et al. (1992) then examined under the label of clinic-based studies, an implicit criterion was that the study be based in a clinic, with its primary function being clinical. It is under this joint constellation of conditions that their review of relevant studies concluded that clinic therapy yielded null effects when compared with clients not receiving treatment.

In the present study, we use a broader therapy database to conduct a meta-analytic test of this joint constellation hypothesis, using Weisz et al.'s (1992) criteria as an initial guide to conceptualizing the distinction between research and clinic therapy. Any effort to test such a hypothesis, however, must deal with questions about the appropriateness of their proposed criteria for clinical relevance. On the one hand, Weisz et al.'s (1992) distinction between clinic and research therapy is certainly debatable and somewhat arbitrary; indeed, we discuss examples of its limitations at the end of this article. Like any simple distinction, it fails to capture the full richness of either kind of therapy, and one could undoubtedly make the distinction in other ways that might yield somewhat different results. Nonetheless, their distinction between research and clinic therapy is both plausible and a good starting point for research, for three reasons.

First, empirical work about the nature of clinic therapy suggests that their criteria accurately describe much clinic practice. For example, surveys of American Psychological Association Divisions 12, 23, and 29 members (Norcross & Prochaska, 1982; Prochaska & Norcross, 1983) indicate that most therapists (98%) had doctorates with an average of 12 to 16 years postdoctoral experience. More of these therapists described their approach as eclectic (30%) than any other orientation; by comparison, most research therapy examines a particular theoretical

orientation such as behavioral or psychodynamic. Most therapists (68%) worked in settings where the primary function was clinical (see also Perlman, 1985). In brief, then, a large number of practicing psychotherapists probably do work in clinic settings; they see referred rather than solicited clients who are heterogeneous in personal and diagnostic characteristics; and they use interventions in which they have little recent training, with no treatment manual, and with no one looking to see they implement therapy "properly." If one were to propose a description of the modal instance of actual therapy, it would probably be very similar to Weisz et al.'s (1992) criteria.

Second, on a practical level, these criteria have already been used with provocative results (Weisz et al., 1992). Their findings, that clinic therapy (by their criteria) may be no more effective than no treatment, have potentially important implications for clinical practice and public policy. Although one might question the validity of their criteria, the attempt to explore the robustness of Weisz et al.'s (1992) findings should probably start by adopting reasonably similar criteria using their "constellation" hypothesis. So we begin the exploration of clinical relevance at the plausible modal instance point suggested by Weisz et al.'s (1992) criteria.

Third, the approach that Weisz et al. (1992) used is a reasonable methodology to study generalizability. What they have done is conceptually consistent with a principle for investigating external validity that Campbell called proximal similarity—that we generalize an experimental effect "with most confidence where treatment, setting, population, desired outcome, and year are closest in some overall way to the original program treatment" (1986, pp. 75–76). In this conceptualization, generalization is warranted through similarity between the object studied and the target of generalization. It is presumably the case that the modal target of generalization for most therapy research—therapy as it is conducted by practicing psychotherapists—has certain characteristics that are plausibly described by Weisz et al.'s (1992) criteria or something much like those criteria. In the Discussion section of this article, we suggest other models of generalization that might also be investigated, but the model implicitly followed by Weisz et al. (1992) is one credible methodology.

However, the present study also expands in some ways on Weisz et al.'s (1992) approach, so it is not a strict replication. First, Weisz et al. limited their review to child and adolescent therapies, but we included a broader array of clients and psychotherapies. Second, Weisz et al. (1992) mostly examined the statistical significance of results of clinically representative therapy; in their more recent work, however, Weisz et al. (1995) did use effect size estimates, and we use effect size estimates as well because statistical significance level is dependent on sample size but effect size is not and because this places the set of clinic therapy studies on the same outcome metric as the meta-analyses. Third, Weisz et al. examined only follow-up assessments in most of the clinic-based studies they reviewed, but we used mostly posttest assessments because most studies have posttests but relatively few have follow-ups, most meta-analyses coded posttests rather than follow-ups, and examining posttests maximizes the chance of finding an effect for clinically representative studies if therapy effects diminish over time. Fourth, the criteria we used for clinic therapy are not identical to those of Weisz et al. (1992) primarily because the meta-analysts whose work we review often used somewhat different criteria than Weisz et al. (1992). Fifth, Weisz et al. had difficulty locating many clinically representative outcome studies in the child and adolescent therapy literature. To locate more studies, we took advantage of the large databases of studies already available in existing meta-analyses of the therapy research literature. To the extent that the authors of those meta-analyses had coded variables pertaining to clinical relevance, they could quickly identify studies with some or all of the desirable characteristics. They could also report effect sizes for those studies more quickly than if the studies had to be located, coded, and analyzed from scratch. A disadvantage of this approach is that these meta-analysts did not code all of Weisz et al.'s criteria. Therefore, the present methodology almost certainly underestimates the number of clinically representative studies in the literature. Another disadvantage is that the present study is limited to secondary analysis of the data already coded by these existing meta-analysts, which means that we cannot easily introduce new variables to further explore the research versus therapy distinction.

Method

We sent letters to 48 authors of all 59 published meta-analyses of therapy outcome studies that we located in a literature review (some authors published more than one meta-analysis). The letters requested that the authors compare their database against three progressively more stringent sets of clinical relevance criteria and that they report the number and posttest effect sizes of studies that met the criteria. The three sets of criteria were described and applied in three cumulative stages, with Stage 2 criteria including all Stage 1 criteria plus new ones, and Stage 3 criteria including both Stage 1 and Stage 2 criteria plus new ones. We considered applying all criteria simultaneously rather than in stages (i.e., to go straight to Stage 3), but such a choice would have prevented examination of any cumulative effect of adding more and more criteria. The division of criteria into three sets was, of course, partly abitrary but reflected (a) an effort to group them into coherent sets; (b) expectations about what variables meta-analysts were likely to have coded, putting more likely variables in earlier stages so as to keep the sample size as high as possible through the stages; (c) expectations that few studies would meet the full set of criteria and that clinically representative information could still be obtained using a smaller set of criteria that could be met by more studies; and (d) an effort to begin with variables that had the most face validity as measures of clinically representative therapy and that proceeded through additional criteria that might be less obviously related but still potentially important. However, we stress that there is no single, correct grouping or ordering of these criteria.

We first asked meta-analysts to choose studies that compared a treated group with a control condition that did not receive an active treatment because this is both the comparison made by Weisz et al. (1992) and the comparison that is crucial to an examination of therapy effectiveness. In Stage 1, we asked the meta-analysts to select studies that (a) were conducted in nonuniversity settings[1] (e.g., outpatient mental health clin-

[1] This criterion departs from that of Weisz et al. (1992) and resulted from an early misreading of their research–clinic distinction as a university–nonuniversity distinction. Many of the nonuniversity settings included in the present study were clinic-based settings, but some were not, such as studies of child interventions in school classroom settings. These latter settings probably do count as a legitimate target of generalization for psychotherapy research but may well not count as occurring in a clinic in the sense that Weisz et al. intended.

ics, community mental health centers, general hospitals, private practice, prisons, school systems), (b) involved patients that were referred through usual clinical routes rather than solicited by the experimenter, and (c) involved experienced, professional therapists with regular caseloads rather than therapists in training or therapists who received training specifically for that research study. In Stage 2, we asked the meta-analysts to select from among those studies that passed Stage 1 criteria any study that also (d) did not use a treatment manual and (e) did not monitor the implementation of treatment. In Stage 3, we asked the meta-analysts to select from among those studies that passed Stage 2 any study that also (f) used clients who were heterogeneous in personal characteristics (e.g., sex, age, race, socioeconomic status), (g) used clients who were heterogeneous in focal presenting problems, (h) used therapists who were not trained immediately before the study in the specific treatment being studied (of course, they may have received training in that treatment at some point in their career but not for purposes of the study in which they participated), and (i) used therapists who were free to use a wide variety of procedures in treatment and were not limited to just one treatment procedure. If a coding system used in a meta-analysis did not include all of these characteristics or if it included slightly different characteristics, researchers were asked to identify studies that came as close as possible to the criteria and to inspect the original studies to make this determination if needed. The meta-analyst provided us with the average study-level effect size at each stage for each study, along with which stage each study passed.

Note, therefore, that the present study is not a meta-analysis; rather, it is a secondary analysis of past meta-analytic data. In secondary analyses of multiple primary sources, a concern is whether the original meta-analysts applied the stage criteria uniformly. To explore this, William R. Shadish, Georg E. Matt, and Ana M. Navarro independently coded the criteria and effect sizes from one randomly sampled study from each of the 15 meta-analyses with clinically representative studies. Reliability was very good. The correlation between our computation of effect size and the original meta-analysts' was .86. For each of the nine clinical relevance criteria (coded dichotomously), percent agreement between our ratings and the original meta-analysts' ratings were as follows: university setting, 93.33%; patient referral, 93.33%; professional therapists, 100%; treatment manual, 80%; monitoring treatment implementation, 80%; heterogeneous client characteristics, 86.67%; heterogeneous presenting problems, 100%; specially trained therapists, 80%; and limited treatment procedure, 100%. Percent agreement on categorization into stages was as follows: Stage 1, 86.67%; Stage 2, 73.33%; and Stage 3, 100%. Ordinarily with dichotomous data, one should measure reliability with coefficient kappa, which corrects for chance agreement. In the present case, however, kappa could not be computed for most variables because agreement was so high that either a row or column had no entries.

Results

Of the 48 authors originally contacted, 24 authors informed us about the outcome of their searches. Of these 24, 11 authors of meta-analyses that included a total of 486 studies reported that none of their studies met the Stage 1 criteria (Allen, Hunter, & Donohue, 1989; Baker, Swisher, Nadenichek, & Popowicz, 1984; Barker, Funk, & Houston, 1988; Blanchard, Andrasik, Ahles, Teders, & O'Keefe, 1980; Dewey & Hansley, 1990; Hahlweg & Markman, 1988; Hill, 1987; Kazdin, Bass, Ayers, & Rogers, 1990; Knight, Lutzky, & Macofsky-Urban, 1993; Lambert, Hatch, Kingston, & Edwards, 1986; Shoham-Salomon & Rosenthal, 1987). The remaining 13 authors (of 15 meta-analyses) identified at least one clinically representative study according to the Stage 1 criteria (Crits-Christoph, 1992; Hazelrigg, Cooper, & Borduin; 1987; Jorm, 1989; Lyons & Woods, 1991; Navarro, 1990, 1993; Nietzel, Russell, Hemmings, & Gretter, 1987; Prout & DeMartino, 1986; Robinson, Berman, & Neimeyer, 1990; Shadish, Montgomery, Wilson, Wilson, Bright, & Okwumabua, 1993; Smith, Glass, & Miller, 1980; Svartberg & Stiles, 1991; Trull, Nietzel, & Main, 1988; Weisz, Weiss, Alicke, & Klotz, 1987; Wittmann & Matt, 1986). Results from these meta-analyses are presented in Table 1, with Stage 1 data presented in Figure 1. It is worth noting that the meta-analyses in Table 1 cover a diverse and relatively comprehensive array of psychotherapeutic interventions (e.g., rational-emotive therapy, brief therapy, marital and family therapy, child therapy, general therapy), of presenting problems (e.g., depression, school problems, agoraphobia, anxiety, neuroticism), and even of cultures (e.g., United States, Germany, Spain, Latin America). By contrast, the 11 meta-analyses in which no studies met the clinical relevance criteria often focused more narrowly on specific problems or therapies, including public speaking anxiety, paradoxical interventions, headaches, caregiver distress, and nonspecific effects. This latter group also had fewer large, comprehensive meta-analyses of therapy. Hence, the 15 meta-analyses that did have studies passing some clinical relevance criteria probably reflect a reasonably comprehensive overview of therapy studies.

The first column of Table 1 presents the average effect size reported in the original meta-analysis and the number of studies on which that was based. Effect sizes and number of studies are presented for those studies passing the criteria in Stages 1, 2, and 3. In general, there are two results of interest. First, no matter how one counts the baseline total number of studies conducted, it is clear that relatively few studies of clinic therapy exist in the therapy research literature. The 15 meta-analyses in Table 1 listed a total of just over 1,000 studies in their bibliographies. A minimum of about 500 of these studies must be independent because the Smith et al. (1980) meta-analysis itself examined about 500 unduplicated studies. Moreover, we can be fairly certain the count is greater than 500 because Smith et al.'s list ends in the mid-1970s, and we can reasonably suppose that a large number of psychotherapy experiments have been conducted in the 20 years since then. Counting the 486 studies in the 11 meta-analyses that failed all the Stage 1 criteria, the total number of unduplicated studies is probably at least 1,000. Even if the count is only 500, however, it would imply no important change to our conclusion that relatively few clinically representative studies have been conducted. Only 56 studies passed the relatively modest Stage 1 criteria of being conducted outside a university, involving patients referred through usual clinical routes and using experienced professional therapists with regular case loads.[2] If we add to these criteria the Stage 2 requirement that the study neither use a treatment manual nor monitor implementation of treatment, the number of clinic ther-

[2] None of six studies that Weisz et al. (1992) described as clinically representative are in the present sample. This occurred because either (a) the studies were not in any of the meta-analyses surveyed or (b) if they were in the meta-analyses, the meta-analysts did not report them as passing the Stage 1 criteria. References for 51 of the 56 Stage 1 studies are listed in the reference section. References for the remaining 5 could not be determined; these 5 were part of the 20 Stage 1 studies from the Smith et al. (1980) meta-analysis. Although the data for these five studies were available, the identity of the reference had been lost.

Table 1
Mean Posttest Effect Sizes (and Standard Errors), and Number of Studies for Psychotherapy Meta-Analyses With Clinically Representative Studies

Author and topic	Effect size (*SE*) and *n*			
	All studies	Stage 1	Stage 2	Stage 3
Crits-Christoph (1992)	0.91[a,b]	0.67[a] (0.18)		
Brief dynamic psychotherapy	*n* = 11	*n* = 2	*n* = 0	*n* = 0
Hazelrigg et al. (1987)	0.48[a]	0.88[a] (0.23)	0.89[a] (0.29)	0.51 (0.61)
Family therapies	*n* = 20	*n* = 3	*n* = 2	*n* = 1
Jorm (1989)	0.53[a]	0.47[a] (0.25)	0.63 (0.42)	
Psychotherapy for trait anxiety and neuroticism	*n* = 63	*n* = 2	*n* = 1	*n* = 0
Lyons & Woods (1991)	1.00[a,b]	1.11[a] (0.22)	1.27 (0.36)	
Rational-emotive therapy	*n* = 70	*n* = 3	*n* = 1	*n* = 0
Navarro (1990)	0.75[a]	0.70[a] (0.29)	0.36 (0.40)	
Spanish-language outcome studies	*n* = 44	*n* = 3	*n* = 1	*n* = 0
Navarro (1993)	0.18[a]	0.24 (0.21)		
Psychotherapy with Latinos in the United States	*n* = 11	*n* = 2	*n* = 0	*n* = 0
Nietzel et al. (1987)	0.71[a,c]	2.34[a] (0.60)		
Psychotherapy for depression	*n* = 31	*n* = 1	*n* = 0	*n* = 0
Prout & DeMartino (1986)	0.58[a,b]	0.67[a] (0.14)	0.67[a] (0.14)	
School-based psychotherapy	*n* = 33	*n* = 6	*n* = 6	*n* = 0
Robinson et al. (1990)	0.73[a]	0.40 (0.33)		
Psychotherapy for depression	*n* = 37	*n* = 1	*n* = 0	*n* = 0
Shadish et al. (1993)	0.51[a]	0.06 (0.25)	−0.08 (0.37)	
Marital and family psychotherapies	*n* = 71	*n* = 3	*n* = 1	*n* = 0
Smith, Glass, & Miller (1980)	0.60[a]	0.62[a] (0.06)		
All psychotherapy[d]	*n* = 476	*n* = 20		
Svartberg & Stiles (1991)	0.32[a,b]	−0.01 (0.18)	−0.02 (0.22)	
Short-term psychodynamic psychotherapy	*n* = 19	*n* = 3	*n* = 2	*n* = 0
Trull et al. (1988)	0.47[a,c]	0.95[a] (0.43)	0.95[a] (0.43)	
Psychotherapy for agoraphobia	*n* = 19	*n* = 1	*n* = 1	*n* = 0
Weisz et al. (1987)	0.79[a,b]	0.67 (0.37)		
Children and adolescents	*n* = 105	*n* = 1	*n* = 0	*n* = 0
Wittman & Matt (1986)	0.34[a]	0.48[a] (0.12)		
German-language outcome studies	*n* = 72	*n* = 5	*n* = 0	*n* = 0
No. (and %) of studies	1,082	56 (5.2%)	15 (1.4%)	1 (0.1%)

[a] The effect size is significantly different from zero, $p < .05$. [b] Unweighted effect estimate in this cell only; means in Stages 1–3 for this meta-analysis were weighted. [c] Some effect sizes in this cell compare treated group to normative data rather than no-treatment control; these were eliminated from cells at Stages 1–3. [d] Weighted estimates obtained from secondary analysis of original data set; Smith et al. (1980) did not code Stages 2 and 3 criteria.

apy studies drops to 15. Only one study (Katz, de Krasinski, Philip, & Wieser, 1975) passed the additional Stage 3 criteria that includes the complete set of criteria for clinic therapy. Despite some ambiguity about the number of unduplicated studies that should serve as the baseline for total number of studies conducted, it is very clear from these figures that the number of studies of clinic therapy is several orders of magnitude smaller than the total number of studies conducted.

Table 1 also reports the effect sizes and standard errors separately for all meta-analyses and for the three stages. In all results reported in this paragraph, average effect sizes for the three stages use Hedges and Olkin's (1985) weighting system, giving more weight to studies with larger sample sizes. This effect size data can (and probably should) be analyzed in a variety of different ways—we do so here to demonstrate that results are very consistent no matter how the data are analyzed. The simplest analysis averages the effect sizes from meta-analyses down each column of Table 1. Doing so yields $\bar{d} = .59$ over all studies, $\bar{d} = .68$ for the 56 Stage 1 studies, $\bar{d} = .58$ for the 15 Stage 2 studies, and $\bar{d} = .51$ for the Stage 3 study—although when interpreting this latter figure, it should be kept in mind that only one study passed the Stage 3 criteria, so the estimate may be quite unstable. In a preliminary report of these data (Matt, Shadish, Navarro, & Siegle, 1994), we also found that in each of the original meta-analyses the average effect size for those studies that passed the Stage 1 criteria was not significantly different from the effect size over all studies in that meta-analysis. Figure 1 illustrates this conclusion, for the overall results in Table 1 fall within the 95% confidence intervals for the Stage 1 data for every case except for Nietzel et al. (1987).[3] Also, although Table 1 does not present follow-up data sepa-

[3] Between that preliminary report and the present one, we changed our treatment of Nietzel et al. (1987) by omitting a large number of studies that had been erroneously included in our preliminary report but that had compared treatment to normative data in test banks.

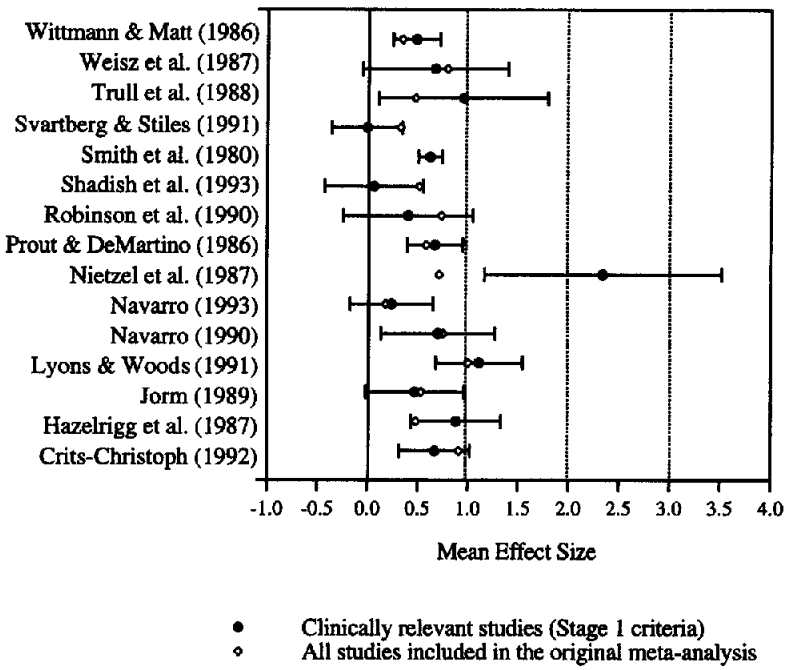

Figure 1. Mean effect sizes and 95% confidence intervals for Stage 1 clinic therapy studies.

rately, three of the meta-analyses in Table 1 had a total of four Stage 1 studies with a follow-up assessment of more than 4 weeks (Navarro, 1990, 1993; Robinson, Berman, & Neimeyer, 1990). The average follow-up effect size over these three meta-analyses was $d = .63$, about the same as the posttest effect sizes. This latter finding is consistent with previous meta-analytic results that follow-up effect sizes do not differ significantly from posttest effect sizes (Nicholson & Berman, 1983).

Taking the average of each column, however, underweights meta-analyses that contributed many studies. A better way to analyze the data is to pool all studies into one group instead of averaging each meta-analysis's average. Doing so yielded $\bar{d} = .56$ ($SE = .04$) for the 56 Stage 1 studies, $\bar{d} = .52$ ($SE = .09$) for the 15 Stage 2 studies, and $\bar{d} = .51$ ($SE = .61$) for the Stage 3 study. In a final refinement to this analysis, we eliminated the two duplicate studies reported by two meta-analyses in Table 1 and kept only randomized trials. Results yielded $\bar{d} = .54$ ($SE = .05$) for the 46 remaining Stage 1 studies and $\bar{d} = .52$ ($SE = .11$) for the 12 remaining Stage 2 studies. However, this analysis eliminated the one Stage 3 study that was a quasi-experiment—not a trivial loss because it means there is no randomized study in the therapy literature reviewed here that seems to meet all the criteria for clinic therapy.

These last figures are probably the best estimates of the effects of clinically representative therapy based on the data available to us, and the figures are quite similar to the overall $\bar{d} = .59$ reported for all therapy studies in the original meta-analyses. However, the effect sizes from clinic studies are about 10% smaller than over all therapy studies, so it is possible there is a very small decrease in effect size in clinic therapy, a possibility that we should leave open given the lack of good Stage 3 studies. However, that decrease might also be artifactual, related to our use of weights in analyzing the Stages 1, 2, and 3 data, something not systematically done by all of the original meta-analysts. Nine of the original meta-analyses used weights, but six meta-analyses used unweighted estimates. With access to the raw data set (Smith et al., 1980), we computed weighted estimates for one of those six and it is reported in the first column of Table 1; the remaining five are footnoted in Table 1. This makes the first column of Table 1 slightly different from the remaining columns where we ensured that weighted effect sizes were used in each case. More importantly, the direction of the bias would be to reduce the size of the effect in the clinic therapy estimates because the use of weights reduces average effect size. This decrease is illustrated by the change in the effect size estimates in the Smith et al. (1980) meta-analysis from their originally reported unweighted $\bar{d} = .85$ to a weighted $\bar{d} = .60$ that we recomputed over all studies in the original sample. Unfortunately, it was not feasible to have the remaining five meta-analysts recompute their original estimates over all studies with weighted least squares methods. Hence, we continue to use $\bar{d} = .59$ as the overall estimate from the first column of Table 1. The figure is probably reasonably accurate, given that most of the original meta-analyses did use weights. However, the reader should note that if weighting would have reduced that figure a little bit, it would probably make the difference between clinic therapy and research therapy even smaller than our data suggest. A related problem is that the overall effect size estimates in Table 1 contain the studies that are also listed for Stages 1 through 3, so comparisons between the overall results and the stages are not independent. However, given that so very few studies from the overall results were included in the data from Stages 1 through 3, this overlap is not likely to have any substantial effect on the comparison.

A final objection to the preceding analyses is that they lump together studies that are too diverse; for example, studies of interventions for children in schools, unspecified therapy for Latinos with unspecified problems, or rational-emotive therapy generally. The objection has force because good therapy research must not be oblivious to such factors as age of onset, course, severity, or responsiveness to differential treatment. In light of this objection, however, it is worth noting that the results of our analyses are remarkably consistent within each row of Table 1. In only 2 of the 15 meta-analyses in Table 1 did the effect size for Stage 1 studies drop substantially to a value close to zero (Shadish et al., 1993; Svartberg & Stiles, 1991). In the remaining 13, the effect size either remained about the same or increased and was almost always moderate or large in size. Similar observations hold for the Stage 2 results, as well. So our conclusions about clinic therapy effectiveness would hold even if we did not average results over all 56 studies.

Discussion

Although our findings are limited, they tentatively suggest that clients who receive therapy under clinical conditions do better than those who do not receive therapy, and that effect sizes for clinic therapy may be comparable with those yielded by past meta-analyses. However, these results are far from definitive, so that future researchers should consider three issues that our research makes salient: criteria for clinical representativeness, the need for more studies with which to test the clinical representativeness hypothesis, and alternative models of generalization.

Is it possible to reach agreement on criteria for which studies should count as clinically representative and which should not? Our own criteria are reasonable but not perfect, as the following six examples illustrate. First, some university clinics cater to a community population of normally referred clients with diverse diagnostic and personal characteristics; such clinics are sometimes major providers of services to a catchment area, particularly in rural areas. If so, the exclusion of university clinics from the clinically representative category might not be warranted. Second, clinical therapists sometimes are recently and intensively trained in the procedures they are about to use with a client. For example, especially as new procedures emerge and become more widely used, practicing therapists take workshops in the procedure. Even with older procedures, they may brush up their skills in a similar way (Norcross & Prochaska, 1982; Prochaska & Norcross, 1983). Such training may encourage these therapists to gravitate more toward use of those procedures after training. Third, rudimentary forms of treatment manuals are sometimes used in clinic therapy. For example, clinic therapists sometimes consult books of therapy procedures or casebooks for particular clinical conditions. Such references may serve a role somewhat like a treatment manual. In addition, recent attention to treatment guidelines (whether flexible or strict) in clinical practice resembles treatment manuals in some ways. Fourth, some therapists limit their practice to relatively homogenous clients. A marital therapist in private practice, for example, might well see mostly middle- to upper-class clients presenting with marital discord as the primary problem. Conversely, although research therapy often solicits clients with a focused problem, those focused problems are often accompanied by other pathology. Fifth, practicing therapists sometimes advertise their specialty in a way quite similar to experimental solicitation for research, for example, by radio or television appearances, and sometimes by the wording of their specialty in advertisements. For instance, a yellow pages entry that lists stress control as a specialty is quite likely to elicit a different kind of client with a more focused problem than is one that just lists clinical psychology as a specialty. Sixth, some managed care scrutiny might approximate the monitoring of therapy that occurs in some research. Various kinds of collegial review, such as occurs in case presentations or staffings, might also serve that function.

In view of this litany, we need improved empirical data about therapy practice so that we can better approximate the features of clinically representative research. Although our assumptions about the modal instance of therapy practice are plausible and supported by data (e.g., Norcross & Prochaska, 1982; Perlman, 1985; Prochaska & Norcross, 1983), those data are sparse and tend to be gathered from psychologists who are members of the American Psychological Association. Primary care physicians, social workers, and other mental health workers needs to be considered, as well. So, much as was the case when Norcross and Prochaska (1982) revisited Garfield and Kurtz's (1974) research on therapy practice, perhaps it is time to update our picture of the activities of current practicing psychotherapists.

The second issue that our research makes salient is the relative lack of studies done under clinical conditions. Few studies met our Stage 1 criteria, and more importantly, only one quasi-experiment met the full set of Stage 3 criteria, so we cannot be confident that the Stage 3 results are either stable and unbiased. The therapy outcome literature needs new experiments designed with clinically representative facets in mind, especially randomized experiments because they are now completely lacking. The minimum needed is random assignment, low attrition, and a reliable and valid posttest (adding a pretest is helpful for other reasons such as sorting out effects of attrition). To mimic clinic practice, such research need not use treatment manuals or monitor treatment (except enough to report what therapy was studied); nor would it need patient selection processes beyond those minimally necessary to meet ethical obligations such as screening for and treating suicidal patients. Of course, therapy researchers should also study clinic therapies with other designs, as well, such as high-quality quasi-experiments (Shadish & Ragsdale, 1996), or single-case designs. Similarly, the recent *Consumer Reports* survey of therapy (Seligman, 1995) yielded qualitative conclusions much like those from many (but not all) meta-analyses of experiments but did so with a sample of patients that was arguably more clinically representative. However, many such studies already exist. It is the high-quality randomized experiments about clinic therapy that are currently lacking.

The third issue that our research makes salient is how to conceptualize and study therapy generalization. Random sampling of patients, therapies, settings, outcome measures, or time is rarely feasible in therapy outcome research. What other options might facilitate generalization? We noted in the introduction that the present study uses proximal similarity, but it has both flaws and alternatives (Cook, 1990; Shadish, 1995). It assumes that superficial features (e.g., the criteria used in this

article) are most responsible for the generalization of results, but such features can be misleading. For example, if possible key features causing therapy outcome, such as the amount of fee charged, are not salient in research reports, that feature might be constant between studies of research and clinic therapy (if neither charged for therapy), biasing both kinds of studies towards similar but incorrect answers relative to actual clinical practice.

To complement the emphasis on superficial features, an alternative is to study the underlying causal processes rather than key superficial features. The assumption is that if one can identify those processes, then one can know exactly what the crucial features are to transfer effective therapy to new settings. Social psychologists have used this model for years (Jones, 1985), with the advantage that one can often investigate many underlying causal processes in the lab before studying whether those mechanisms work well in clinical conditions (Cohen, 1994). Therapy research needs such studies (e.g., Kendall & Southam-Gerow, 1995).

Some therapy researchers may use a technology transfer model of generalization like that in some medical (Greenwald & Cullen, 1984; Office of Technology Assessment, 1982) or psychiatric literatures (Hoagwood, Hibbs, Brent, & Jensen, 1995). In the medical case, for example, interventions are first pilot tested for feasibility, acceptability, and risks. In a second phase, they are evaluated in controlled clinical trials (e.g., case studies, quasi-experiments, randomized trials) of whether they are efficacious under ideal conditions (i.e., efficacy studies); most of the studies in the current therapy outcome literature may be of this kind. The third phase involves defined population studies to measure the impact of an intervention in sizable, distinct, and well-characterized populations outside the laboratory. The kinds of studies recommended by Division 12 of the American Psychological Association in their report on empirically validated psychological treatments (Chambless et al., 1996; Task Force on Promotion and Dissemination of Psychological Procedures, 1995) probably fall into both these second and third phases. The fourth phase consists of demonstration and implementation studies to apply the intervention and measure its impact on public health. Interventions with demonstrable public health benefits may be recommended for adoption in a particular population and under specific conditions, accompanied by education and training. In a final phase, an intervention can be evaluated as it is used in clinical practice (i.e., effectiveness studies). Of course, effectiveness studies need not be preceded by efficacy studies. However, a technology transfer model calls for a program of research that goes from laboratory to field. Such technology transfer programs of research are mostly lacking in therapy research—although the recently approved template for the evaluation of treatment guidelines of the American Psychological Association is consistent with the spirit of such a model (Task Force on Psychological Intervention Guidelines, 1995).

Conclusion

The generalizability of therapy research continues to concern practitioners and researchers for two reasons. First, the tradeoff between high-quality causal inference (internal validity) and generalization of that inference (external validity) is to some extent part of the structure of therapy experiments, as features that facilitate high-quality causal inference (e.g., a population that will accept random assignment) can reduce generalization (e.g., those willing to be randomly assigned may differ from the population of interest). Second, the pertinent policy issue—whether the research literature supports the effectiveness of clinical practice for purposes such as insurance reimbursement—should be salient in future policy debates. For such reasons, we hope the present article stimulates more conceptual and empirical attention to this issue.

References

(References preceded by an asterisk were among the 56 studies identified in Stage 1 of the meta-analysis)

Allen, M., Hunter, J. E., & Donohue, W. A. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. *Communication Education, 38,* 54–76.

*Amerikaner, M., & Summerlin, M. L. (1982). Group counseling with learning disabled children: Effects of social skills and relaxation training on self-concept and classroom behavior. *Journal of Learning Disabilities, 15,* 340–343.

Andrews, G., & Harvey, R. (1981). Does psychotherapy benefit neurotic patients? *Archives of General Psychiatry, 38,* 1203–1208.

*Antons, K. (1972). Soziometrische Kurwirkungen. Eine Test-Retestuntersuchung mit zwei Persönlichkeitsinventaren am Vorsorgeheim Hundseck der Bundesknappschaft Bochum [Sociometric effects of a health intervention. A test–retest study with two personality inventories at the Hundseck clinic of the miners' health organization Bochum]. *Zeitschrift fur psychosomatische Medizin und Psychoanalyse, 18,* 369–389.

Baker, S. B., Swisher, J. D., Nadenichek, P. E., & Popowicz, C. L. (1984). Measured effects of primary prevention strategies. *Personnel and Guidance Journal, 62,* 459–464.

Barker, S. L., Funk, S. C., & Houston, B. K. (1988). Psychological treatment versus nonspecific factors: A meta-analysis of conditions that engender comparable expectations for improvement. *Clinical Psychology Review, 8,* 579–594.

*Barron, F., & Leary, T. F. (1955). Changes in psycho-neurotic patients with an without psychotherapy. *Journal of Consulting and Clinical Psychology, 19,* 239–245.

Blanchard, E. B., Andrasik, F., Ahles, T. A., Teders, S. J., & O'Keefe, D. (1980). Migraine and tension headache: A meta-analytic review. *Behavior Therapy, 14,* 613–631.

*Block, J. (1978). Effects of a rational-emotive mental health program on poorly achieving, disrupting high school students. *Journal of Counseling Psychology, 25,* 61–65.

*Brodaty, H., & Andrews, G. (1983). Brief psychotherapy in family practice: A controlled prospective intervention trial. *British Journal of Psychiatry, 143,* 11–19.

Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 67–77). San Francisco: Jossey-Bass.

*Cattell, R. B., Rickels, K., Weise, C., Gray, B., & Yee, R. (1966). The effects of psychotherapy upon measured anxiety and regression. *American Journal of Psychotherapy, 20,* 261.

Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49,* 5–18.

*Coche, E., & Flick, A. (1975). Problem-solving training groups for hospitalized patients. *Journal of Psychology, 91,* 19–29.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

*Coleman, M., & Glofka, P. T. (1969). Effect of group therapy on self-concept of senior nursing students. *Nursing Research, 18,* 274–275.

*Comas-Diaz, L. (1981). Effects of cognitive and behavioral group treatment on the depressive symptomalogy of Puerto Rican women. *Journal of Consulting and Clinical Psychology, 49,* 627–632.

Consumer Reports. (1995, November). Mental health: Does therapy help? pp. 734–739.

Cook, T. D. (1990). The generalization of causal connections: Multiple theories in search of clear practice. In L. Sechrest, E. Perrin, & J. Bunker (Eds.), *Research methodology: Strengthening causal interpretations of nonexperimental data* (DHHS Publication No. PHS 90-3454, pp. 9–31). Rockville, MD: Department of Health and Human Services.

Crits-Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: A meta-analysis. *American Journal of Psychiatry, 149,* 151–158.

Depression Guideline Panel. (1993, April). *Depression in Primary Care: Vol. 2. Treatment of major depression.* Clinical Practice Guideline, No. 5. (AHCPR Publication No. 93-0551). Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Dewey, D., & Hansley, J. (1990). The effects of marital adjustment and spouse involvement on the behavioral treatment of agoraphobia: A meta-analytic review. *Anxiety Research, 2,* 69–83.

Dole, A. A., Rockey, P. B., & DiTomasso, R. (1983, April). *Meta-analysis of outcome research in reducing test anxiety: Interventions, rigor and inertia.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

*Endicott, N. A., & Endicott, J. (1964). Prediction of improvement in treated and untreated patients using the Rorschach prognostic rating scale. *Journal of Consulting Psychology, 28,* 342–348.

*Ezzo, F. R. (1980). A comparative outcome study of family therapy and positive parenting with court-referred adolescents. *Dissertation Abstracts International, 40A,* 6198. (University Microfilms No. 80-13836)

*Fernandez-Rodriguez, C., & Perez-Alvarez, M. (1987). Modificacion de conducta y mejora en el cumplimiento del tratamiento en diab ticos tipo II [Behavior modification and improvement of adherence to treatment among Type II diabetes patients]. *Revista Espanola de Terapia del Comportamiento, 5,* 233–248.

*Fischer, J., Anderson, J., Arveson, E., & Brown, S. (1978). Alderian family counseling: An evaluation. *International Journal of Family Counseling, 6,* 42–44.

*Florin, I., Cohen, R., & Meyer-Osterkamp, S. (1973). Eine Untersuchung zum operanten Konditionieren sozialen Verhaltens bei chronisch Schiziphrenen [A study of operant conditioning of social behavior in chronic schizophrenics]. *Zeitschrift fur Klinische Psychologie, 2,* 243–277.

Garfield, S. L., & Kurtz, R. (1974). A survey of clinical psychologists: Characteristics, activities, and orientations. *The Clinical Psychologist, 28,* 7–10.

*Gerler, E. R. (1980). A longitudinal study of multimodal approaches to small group psychological education. *School Counselor, 27,* 184–190.

Greenwald, P., & Cullen, J. W. (1984). The scientific approach to cancer control. *CA-A Cancer Journal for Clinicians, 34,* 328–332.

Hahlweg, K., & Markman, H. J. (1988). Effectiveness of behavioral marital therapy: Empirical status of behavioral techniques in preventing and alleviating marital distress. *Journal of Consulting and Clinical Psychology, 56,* 440–447.

*Hannemann, E. (1979). Short-term family therapy with juvenile status offenders and their families. *Dissertation Abstracts International, 40,* 1894B. (University Microfilms No. 79-22867)

Hazelrigg, M. D., Cooper, H. M., & Borduin, C. M. (1987). Evaluating the effectiveness of family therapies: An integrative review and analysis. *Psychological Bulletin, 101,* 428–442.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Hill, K. A. (1987). Meta-analysis of paradoxical interventions. *Psychotherapy, 24,* 266–270.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to special section: Efficacy and effectiveness studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 683–687.

Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1, pp. 47–107). New York: Random House.

*Jones, F. D., & Peters, H. N. (1952). An experimental evaluation of group psychotherapy. *Journal of Abnormal and Social Psychology, 47,* 345–353.

Jorm, A. F. (1989). Modifiability of trait anxiety and neuroticism: A meta-analysis of the literature. *Australian and New Zealand Journal of Psychiatry, 23,* 21–29.

*Kassinove, H., Miller, N., & Kalin, M. (1980). Effects of pretreatment with rational emotive bibliotherapy and rational emotive audiotherapy on clients waiting at community mental health centers. *Psychological Reports, 46,* 851–857.

*Katz, A., de Krasinski, M., Philip, E., & Wieser, C. (1975). Change in interactions as a measure of effectiveness in short-term family therapy. *Family Therapy, 2,* 31–56.

Kazdin, A. E. (1978). Evaluating the generality of findings in analogue therapy research. *Journal of Consulting and Clinical Psychology, 46,* 673–686.

Kazdin, A. E. (1986). The evaluation of psychotherapy: Research design and methodology. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed., pp. 23–68). New York: Wiley.

Kazdin, A. E., Bass, D., Ayers, W. A., & Rogers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58,* 729–740.

Kendall, P. C., & Southam-Gerow, M. A. (1995). Issues in transportability of treatment: The case of anxiety disorders in youth. *Journal of Consulting and Clinical Psychology, 63,* 702–708.

*Kilman, P. C., Henry, S. E., Scarbro, H., & Laughlin, J. E. (1977). The impact of affective education on elementary school underachievers. *Psychology in the Schools, 16,* 217–233.

Knight, B. G., Lutzky, S. M., & Macofsky-Urban, F. (1993). A meta-analytic review of interventions for caregiver distress: Recommendations for future research. *Gerontologist, 33,* 240–248.

*Krop, N., Calhoon, B., & Verrier, R. (1971). Modification of the "self-concept" of emotionally disturbed children by covert reinforcement. *Behavior Therapy, 2,* 201–204.

Lambert, M. J., Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. *Journal of Consulting and Clinical Psychology, 54,* 54–59.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209.

*Lipsky, M. J., Kassinove, H., & Miller, N. J. (1980). Effects of rational-emotive therapy, rational role reversal and rational-emotive imagery on the emotional adjustment of community mental health center patients. *Journal of Consulting and Clinical Psychology, 48,* 366–374.

Lyons, L. C., & Woods, P. J. (1991). The efficacy or rational-emotive

therapy: A quantitative review of the outcome research. *Clinical Psychology Review, 11,* 357–369.

*Macia, D., & Mendez, F. X. (1986). Programa de intervencion conductual para el cumplimiento de las prescripciones medicas [Behavioral intervention program to enhance adherence to medical prescriptions]. *Revista de Psicologia General y Aplicada, 41,* 369–377.

*Maldonado, A. (1984). Terapia de conducta y depresion: Un an lisis experimental de las interacciones entre tratamientos cognitivos y conductuales con tratamientos farmacologicos en sujetos depresivos [Behavior therapy and depression: An experimental analysis of the interactions between cognitive–behavioral and pharmacological treatments in depressive patients]. *Revista de Psicologia General y Aplicada, 39,* 517–535.

Matt, G. E., Shadish, W. R., Navarro, A. M., & Siegle, G. (1994, August). *The effects of psychotherapy conducted under clinically representative conditions: A reanalysis of psychotherapy meta-analyses.* Paper presented at the 102nd Annual Convention of the American Psychological Association, Los Angeles.

*Mavissakalian, M., & Michelson, L. (1983). Self-directed in vivo exposure practice in behavioral and pharmacological treatments of agoraphobia. *Behavior Therapy, 14,* 506–519.

*Meyer, A. E. (1981). The Hamburg Short Psychotherapy Comparison Experiment. *Psychotherapy and Psychosomatics, 35,* 77–212.

*Moleski, R., & Tosi, D. J. (1976). Comparative psychotherapy: Rational-emotive therapy versus systematic desensitization in the treatment of stuttering. *Journal of Consulting and Clinical Psychology, 44,* 309–311.

*Morrow, G. R. (1986). Effect of the cognitive hierarchy in the systematic desensitization treatment of anticipatory nausea in cancer patients: A component comparison with relaxation only, counseling, and no treatment. *Cognitive Therapy and Research, 10,* 421–466.

Navarro, A. M. (1990). Eficacia de las terapias psicológicas: Metaanálisis de estudios en paises de habla hispana [Effectiveness of psychological therapies: Meta-analysis of outcome studies in Spanish-speaking countries]. *Revista de Psicología Social y Personalidad, 6,* 87–99.

Navarro, A. M. (1993). Efectividad de las psicoterapias con Latinos en los Estados Unidos: Una revisión meta-analítica [Effectiveness of psychotherapy with Latinos in the United States: A meta-analytic review]. *Interamerican Journal of Psychology, 27,* 131–146.

Nicholson, R. A., & Berman, J. S. (1983). Is follow-up necessary in evaluating psychotherapy? *Psychological Bulletin, 93,* 261–278.

Nietzel, M. T., Russell, R. L., Hemmings, K. A., & Gretter, M. L. (1987). Clinical significance of psychotherapy for unipolar depression: A meta-analytic approach to social comparison. *Journal of Consulting and Clinical Psychology, 55,* 156–161.

Norcross, J. C., & Prochaska, J. O. (1982). A national survey of clinical psychologists: Characteristics and activities. *Clinical Psychologist, 35,* 5–8.

Office of Technology Assessment. (1982). *Technology transfer at the National Institutes of Health. A technical memorandum* (Library of Congress Catalog Card Number 82-600529). Washington, DC: U.S. Government Printing Office.

Office of Technology Assessment. (1994, September). *Identifying health technologies that work: Searching for evidence* (Publication no. OTA-H-608). Washington, DC: U.S. Government Printing Office.

*Omizo, M. M., & Michael, W. B. (1982). Biofeedback-induced relaxation training and impulsivity, attention to task, and locus of control among hyperactive boys. *Journal of Learning Disabilities, 15,* 414–416.

*Parson, B. V., & Alexander, J. F. (1973). Short-term family intervention: A therapy outcome study. *Journal of Consulting and Clinical Psychology, 41,* 195–201.

Perlman, B. (1985). A national survey of APA-affiliated masters-level clinicians: Description and comparison. *Professional Psychology: Research and Practice, 16,* 553–564.

*Persons, R. W., & Pepinsky, H. B. (1966). Convergence in psychotherapy with delinquent boys. *Journal of Counseling Psychology, 13,* 329–334.

*Piper, W. E., Azim, H. F., McCallum, M., & Joyce, A. S. (1990). Patient suitability and outcome in short-term individual psychotherapy. *Journal of Consulting and Clinical Psychology, 58,* 475–481.

*Poser, E. G. (1966). The effects of therapists' training on group therapeutic outcome. *Journal of Consulting Psychology, 30,* 283–289.

Prochaska, J. O., & Norcross, J. C. (1983). Contemporary psychotherapists: A national survey of characteristics, practices, orientations, and attitudes. *Psychotherapy: Theory, Research & Practice, 20,* 161–173.

Prout, H. T., & DeMartino, R. A. (1986). A meta-analysis of school-based studies of psychotherapy. *Journal of School Psychology, 24,* 285–292.

Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin, 108,* 30–49.

*Sacks, J. M., & Berger, S. (1954). Group therapy techniques with hospitalized chronic schizophrenic patients. *Journal of Consulting Psychology, 18,* 297–302.

*Schandl, V., & Loschenkohl, E. (1980). Kind im Krankenhaus: Evaluierung eines Interventionsprogrammes bei Verhaltensstorungen [Children in hospitals: An evaluation of an intervention program for behavioral disorders]. *Praxis der Kinderpsychologie und Kinderpsychiatrie, 29,* 252–258.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50,* 965–974.

Shadish, W. R. (1995). The logic of generalization: Five principles common to experiments and ethnographies. *American Journal of Community Psychology, 23,* 419–428.

Shadish, W. R., Montgomery, L. M., Wilson, P., Wilson, M. R., Bright, I., & Okwumabua, T. (1993). Effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology, 61,* 992–1002.

Shadish, W. R., & Ragsdale, K. (1996). Random versus nonrandom assignment in psychotherapy experiments: Do you get the same answer? *Journal of Consulting and Clinical Psychology, 64,* 1290–1305.

Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92,* 581–604.

*Shefler, G., & Dasberg, H. (1989, June). *A randomized controlled outcome and follow-up study of James Mann's time-limited psychotherapy in a Jerusalem community mental health center.* Paper presented at the annual convention of the Society for Psychotherapy Research, Toronto, Ontario, Canada.

Shoham-Salomon, V., & Rosenthal, R. (1987). Paradoxical interventions: A meta-analysis. *Journal of Consulting and Clinical Psychology, 55,* 22–28.

*Sloan, R. B., Staples, F. R., Cristol, A. H., Yorkston, N. J., & Whipple, K. (1975). *Psychotherapy versus behavior therapy.* Cambridge MA: Harvard University Press.

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

*Steier, F. (1983). Family interaction and properties of self-organizing systems: A study of family therapy with addict families. *Dissertation Abstracts International, 44,* 863A. (University Microfilms No. 83-16093)

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63,* 182–196.

Steinbrueck, S. M., Maxwell, S. E., & Howard, G. S. (1983). A meta-

analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults. *Journal of Consulting and Clinical Psychology, 51,* 856–863.

*Stotsky, B. A., Daston, P. G., & Vardack, C. N. (1955). An evaluation of the counseling of chronic schizophrenics. *Journal of Counseling Psychology, 2,* 248–255.

*Stover, L., & Guerney, B. (1967). The efficacy of training procedures for mothers in filial therapy. *Psychotherapy: Theory, Research, and Practice, 4,* 110–115.

Svartberg, M., & Stiles, T. C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 59,* 704–714.

*Szapocznik, J., Santisteban, D., Rio, A., Perez-Vidal, A., Santisteban, D., & Kurtines, W. M. (1989). Family effectiveness training: An intervention to prevent drug abuse and problem behaviors in Hispanic adolescents. *Hispanic Journal of Behavioral Sciences, 11,* 4–27.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995, Winter). Training in and dissemination of empirically validated psychological treatments: Report and recommendations. *The Clinical Psychologist, 48,* 3–24.

Task Force on Psychological Intervention Guidelines. (1995, July 21). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders.* Washington DC: American Psychological Association.

*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.

*Truax, C. B., Wargo, D. G., & Silber, L. D. (1966). Effects of group psychotherapy with high accurate empathy and nonpossessive warmth upon female institutionalized delinquents. *Journal of Abnormal Psychology, 71,* 267–274.

Trull, T. J., Nietzel, M. T., & Main, A. (1988). The use of meta-analysis to assess the clinical significance of behavior therapy for agoraphobia. *Behavior Therapy, 19,* 527–538.

*Ullrich de Muynck, R., & Ullrich, R. (1980). Erster Effizienznachweis des Assertiveness-Training-Programm (ATP) [First evidence for the effectiveness of the Assertiveness-Training-Program (ATP)]. In R. Ullrich (Ed.), *Soziale Kompetenz* (Bd. II, pp. 1–20). Munich: Pfeiffer.

*Ullrich de Muynck, R., & Ullrich, R. (1980). Spezifische Effekte des ATP im Arbeits-und Leistungsbereich [Specific effects of ATP in the areas of work and performance]. In R. Ullrich (Ed.), *Soziale Kompetenz* (Bd. II, pp. 21–32). Munich: Pfeiffer.

*Webster-Stratton, C. (1984). Randomized trial of two parent-training programs for families with conduct-disordered children. *Journal of Consulting and Clinical Psychology, 52,* 666–678.

*Weissman, M. M., Prusoff, B. A., DiMascio, A., Neu, C., Goklaney, M., & Klerman, G. L. (1979). The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes. *American Journal of Psychiatry, 136,* 555–558.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 688–701.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542–549.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47,* 1578–1585.

*White, K., & Allen, R. (1971). Art counseling in an educational setting: Self-concept change among pre-adolescent boys. *Journal of School Psychology, 9,* 218–225.

Wittmann, W. W., & Matt, G. E. (1986). Meta-Analyse als Integration von Forschungsergebnissen am Beispiel deutschsprachiger Arbeiten zur Effektivität von Psychotherapie [Meta-analysis as a method for integrating psychotherapy studies in German-speaking countries]. *Psychologische Rundschau, 37,* 20–40.

*Wolk, R. L., & Goldfarb, A. I. (1967). The response to group psychotherapy of aged recent admissions compared with long-term mental hospital patients. *American Journal of Psychiatry, 123,* 1251–1257.

Received March 27, 1996
Revision received July 30, 1996
Accepted September 17, 1996 ■