META-ANALYSIS OF MFT INTERVENTIONS

William R. Shadish and Scott A. Baldwin
The University of Memphis

This article briefly reviews 20 meta-analyses of marital and family interventions. These meta-analyses support the efficacy of both MFT for distressed couples, and marital and family enrichment. Those effects are slightly reduced at follow-up, but still significant. Differences among kinds of marital and family interventions tend to be small. MFT produce clinically significant results in 40-50% of those treated, but the effects of MFT in clinically representative settings have not been much studied. The article also introduces the concept of meta-analytically supported treatments (MASTs), which are treatments that meet certain criteria for efficacy in meta-analysis, and which remedy certain problems in the empirically supported treatment (EST) literature. The article concludes with recommendations for doing better meta-analyses.

Moses Herzog, the fictional academic whose moniker is the title for Saul Bellow's (1964) novel, once said "What this country needs is a good five-cent synthesis" (p. 207). Well, we are pleased to report that we are halfway toward that goal. In meta-analysis, we do indeed have a good methodology for the synthesis of scientific results. Unfortunately, as we will see later in this chapter, doing meta-analysis nowadays costs a lot more than just five cents.

The development and widespread use of meta-analysis is significant to both researcher and clinician in marriage and family interventions. Researchers benefit from a statistical tool that can be used to summarize the increasingly large research literature on such interventions, and which points to gaps in the literature that future research should address. Clinicians benefit in three ways: first in getting evidence they can show to third-party payers that the work they do is effective, second in having a practical way to inform themselves about the effectiveness and efficacy of marriage and family interventions, and third in a host of specific conclusions that may help them in choosing treatments proven to be effective for different problems. These uses resemble the movement in medicine and public health toward evidence-based medicine, a good model from which to view the meta-analytic literature.

This chapter has the following structure. First, we provide a brief history of meta-analysis, and summarize the key statistical feature of meta-analysis—the effect size. The latter material can be skipped by those with little interest in the methodology of meta-analysis. Second, we describe 20 meta-analyses that have already been done on the effects of both therapy and enrichment interventions with couples and families. In this second section, we summarize the overall results of these 20 meta-analyses, and then present more detailed discussions of the effects of different kinds of marriage and family interventions, the effects of marriage and family interventions compared to other kinds of intervention such as individual therapy, the clinical significance of these effects, the clinical representativeness of this research, and some intriguing findings about variables that may influence how effective marriage and family interventions may be. Also in this second section, we present the idea of Meta-Analytically Supported Treatments (MASTs); that is, treatments that have been shown to be effective in meta-

William R. Shadish, PhD, and Scott A. Baldwin, MS, Department of Psychology, The University of Memphis.

Correspondence concerning this article should be addressed to William R. Shadish, College of Social Sciences, Humanities, and Arts, The University of California at Merced, PO Box 2039, Merced, California, 95344. E-mail: w.shadish@mail.psyc.memphis.edu

analytic work. Third, we review methodological problems in this research, and offer a set of suggestions for improving future meta-analyses in this area. Fourth, we present evidence about the costs of meta-analytic research, and review possible funding sources.

META-ANALYSIS

A Brief History of Meta-Analysis

If you have been around long enough as a researcher, you have seen many innovations occur. Some of those innovations are passing fads, some are interim developments until something better comes along, and some become a minor but permanent feature of the scientific landscape. Occasionally, however, an innovation changes the very landscape of the field. Meta-analysis is the latter kind of innovation. Meta-analysis has become a nearly essential element in reviewing literatures on treatment effectiveness. By the 1990s, at least 1,000 meta-analyses had been done, and today the number is so large that there is no authoritative count.

Meta-analysis first came into the scientific community's spotlight with Smith and Glass's (1977) review of the effects of psychotherapy. Smith and Glass reported the results of a massive quantitative summary of 375 studies of psychotherapy effectiveness (later expanded to 475 studies; Smith, Glass, & Miller, 1980). The studies were conducted over many years, over different geographic areas with diverse clients, using different types of therapies and outcome measures. Over all this variability, clients receiving psychotherapy had better outcomes than control clients who received no treatment. Therapy outcomes were similar whether the therapies were behavioral or nonbehavioral, over varying levels of therapist experience, and over both brief and long-term therapy. With a few exceptions, the effects of therapy generalized over great heterogeneity in many moderators across these studies.

Today, 25 years later, we are aware of more than 140 meta-analyses in the field of psychotherapy alone.[1] The questions asked in these reviews range from very broad inquiries into the effects of psychotherapy in general (Smith & Glass, 1977), to very narrow examinations like Shoham-Salomon and Rosenthal's (1987) meta-analysis of the effects of including paradoxical interventions in psychotherapy. Among all the many psychotherapy meta-analyses, a number address aspects of marriage and family interventions, both with distressed and nondistressed populations. This latter set of meta-analyses is the topic of this chapter. Therefore, although this chapter is in some respects an update of the Shadish, Ragsdale, Glaser, and Montgomery (1995) article on marriage and family meta-analysis, it differs from that article in one important way. Shadish et al. (1995) described the results of a single meta-analysis, but the present chapter reviews results from many meta-analyses. In that sense, this chapter is not a meta-analysis, but a "meta-meta-analysis," attempting to provide both quantitative and qualitative summaries of what we have learned from the many meta-analyses reported herein.

What Is Meta-Analysis?

Meta-analysis is the use of quantitative techniques to summarize the results of scientific studies on the same question. This practice has a long history. In the early 18th century, the English mathematician Roger Cotes computed weighted averages of measurements made by different astronomers. In 1904, Sir Karl Pearson used quantitative methods to average results from six studies of the effects of a newly developed inoculation against typhoid (Shadish & Haddock, 1994; see also Cooper & Hedges, 1994; Dickerson, Higgins, & Meinert, 1990; and Hunter & Schmidt, 1990, for other historical examples). However, widespread adoption of such methods did not occur until Glass (1976) coined the term meta-analysis to describe quantitative techniques for cumulating results over studies.

Glass (1976) distinguished between primary, secondary, and meta-analysis, with these three phrases intended to differentiate between three different types of statistical analyses. According to Glass, primary analysis is "the original analysis of data in a research study" (p. 3). Secondary analysis is "the re-analysis of data for the purposes of answering the original research question with better statistical techniques, or answering new questions with old data" (p. 3). Meta-analysis is "the statistical

analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3).

The essential innovation in meta-analysis was the use of an effect size as a common metric over studies to measure how large is the effect of a treatment. A common metric is needed because different studies rarely use identical outcome measures, even if they address similar questions and invoke similar outcome constructs. Thus, one study of psychotherapy for depression might have used the Beck Depression Inventory while another used the Minnesota Multiphasic Personality Inventory (MMPI) Depression Scale. These measures have different metrics with different means and standard deviations, so averaging scores without converting them into a common metric would yield nonsensical results. Meta-analysis converts each study's outcomes to a common effect size metric, so that different outcomes have the same means and standard deviations and can be more readily averaged across studies.

Many different effect size measures are possible (Fleiss, 1994; Rosenthal, 1994). The two most appropriate effect size measures for meta-analyses of experiments are the standardized mean difference statistic (d) for continuous outcomes (Smith et al., 1980, Appendix 7; Shadish, Robinson, & Lu, 1999), and the odds ratio (o) for dichotomous outcomes (Haddock, Rindskopf, & Shadish, 1998). In the social sciences, d is the most common, and when both continuous measures (e.g., the Dyadic Adjustment Scale) and dichotomous measures (marriage or divorce at the end of therapy) were used as outcomes, the odds ratio is typically converted into d so that all results are in the same effect size metric. Cohen (1988) suggested that $d = .20$ is a small effect, $d = .50$ is a medium effect, and $d = .80$ is a large effect, norms that seem to match empirical observations of the prevalence of effects fairly well (Lipsey & Wilson, 1993).

A common concern of those who first encounter meta-analysis is the risks of combining diverse measures. After all, simply converting diverse measures into a common metric does not mean that it makes sense conceptually to combine them. To pick an extreme example, suppose an experiment measures family therapy outcome as the number of fights a couple has in a week and also as the child's grade point average. Does it make sense to combine these outcomes, especially if they are likely to be uncorrelated? This question is worth chapter length treatment itself, but here are some brief observations. First, it makes no difference whether the measures are correlated. Meta-analysis is not scale development in which items are expected to measure the same thing, and it is an empirical question whether the effects of therapy on both measures are similar despite their apparent conceptual differences. Second, meta-analysis has methods for testing whether it makes sense to combine effect sizes (e.g., heterogeneity tests), and recommended procedures for combining measures that violate these tests (e.g., random effects models). Third, the meta-analyst can always separate effect sizes into categories (e.g., self-report versus observational measures) to test whether results are significantly different over categories; if they are not different, combining them may still make sense. Fourth, it is nearly always possible to construct higher order constructs to account for diverse measures. Although the highest order constructs, such as "mental health" or even "family therapy outcome" may be quite broadly defined, they still unite the vast majority of measures included in outcome research. In the end, then, meta-analysts tend to treat this as an empirical question, combining results until critics can point to solid statistical or conceptual reasons why it should not be done.

Effectiveness/Efficacy for Marriage and Family Interventions: Meta-Analytic Findings

We searched for meta-analyses of the effects of marriage and family interventions through *PsychINFO* and *Dissertation Abstracts International,* by reviewing the reference sections of previous reviews on these topics, and by doing issue-by-issue hand searches of recent issues of journals that commonly publish such reviews (e.g., *Journal of Consulting and Clinical Psychology, Journal of Marital and Family Therapy,* and *Behavior Therapy*). Table 1 lists the $N = 20$ meta-analyses on the effects of marriage and family interventions that this search found. These meta-analyses are listed in the reference section with an asterisk (*) preceding them. These meta-analyses have been appearing about once a year for the last 15 years. Two meta-analyses are reported in both dissertation and publication form (Dunn, 1994; Dunn & Schwebel, 1995; Giblin, 1985, 1986; Giblin, Sprenkle, & Sheehan, 1985); we combine

Table 1

Meta-Analyses of the Effects of Marriage and Family Interventions

Authors	Topic	Number of Studies
1. Butler & Wampler, 1999 (see also Wampler, 1982)	Couple Communication Enrichment	16
2. Cedar & Levant, 1990	Parent Effectiveness Training	26
3. Dunn & Schwebel, 1995 (see also Dunn, 1994)	Marital Therapy (Behavioral Marital Therapy, Cognitive Behavioral Marital Therapy, Insight Oriented Marital Therapy)	15
4. Dutcher, 2000	Marital Therapy (Behavioral Marital Therapy, Cognitive Behavioral Marital Therapy, Insight Oriented Marital Therapy)	17
5. Edwards & Steinglass, 1995	Family Therapy for Alcoholism	21
6. Giblin, 1985 (see also Giblin et al., 1985; Giblin 1986)	Premarital, Marital, and Family Enrichment	85
7. Hahlweg & Markman, 1988	Behavioral Marital Therapy,	17
	Behavioral Premarital Intervention	7
8. Hazelrigg et al., 1987	Family Therapy	20
9. Hight, 2000	Couple Enrichment	111
10. Johnson et al., 1999	Emotionally Focused Couples Therapy	4
11. Mari & Streiner, 1994	Family Therapy for Schizophrenia	6
12. Markus et al., 1990[a]	Family Therapy	19
13. Montgomery, 1991[b]	Family Therapy for Child Identified Problems	47
14. Pitschel-Walz et al., 2001	Family Therapy for Schizophrenia	25
	Family Interventions vs. Other Interventions	
	Family Interventions vs. Usual Care	
15. Plattor, 1991	Marital Therapy	25
16. Shadish et al., 1993 (see also Shadish, 1992)	Marital and Family Therapy	163
	Treatment-Control Comparisons	
	Treatment-Treatment Comparisons	
17. Stanton & Shadish, 1997	Family Therapy for Drug Abuse	15
18. Sweeney, 1991[b]	Marital and Family Therapy	163
19. Wilson, 1989[b]	Marital Therapy	62
20. Wilson, 1994	Marital and Family Therapy vs. Individual	
	Therapy	44

[a]   Some details of the methods used in this meta-analysis are only available in
       an unpublished foreign language manuscript that was unavailable to us (Markus, 1988).
[b]   Montgomery (1991), Sweeney (1991), and Wilson (1989) report meta-analyses of a subset of the
       Shadish et al. (1993) meta-analysis.

these two forms because both the dissertation and the publication presented the same information. One meta-analysis (Butler & Wampler, 1999) incorporates all the data from a previous meta-analysis by the same author (Wampler, 1982), so these two reports are counted as one meta-analysis. Three more meta-analyses (Montgomery, 1991; Sweeney, 1991; Wilson, 1989) examined subsets of studies in a larger meta-analysis (Shadish et al., 1993); we report these as four separate meta-analyses since they provide mostly different meta-analytic information.
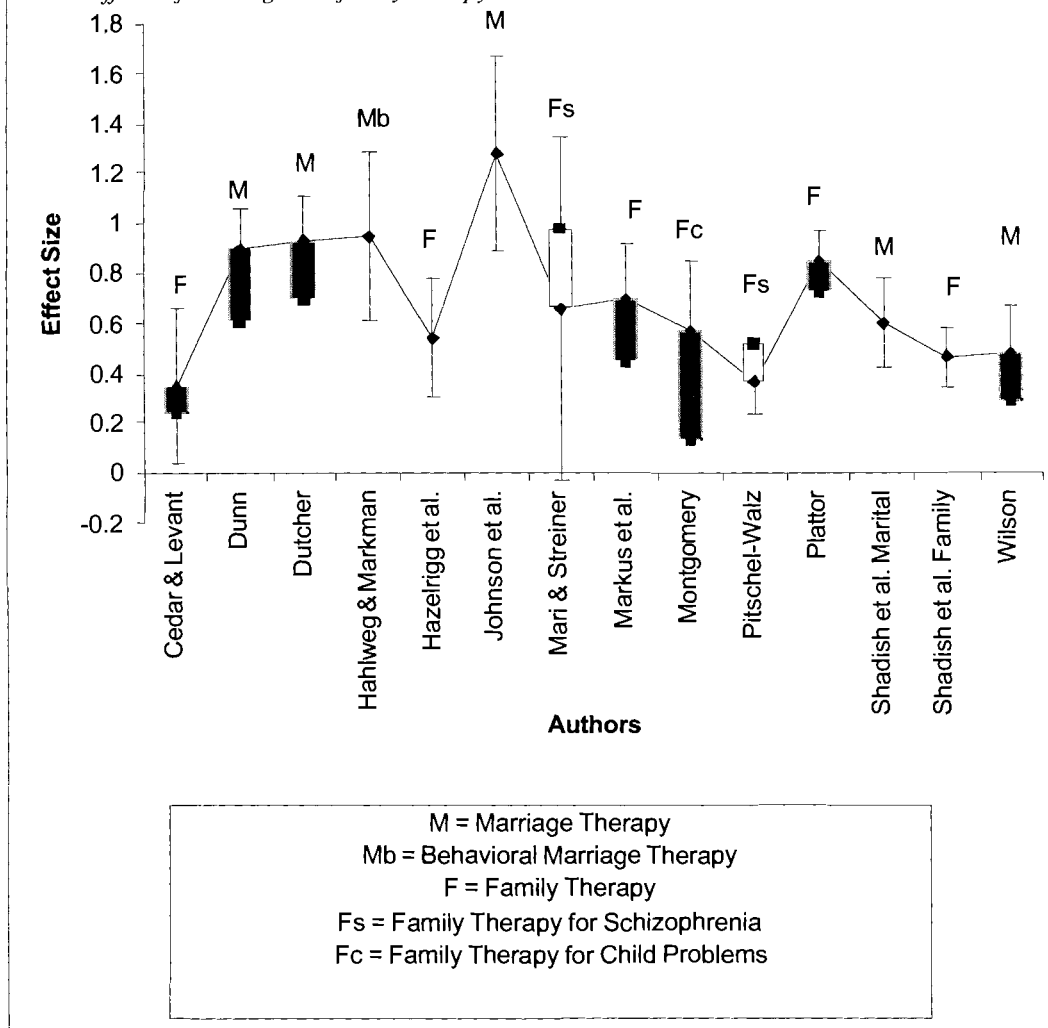
Table 2 summarizes some general characteristics of these 20 meta-analyses. A number of features

Table 2
Characteristics of 20 Meta-Analyses of the Effects of Marriage and Family Interventions

Characteristic	Number of Meta-Analyses
Publication Status of Meta-Analysis	
Published Journal Article	11
Unpublished Thesis or Dissertation	7
Both Published and Unpublished Forms	2
Therapy vs Enrichmenta	
Therapy	17
Enrichment	4
Substantive Focus	
Both Marriage and Family	5
General Marriage and Family Therapy ($n = 3$)	
Marriage and Family Therapy for Drug Abuse ($n = 1$)	
Marriage and Family Enrichment ($n = 1$)	
Marriage Only	8
General Marriage Therapy ($n = 4$)	
Behavioral Marital Therapy ($n = 1$)	
Emotionally Focused Therapy ($n = 1$)	
Marriage Enrichment ($n = 2$)	
Family Only	8
General Family Therapy ($n = 2$)	
Family Therapy for Alcoholism ($n = 1$)	
Family Therapy for Child Presenting Problems ($n = 1$)	
Family Therapy for Schizophrenia ($n= 2$)	
Parent Effectiveness Training ($n = 1$)	
Family Enrichment ($n = 1$)	

a   Hahlweg and Markman (1988) included both therapy and enrichment studies, and so is included in both categories.

of these two tables are worth comment. First, meta-analyses are increasingly popular as graduate student theses and dissertations. More than a third of the meta-analyses we located were student theses and dissertations, most of which never made it into print (Dutcher, 2000; Hight, 2000; Montgomery, 1991; Plattor, 1991; Sweeney, 1991; Wilson, 1989, 1994). So if one wishes to locate all meta-analyses on a topic, clearly one must search for unpublished work. Second, while most of the meta-analyses pertained to therapy for distressed couples, we located four meta-analyses about marital or family enrichment interventions (Butler & Wampler, 1999; Giblin, 1985; Hahlweg & Markman, 1988; Hight, 2000). The concepts of therapy and enrichment seem distinct, but in practice the two concepts sometimes overlap. One enrichment meta-analysis noted, for example, that "studies of clinic couples or families were included when the thrust was greater than symptom removal and processes of enrichment were employed" (Giblin et al., 1985, p. 259). Third, the 20 meta-analyses are spread evenly over both marriage and family therapy taken separately; and a number of more specific meta-analyses have examined the effects of marriage and family therapies on such problems as drug abuse, alcoholism, child problems, and schizophrenia. Fourth, most of the meta-analyses are fairly small, including less than 30 primary studies; but some are quite large, with the largest (Shadish et al., 1993) locating 163 randomized experiments on the topic. Clearly, then, the literature on marriage and family interventions is quite large, probably numbering several hundred experiments by now.
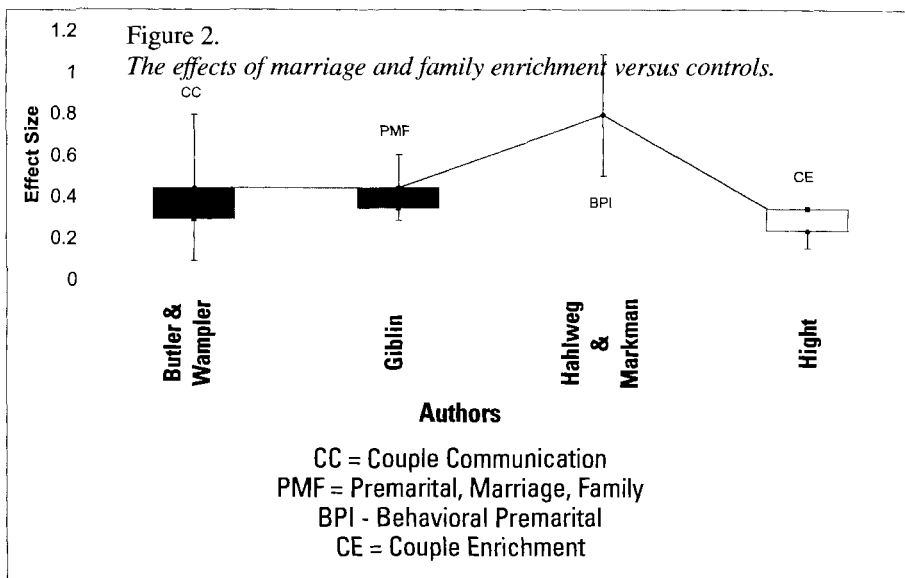
Figure 1.
The effects of marriage and family therapy versus controls.

M = Marriage Therapy
Mb = Behavioral Marriage Therapy
F = Family Therapy
Fs = Family Therapy for Schizophrenia
Fc = Family Therapy for Child Problems

So what have we learned about the effects of marriage and family interventions from these meta-analyses? Frankly, it is very difficult to summarize all these varied meta-analytic results. Results are rarely comparable from meta-analysis to meta-analysis given variations in (a) the questions asked, (b) which primary studies are included and which are excluded, (c) how the same construct is differentially operationalized in different meta-analyses, (d) how effect sizes are computed, (e) how effect sizes are weighted prior to being combined, (f) the kinds of analyses that are then done, and (g) how all these results are reported. Nonetheless, despite all the many good reasons that argue against trying to summarize these very disparate results, we do exactly that in the sections that follow.

Marriage and Family Interventions Versus Controls

Figures 1 and 2 summarize results for those meta-analyses that compared marriage and family interventions to a no treatment or waitlist control. In both figures, error bars approximate 95% confidence intervals;[2] and drop bars indicate the difference between posttest and follow-up effect sizes, with drop bars in black indicating decreases at posttest and drop bars in white indicating increases at

Figure 2.

The effects of marriage and family enrichment versus controls.

CC = Couple Communication
PMF = Premarital, Marriage, Family
BPI - Behavioral Premarital
CE = Couple Enrichment

posttest (we could not estimate follow-up effects for some meta-analyses). For example, the Cedar and Levant (1990) meta-analysis is depicted first in Figure 1, on the left side of the figure. That meta-analysis reported a mean posttest effect size of d = .35, marked by a diamond in Figure 1, and a followup effect size of d .24, marked by a square, with a black drop bar going from the diamond to the square to indicate the size of the decrease. Surrounding the posttest effect size of d .35 is the error bar for this effect (depicted as a vertical line that has horizontal lines at each end). The error bar marks the 95% confidence interval for the Cedar and Levant effect size. That interval is $.036 \leq \Delta \leq .664$, meaning we are 95% confident that the true effect lies in that interval. Since the confidence interval does not include zero, the posttest effect size of d .35 is significantly different from zero at the $\alpha = .05$ level.

Both figures suggest that marriage and family interventions are effective. Figure 1 shows the results for marriage and family therapy. Note several features of that figure. First, the 95% confidence intervals indicated by the error bars exclude $d = 0$ in all but one case (Mari & Streiner, 1994), suggesting that the effects of marriage and family therapy differ significantly from zero. Second, the drop bars show that effects at follow-up are generally smaller than those at posttest, though the difference between posttest and follow-up effects are relatively small. The simple average of the 12 meta-analyses yields $d = .65$ at posttest and $d = .52$ at follow-up. The only two drop bars that show an increase in effects from posttest to follow-up are for family therapy with schizophrenia. Third, note that the average effects for marriage therapy $(d = .84)$ tend to be higher than the effects for family therapy $(d = .58)$. This has been a very persistent finding in the meta-analytic literature, and some evidence suggest that this difference is due to the fact that family therapy is typically applied to less tractable problems than is marriage therapy. In those rare cases in which marriage therapy and family therapy were applied to similar problems, their effects were about the same (Shadish et al., 1993).

The results for marriage and family enrichment are similar (Figure 2), although the effects of these interventions are a bit smaller than for marriage and family therapy, with $d = .48$ at posttest and $d = .32$ at follow-up. Perhaps most noteworthy in Figure 2 is the high effect for behavioral premarital intervention, mirroring Figure 1's finding that the highest average effect size came from behavioral marriage therapy. This invokes, of course, the long-standing debate in psychotherapy meta-analysis about whether behavioral interventions tend to produce larger effects than do other interventions. We will return to this question at the end of this chapter. On the one hand, it is clear that much behavior therapy research is conducted with methods that tend to yield larger effects, especially the use of behavioral dependent measures that systematically yield larger effects. On the other hand, the finding of

Table 3
The Effects of Various Social, Educational, and Medical Treatments

Topic	Effect Size *d*
Electroconvulsive Therapy for Depression	.80
Coronary Bypass Surgery for Angina	.80
Marriage and Family Therapy	.65
Marriage and Family Enrichment	.48
AZT for AIDS Mortality	.47
Dipyridamole Medication for Angina	.39
Neuroleptic Drugs for Dementia	.37
Anticoagulant Therapy for Thromboembolism	.30
Aortocoronary Bypass Surgery on Mortality	.15
Intravenous Streptokinase for Myocardial Infarction Mortality	.08

behavioral superiority has emerged with some consistency across many psychotherapy literatures, and it occasionally withstands efforts to adjust away that superiority with regression models that take such differential methods into account (e.g., Shadish, Matt, Navarro, & Phillips, 2000). So this finding is worth more study.

To put these effect sizes into better perspective, Table 3 shows how they compare to effect sizes from meta-analyses of the effects of various other social, educational, and medical treatments (these examples are drawn from Lipsey & Wilson, 1993). Effect sizes in that table range from a high of $d =$ .80 for the effects of electroconvulsive therapy for depression to a low of $d = .08$ for the effects of intravenous streptokinase to prevent mortality after a myocardial infarction. Of course, the interventions and outcomes in Table 3 are not strictly comparable to the family and marriage interventions and outcomes summarized in this chapter; but Table 3 does give a sense of the range of effect sizes found in a range of interventions, against which the effect sizes for marriage and family interventions fare reasonably well.

Direct Comparisons of Different Kinds of Family and Marriage Interventions

Studies that directly compare two or more different kinds of marriage and family interventions provide particularly compelling evidence about the relative effectiveness of those therapies. Four meta-analyses did these kinds of comparisons, and on the whole, any such differences were small and usually nonsignificant. Butler and Wampler (1999) found that the Couple Communication program performed about the same as similar programs. Pitschel-Walz, Leucht, Bauml, Kissling, and Engel's (2001) review of family therapy for schizophrenia did find some differences among family therapies, including "studies testing brief interventions, which made in general a poorer showing, and studies testing the multifamily format, which seems to be more successful in the long-run than the single-family format" (p. 83). Shadish et al. (1993) found few differences among different theoretical orientations to marriage and family interventions, and those differences disappeared once other covariates were controlled. Wilson (1989) found no significant differences among different kinds of marriage therapies.

Family and Marriage Interventions Versus Alternative Treatments

How well do marriage and family interventions fare against alternative treatments? Figure 3, which includes both therapy and enrichment interventions, shows that at posttest they perform at least as well as, and sometimes better than, alternatives that include individual psychotherapy, placement in halfway houses, problem solving training, bibliotherapy, hospitalization, peer counseling, or group therapy. However, these effect sizes are quite small, tend to get even smaller at follow-up, and are usually nonsignificant.

Figure 3.
The effects of marriage and family therapy versus alternative treatments.
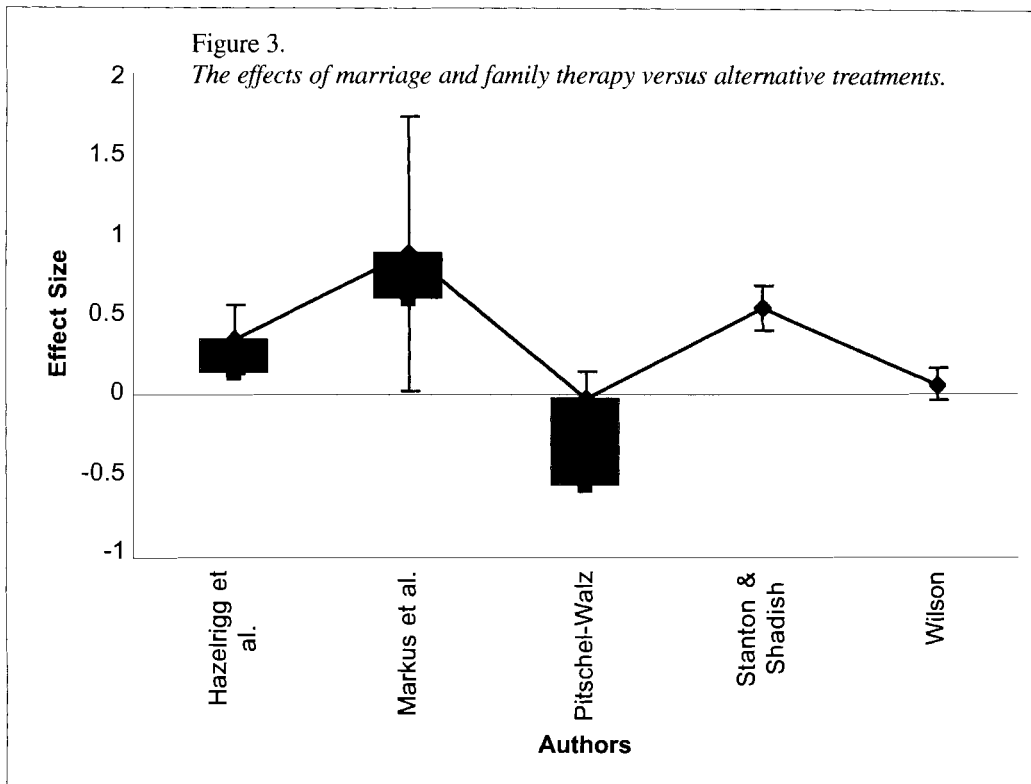
Meta-Analytically Supported Treatments (MASTs)

In recent years, many therapists and researchers have discussed what has come to be known as empirically supported treatments (ESTs). These are clearly specified therapies (i.e., with a treatment manual or equivalent) that are efficacious in controlled research with a well-delineated population (Chambless & Hollon, 1998). The accumulated literature now includes many ESTs, and the associated research has contributed much to our understanding of effective treatments and good outcome research. In addition, ESTs are a valuable resource to the practitioner who has a client with a presenting problem, and is searching for a treatment that is shown to be effective in treating it. This specificity of ESTs is laudable, and to be encouraged.

However, ESTs have a number of problems. First is a problem of construct validity, that the label does not well represent what was actually done in ESTs. What makes ESTs uniquely identifiable is not being empirically supported—a trait they have in common with many therapies that are not ESTs—but rather having clearly specified manuals and clearly delineated populations. A label that more accurately reflects the operations unique to these treatments might be "effective, manualized, population-specific treatments" (EMPS). In addition, the EST label implies boundaries outside of which treatments are not empirically supported, even though meta-analytic reviews of psychotherapy suggest great empirical support for a huge array of therapies that are not on the official EST list.

A second problem is related to the first: that the EST literature misleadingly fosters the impression that other treatments have not been evaluated scientifically. For example, the EST Web site on borderline personality disorders includes a typical statement about ESTs: "While other psychotherapies may be helpful for treatment of borderline personality disorder, they have not been evaluated scientifically in the same way as the treatment listed here."[3] How many people, especially in the target audience of potential psychotherapists or their clients, will understand the subtle nuances of the last few words of that statement: "in the same way as the treatment listed here"? How many will presume that other treatments simply have not been evaluated scientifically in a credible way, despite the fact that 25 years of

accumulated meta-analytic research shows that a huge number of psychotherapies have been scientifically evaluated as effective?

Third, an undue focus on ESTs obscures the fact that both practice and policy value effective treatments even if the studies do not utilize treatment manuals or the like. Examples include the Cochrane Collaboration's[4] meta-analytic summaries of the effects of medical and public health interventions, various summaries by the U.S. Department of Education that identify best practices in education, and the various Evidence-Based Practice Centers funded by the U.S. Government to do meta-analytic summaries of effective medical treatments. None of these require the treatment or population specificity of ESTs, yet all have had great impact on both policy and practice.

A fourth problem with the EST literature is statistical power. Any given study may fail to show statistically significant results by, for example, having too few participants or less reliable measures. Current incarnations of EST criteria require two studies of a proposed EST with statistically significant results. Many proposed ESTs fail this criterion not because they lack two studies, but because results were not statistically significant. Yet a meta-analytic summary of two studies may yield an effect size that is significantly different from zero even when each study by itself is nonsignificant. EST methodology could and should shift from an emphasis on significance testing to use of effect sizes and confidence intervals, a shift that would be consistent with other trends in statistics (Wilkinson & the Task Force on Statistical Inference, 1999).

To provide a complement to empirically supported treatments, therefore, we propose the concept of *meta-analytically supported treatments*, or MASTs. MASTs have four characteristics: (a) effect sizes from more than one study of the treatment construct have been combined meta-analytically, (b) all of the studies contributing to that synthesis are randomized trials comparing the treatment to a control group, (c) the meta-analysis reported an effect size and significance test showing that the treatment produced pooled meta-analytic effects that are significantly larger than expected by chance, and (d) the meta-analysis used sound meta-analytic statistical methods such as aggregating effect sizes to the study level. Table 4 presents a preliminary list of the treatments that qualify as MASTs according to these criteria among the 20 meta-analyses of the effects of marriage and family interventions. The list contains 24 MASTs, including broad treatment constructs such as family therapy in general, and more narrow treatment constructs such as behavioral family therapy for child identified problems.

Not surprisingly, this list of MASTs and the list of ESTs that appears on the American Psychological Association Division 12 Web site[5] overlap in some cases, including family therapy for schizophrenia, family therapy for child conduct disorders, and behavioral marital therapy for marital distress. On the other hand, some MASTs are not on the EST list. For example, on the EST site for drug and alcohol abuse,[6] family therapy is not mentioned as an effective treatment; and the EST site for marital distress[7] omits systemic and eclectic marital therapies, both of which have clear empirical support in meta-analytic research.

We do not mean to minimize the importance of the two key operational differences between MASTs and ESTs; namely, that ESTs are described in written treatment manuals and use clients who have clearly delineated problems. These two differences allow therapists to make particular treatment recommendations to particular types of clients. On the other hand, such specificity can be misleading. Rosen and Davison (in press) pointed out that there are certain empirically supported principles, such as the use of exposure in the treatment of phobias, that are so well-established and superior to other treatment principles that to not use them might be unethical no matter whether they are represented on the list of ESTs. Meta-analysis would be an ideal methodology for demonstrating the effectiveness of such principles, because by its very nature meta-analysis reaches conclusions about principles that are presumed to be common to diverse treatment operations. Moreover, the specificity of ESTs tends to hide the empirical finding in meta-analysis that the effects of psychotherapy tend to generalize robustly across variations in presenting problem, over whether or not manuals are used, and indeed, over a very large number of other potential moderators.

We do not mean that practitioners can use any therapy with any problem. To the contrary, when ESTs or other demonstrably effective treatments or principles are available to treat a specific problem,

Table 4

Meta-Analytically Supported Treatments (MASTs) for Marriage and Family Interventions

Treatment Constructs	Supporting Meta-Analyses
Family Interventions	
Family Behavioral Therapy for Alcoholism	Edwards & Steinglass (1995)
Family Therapy for Schizophrenia[a]	Mari & Streiner (1994); Pitschel-Walz et al. (2001)
Family Therapy for Drug Abuse	Stanton & Shadish (1997)
Family Therapy for Child Identified Problems[a]	Montgomery (1991)
• Behavioral/Psychoeducational/ Problem Solving	
• Systemic	
• Humanistic	
• Eclectic	
Family Therapy: Behavioral	Shadish et al. (1993)
Family Therapy: Systemic	Shadish et al. (1993)
Family Therapy: Eclectic	Shadish et al. (1993)
Family Therapy: General	Shadish et al. (1993)
Marital Interventions	
Marital Therapy: Behavioral[a]	Dunn (1994); Dunn & Schwebel (1995); Dutcher (2000); Shadish et al. (1993); Wilson (1989)
Marital Therapy: Cognitive Behavioral	Dunn (1994); Dunn & Schwebel (1995); Dutcher (2000)
Marital Therapy: Systemic	Shadish et al. (1993)
Marital Therapy: Eclectic	Shadish et al. (1993); Wilson (1989)
Marital Therapy: Emotionally Focused Therapy[a]	Johnson et al. (1999)
Marital Therapy	Plattor (1991)
• Contracting	
• Communication Training	
• Contracting plus Communication Training	
• Insight Oriented[a]	
Marital Therapy: General	Dunn (1994); Dunn & Schwebel (1995); Dutcher (2000); Plattor (1991); Shadish et al. (1993)
Marital Enrichment Interventions	Hight (2000)

a   These treatments also appear on the list of empirically supported therapies at http://www.apa.org/divisions/div12/rev_est/index.shtml, as of June, 2002

there is much to be said for using them. But lacking such specific information, it is still sensible and ethical to refer clients to therapies that seem to work in general. After all, although the general prescription to do marital or family interventions is far less specific than an EST manual, it still points to a set of practices that are well differentiated from other interventions such as individual psychotherapy, consulting with clergy, or vacationing in the Florida Keys. The construct of marriage and

family therapy also appeals to a set of practices in which clinicians have experience or training, or can locate further sources of training such as books and workshops.

In view of these points, practicing clinicians might use ESTs and MASTs in the following complementary way. When a practitioner has a client with a particular problem, they might choose an EST if one exists to treat that problem and if it seems contextually appropriate. If no EST for that problem has been approved, the clinician might review lists of MASTs for therapies that have been shown effective with that problem, or as a last resort, for MASTs that have general support even if not for that specific problem.

We suspect that meta-analytic searches for further ESTs and MASTs would prove fruitful in two respects. One is that meta-analysts could routinely code studies for EST characteristics but combine them meta-analytically rather than relying on the individual significance tests reported in single studies. This should increase the number of ESTs available to clinical practice. The second would be to search for treatments that are shown to affect the outcome at issue in a particular presenting problem. For example, a MAST might exist that reduces childhood depression, even though the clients in the studies might not meet the criteria for depression that ESTs require. If so, it would be reasonable to use that MAST for the depressed child in lieu of the absent EST.

Table 4 is just a first effort at identifying MASTs, and then just for family and marriage interventions. For example, we would like to recode the studies contributing to MASTs from scratch rather than relying on coding done by other researchers, to make sure the codes are accurate and the effect sizes were computed correctly. Even so, this list of MASTs makes the general point that marriage and family interventions are empirically supported treatments, even if they are not always on someone's list with that name.

Clinical Significance

A criticism of meta-analysis has been that effect sizes do not convey much information about clinical significance. A number of translations of effect size into other metrics have partially assessed this, such as comparing effect sizes to Cohen's norms of small, medium, or large effect sizes (Butler & Wampler, 1999; Markus, Lange, & Pettigrew, 1990), using Rosenthal and Rubin's (1982) binomial effect size display to convert effect sizes into a fourfold table that estimates the percent of successes and failures in both treatment and comparison groups (Pitschel-Walz et al., 2001), and conversions of effect sizes back into the metrics of original scales like the Beck Depression Inventory. However, three meta-analyses used more interesting methods to examine clinical significance.

One approach is illustrated by Edwards and Steinglass (1995), who examined the clinical significance of the effects of family therapy on alcoholism. Their approach took advantage of the fact that these studies report the same outcome measure: abstinence rates among alcoholics. Spontaneous abstinence rates are less than 5% among untreated alcoholics, so Edwards and Steinglass set 50% abstinence as a clinically significant effect for family therapy with alcoholics. By this criterion, 83% of the family or marriage therapy interventions yielded clinically significant results, compared to 50% of the nonfamily therapy treatment alternatives.

Hahlweg and Markman (1988) built on the fact that the Marital Adjustment Scale (MAS) or the Dyadic Adjustment Scale (DAS) are commonly used in many outcome studies of marital therapy. Scale norms suggest that a score of 100 on the MAS and a score of 97 on the DAS are considered indicative of good marital adjustment. Therefore, Hahlweg and Markman computed effect sizes by substituting those norms in the d statistic, so that a clinically significant effect size would then be $d > 0$. For nine studies that provided enough data to do this calculation, $d = 0.15$ ($SD = .41$). That is, the average treatment group mean achieved an MAS or DAS score that was higher than the norms for good marital adjustment, although clearly many clients were still distressed. If we translate this into percent success using Rosenthal and Rubin's (1982) binomial effect size display, we find that this translates into a clinically significant success rate of 54%.

Shadish et al. (1993) used a third method. Their database contained 19 studies that reported

treatment and control group means on the MAS or the DAS at both pretest and posttest. They then computed the percent of studies that moved groups from pretest means below the norms for good marital adjustment to posttest means above those norms. Shadish et al. found that none of the control groups produced clinically significant results by this criterion, compared to 41% of the treatment groups. The latter figure is comparable to the 35.5% clinical significance rate reported by Jacobson et al. (1984) in their original study.

In all, then, marriage and family therapies produce clinically significant improvements in distressed clients, with success rates of 40 to 50%. Further, these three methodologies can be used in any literature in which the same outcome measure is used across many of the studies in the meta-analysis, and where that scale either has a cutoff for clinical significance or such a cutoff can be plausibly suggested. Further development of such methodologies would be a very welcomed advance in meta-analytic work.

Clinical Representativeness

Recent reviews of the psychotherapy outcome research literature show that most studies have been done in ways that do not well represent the conditions of actual clinic practice (e.g., Shadish et al., 2000; Weisz, Donenberg, Han, & Weiss, 1995; Weisz, Weiss, & Donenberg, 1992); for example, using clients referred through usual routes and experienced therapists in actual clinic settings. The issue is important because some meta-analyses in child psychotherapy suggest that such therapy is completely ineffective in clinically representative conditions. Further, in part appealing to the efficacy-effectiveness distinction from public health, federal agencies have become more interested in funding research that is conducted under clinically representative conditions in order to understand whether therapy that is efficacious under ideal conditions is also effective under conditions of actual practice.

None of the marriage and family therapy meta-analyses have directly attended to this issue. Some have investigated individual moderators that would be part of clinical representativeness; for example, comparing therapy in university versus nonuniversity settings (Montgomery, 1991; Stanton & Shadish, 1997; Wilson, 1989), therapy that is manualized or not (Montgomery, 1991; Shadish et al., 1993; Stanton & Shadish, 1997; Wilson, 1989), referral source of clients (Hahlweg & Markman, 1988; Wilson, 1989), or using community-based therapists (Stanton & Shadish, 1997). But the question of whether marriage and family therapy works in clinically representativeness conditions can only be answered by finding studies that combine all these clinically representative characteristics in one study. Future marriage and family meta-analyses would be better if they incorporated codes about clinical representativeness from previous meta-analyses in other areas of psychotherapy (e.g., useful codes for this can be found in Shadish et al., 2000, Appendix 1).

Moderators of Treatment Effects

Historically, meta-analysis has tended to focus on main effects, on whether or not a treatment works. Meta-analysis has been less helpful in answering the famous question posited by Gordon Paul (1967): "What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?" (p. 111). This is a hard enough question to answer in primary studies. Answering this question with meta-analysis requires examining moderators of treatment effects; that is, those variables that interact with treatment by varying the size or direction of that effect.

In general, meta-analysts have not studied moderators extensively. A graduate student in our laboratory at The University of Memphis has looked at this issue for 52 psychotherapy meta-analyses (Phillips, 1998; Phillips & Shadish, 2002). He found that the modal number of moderators examined in psychotherapy meta-analyses is one, and the median is five. For the 20 marriage and family meta-analyses that we reviewed in the present case, the situation is similar, with most meta-analyses examining fewer than five potential moderators of therapy effectiveness. It turns out to be surprisingly difficult to summarize those moderator effects because different meta-analyses use different labels for the same moderator construct, different constructs under the same label, different operationalizations of the same construct, and different ways of reporting all this. Even so, searches for such moderators are worth pursuing. Consider two examples.

First, a common finding is that the kind of measurement used in primary studies makes a large difference to effect size. For example, five meta-analyses found that behavioral measures produced larger effect sizes than other measures (Butler & Wampler, 1999; Giblin, 1985; Hahlweg & Markman, 1988; Hight, 2000; Shadish et al., 1993). Another five meta-analyses found that effect sizes were higher when outcome measures were specifically tied to what was done in therapy (Cedar & Levant, 1990; Hight, 2000; Montgomery, 1991; Shadish et al., 1993; Wilson, 1989). This measurement specificity effect has been found in psychotherapy meta-analyses in general (Matt, 1993).

Second, Sweeney (1991) focused entirely on interactions in her master's thesis on the effects of marriage and family therapies. Table 5 presents one of her results (Shadish & Sweeney, 1991) pertaining to the frequent finding that behavior therapies have higher effect sizes than nonbehavior therapies. The question is whether this finding reflects better patient outcomes, or whether it reflects artifactual confounds. Sweeney found that behavioral marriage and family therapy studies have unusually high effect sizes when they are conducted in university settings with measures that are reactive, specific, and manipulable, and have relatively few participants. Behavior therapies without these characteristics have effect sizes that are no better and no worse than nonbehavior therapies. That is, the larger effect sizes produced by behavior therapies may not reflect better patient outcomes, but rather may reflect artifactual differences in how studies are conducted.

Both of these examples suggest regularities that may underlie diverse meta-analytic results; however, regularities may not be salient because we have approached meta-analysis as a brute force empirical tool for finding out what works. We need better conceptual schemes and theory to clarify the conditions under which moderator effects occur. For example, the measurement specificity effect goes under many different names. One author calls it specificity, another calls it congruence, and yet another does not use any label at all. It takes experience with meta-analysis to see the common concept underneath the diverse labels. Too few researchers approach meta-analysis from this conceptual perspective.

Further, both these moderator findings have implications for how we should interpret the results of

Table 5
Variables That Moderated the Effects of Theoretical Orientation

Moderator	Behavioral d	Nonbehavioral d
Setting		
University	.73	.36
Nonuniversity	.35	.36
Measurement Reactivity		
High	.68	.58
Medium	.48	.39
Low	.07	.48
Measurement Specificity		
Treatment Specific	.72	.46
General Family/Marital Measure	.50	.44
General Measure	.13	.58
Measurement Manipulability		
Not Very	.15	.76
Moderately	.58	.55
Very	.55	.46
Number of Subjects		
Below Median	.77	.39
Above Median	.45	.48

outcome research. For example, a common experience is to read a report of an experiment in which a treatment produces, say, a particularly large effect relative to other therapies. It is tempting to conclude that the *treatment* is more effective than other treatments. Yet that may be the wrong conclusion to draw. The large effect may simply reflect *how the outcome was measured.* If all therapies were measured with the same kinds of outcomes assessments, their results might be considerably more uniform—indeed, this is found to be the case routinely when such moderators are used in regression equations to adjust for apparent treatment orientation differences (e.g., Shadish et al., 1993). In this regard, it would greatly improve our capacity to interpret research if all studies were to adopt at least some uniform measures across experiments. For example, each marital study might include the MAS or DAS as a more general measure, and also a more specific measure tied closely to what was done in therapy.

Summary for Clinicians

This section has reviewed a great deal of material, so Table 6 summarizes some of the more important points for clinical practice.

METHODOLOGICAL ISSUES

This section has two purposes. The first is to highlight some of the key methodological issues in the conduct of meta-analysis that warrant improvement in future meta-analyses of marriage and family interventions. The second is to help clinicians know how to read and criticize the quality of the meta-analyses they may be reading. For those less interested in the details of these methodological issues, we summarize the key points in a table at the end of this section.

Study Design

Following the lead of Smith and Glass (1977), psychotherapy meta-analyses have typically included primary outcome studies with many different kinds of designs. The justification was that the

---

Table 6
Summary for Clinicians

1. Marriage and family interventions, both therapy and enrichment, are more effective than no treatment. Those effects tend to be maintained at followup.

2. Marriage therapy tends to have better outcomes than family therapy, but this seems to occur because family therapists often deal with more difficult problems (e.g., schizophrenia).

3. Different kinds of marriage and family interventions tend to produce similar results.

4. The effects of marriage and family interventions are comparable to or larger than those obtained by alternative interventions ranging from individual therapy to medical interventions.

5. Meta-analytically supported treatments (MASTs) exist that have strong empirical support. While marriage and family therapists should be encouraged to use those therapies called empirically supported treatments (ESTs) when ESTs are available, it makes excellent scientific and professional sense to use MASTs when ESTs are not available for a problem, as will often be the case.

6. Marriage and family therapies produce clinically significant results in 40%–50% of those treated.

7. The effects of marriage and family interventions in clinically representative conditions have not been studied much.

8. We do not know much about variables that moderate the effects of marriage and family interventions, although available evidence suggests that *how* the research is done has as strong an impact on outcome as *what kind* of treatments are used.

question of whether or not design makes a difference to effect size should be an empirical question, and that mixing designs is not a problem if one can show that different designs yield similar results. Among the 20 marriage and family meta-analyses reviewed here, only 9 limited their sample to randomized experiments (Johnson, Hunsley, Greenberg, & Schindler, 1999; Mari & Streiner, 1994; Montgomery, 1991; Pitschel-Walz et al., 2001; Plattor, 1991; Shadish et al., 1993; Stanton & Shadish, 1997; Sweeney, 1991; Wilson, 1989). The remaining 11 meta-analyses were split into two groups.

Six meta-analyses included both randomized and nonrandomized experiments (Cedar & Levant, 1990; Edwards & Steinglass, 1995; Hahlweg & Markman, 1988; Hazelrigg, Cooper, & Borduin, 1987; Hight, 2000; Wilson, 1994[8]); but only four of them tested for differences between the two methodologies (Giblin, 1985; Hazelrigg et al., 1987; Hight, 2000; Wilson, 1994). All four found meaningful differences; yet in three of these cases, the authors combined results despite those differences (Giblin, 1985; Hazelrigg et al., 1987; Hight, 2000). This practice is questionable. Given other evidence that psychotherapy quasi-experiments underestimate effects (Shadish & Ragsdale, 1996; Shadish et al., 2000), meta-analysts should test for design differences, and be very cautious about combining estimates from studies of different design if those estimates appear statistically or practically different.

Finally, five meta-analyses included not just randomized and nonrandomized experiments, but also pretest-posttest only designs with no independent control group (Butler & Wampler, 1999; Dunn, 1994[9]; Dutcher, 2000[10]; Giblin et al., 1985; Markus et al., 1990). Evidence suggests that the latter designs overestimate effect sizes substantially (e.g., Lipsey & Wilson, 1993). A technology for combining uncontrolled studies with results from controlled studies exists (Li & Begg, 1994), as does a technology for combining results from uncontrolled studies with each other (Shadish, Cook, & Campbell, 2002, p. 426, footnote 3). None of the present meta-analyses used those technologies. Consequently, these five meta-analyses may have substantially overestimated effects.

Study Quality Scales

Five meta-analyses (e.g., Butler & Wampler, 1999; Cedar & Levant, 1990; Hight, 2000; Mari & Streiner, 1994; Stanton & Shadish, 1997) used multi-item scales to see if methodological quality was related to effect size. Four of the five tested the relationship between quality and effect size. Three of the four found that effect sizes increased in studies with higher quality ratings (Butler & Wampler, 1999; Cedar & Levant, 1990; Hight, 2000), but the fourth found the opposite effect (Stanton & Shadish, 1997).

While it is worth exploring such scales, we have several reservations about the routine use of any scale that results in a single number to represent study quality. Scales that result in a single number representing design quality tend to combine multiple items into a total score that subsumes a very wide array of methodological variables related to quality. These can include design, sample size, measurement reliability, and representativeness of participants, to name just a few research characteristics. Clearly, these diverse items do not assess the same kind of "quality." Rather, some assess internal validity, some statistical conclusion validity, some construct validity, and some external validity. Thus, global quality assessments can assign identical total scores to studies with decidedly different validity characteristics. Lumping such diversity together may yield a total score of questionable utility or validity. Further, such scales are rarely if ever well developed psychometrically, with no evidence of factorial structure to justify scoring, or even evidence of simple internal consistency reliability. A recent empirical study found that the conclusion reached about study quality, and the conclusion reached by the review itself, can differ considerably depending on which quality scale is used (Jüni, Witschi, Bloch, & Egger, 1999). Global decisions about what makes a study good or bad are fraught with difficulty. Even the most sophisticated researchers can disagree about the dimensions that define quality and how these dimensions apply to each study.

In general, then, items measuring study quality should be scored individually rather than being summed into a total score. We also encourage further development of *multi-dimensional* quality scales

as an important part of reviews. However, researchers who develop such scales should (a) clarify the kind of quality they believe the scale and subscales assess; (b) provide psychometric data about the reliability, factors structure, and validity of the scales where multiple items scales are used; and (c) explore how conclusions about the relationship between quality and outcome would vary depending on whether the total score or the individual items were used.

Attrition

In randomized experiments, it is rare for outcome measurements to be obtained on all clients. Such attrition can seriously compromise the internal validity of inferences from randomized studies. Those few meta-analyses that looked at the question found that attrition was related to effect size (Cedar & Levant, 1990; Montgomery, 1991; Shadish et al., 1993; Wilson, 1994). In fact, one meta-analysis that looked at this question in detail found that family therapy was more successful at retaining clients than were other forms of therapy, especially at retaining more clients with the worst prognosis (Stanton & Shadish, 1997). This made family therapy outcomes look worse than comparison outcomes because comparison treatments were losing more of the bad prognosis clients.

In at least some cases, meta-analysts can estimate the effects of treatments taking attrition into account. For example, when authors of primary studies reported results on both treatment completers and an intent-to-treat analysis, Pitschel-Walz et al. (2001) always used the latter to get effect size estimates that took attrition into account. Taking this one step further, Mari and Streiner (1994) analyzed their data both as the authors reported it, and as an intent-to-treat analysis in which dropouts were treated as failures. They were able to do this because they limited their meta-analysis to relapse rates. Because relapse is dichotomous at the patient level, missing data can easily be rescored as a failure and the overall results recomputed. Mari and Streiner found that family therapy was effective when analyzed as the authors reported it, but not effective when dropouts were rescored as failures and included in the analysis.

Another meta-analysis (Stanton & Shadish, 1997) that included dichotomous outcomes in drug abuse also assumed that dropouts were failures, and estimated effects under that assumption. For continuous outcomes, Stanton and Shadish also wrote authors to obtain raw data, and then assumed that drug abusers who dropped out had the same drug use status as their pretest status. The pretest scores were then substituted for missing data. The meta-analysts then developed a general method for estimating the likely bounds of effects on dichotomous outcomes by making different assumptions about what happened to those who dropped out, including a user-friendly PC program for implementing those analyses (Shadish, Hu, Glaser, Kownacki, & Wong, 1998).[11] By implementing all of these methods, those authors found that the effects of family therapy with drug abusers were likely to be robust no matter what happened to those who dropped out of therapy. Given the prevalence of attrition, more attention to attrition analyses would greatly improve our confidence in meta-analytic results.

Publication Bias

Thirteen meta-analyses included both published and unpublished material (Butler & Wampler, 1999; Cedar & Levant, 1990; Dunn & Schwebel, 1995; Dutcher, 2000; Edwards & Steinglass, 1995; Giblin, 1985; Hight, 2000; Johnson et al., 1999; Plattor, 1991; Shadish et al., 1993; Stanton & Shadish, 1997; Wilson, 1989, 1994). Some other meta-analysts searched for unpublished material (e.g., Markus et al., 1990) but did not find any that met their selection criteria. The standard publication bias was found in the majority of meta-analyses that looked at the question (Cedar & Levant, 1990; Giblin, 1985; Hight, 2000; Montgomery, 1991; Shadish et al., 1993; Wilson, 1989); that is, published works have higher effects than unpublished works. Only Butler and Wampler (1999) found no difference between published and unpublished works. In general, therefore, those meta-analyses that excluded unpublished work may have overestimated the size of the effect (e.g., Hahlweg & Markman, 1988; Hazelrigg et al., 1987; Mari & Streiner, 1994). To ameliorate the latter problem, we recommend a search for unpublished dissertations in every meta-analysis. Many of them can be obtained at no charge using inter-library loan, and can help the researcher to assess whether publication bias is present.

Methods for Testing Posttest Versus Follow-up Effect Sizes

A number of meta-analyses reported separate posttest and follow-up effect sizes or tested for differences between them (e.g., Cedar & Levant, 1990; Hight, 2000; Mari & Streiner, 1994; Markus et al., 1990), and often observed a decrease in effects from posttest to follow-up. Very few of these meta-analyses, however, limited the test to studies that had both posttest and follow-up measures (Hight, 2000), and fewer still then used an appropriate multivariate procedure to perform that within-study test (Montgomery, 1991; Shadish et al., 1993; Stanton & Shadish, 1997; Wilson, 1989). Without this limitation, any observed differences between posttest and follow-up may be confounded with unknown between-study differences. The more appropriate analyses nearly always show no difference between posttest and follow-up, although the power of the test is compromised by the small number of primary studies with both posttest and follow-ups.

Effect Size Computation

The foundational statistic in a meta-analysis is the effect size. If it is estimated inaccurately, the accuracy of the meta-analysis is in doubt. Ironically, few of the meta-analyses reviewed in this chapter describe how effect size was computed, especially in cases where the usual means, standard deviations, and sample sizes were not available. For example, authors often say only that a criterion for including a primary study in the meta-analysis is that it provides sufficient information to enable effect size calculations (e.g., Giblin et al., 1985; Hahlweg & Markman, 1988; Hazelrigg et al., 1987; Hight, 2000; Markus et al., 1990). What this means, however, depends on how much the author knows about effect size calculations. Authors may exclude large numbers of pertinent studies for which good effect sizes could have been computed had the authors only known the appropriate method. Similarly, when the outcome is dichotomous, both *d* and *r* usually underestimate effects considerably, and a statistic like the odds ratio is often preferable (Haddock et al., 1998). Yet few authors state exactly how they treat dichotomous data. Finally, some authors use gain score means rather than posttest means in effect size calculations, either to adjust for selection bias at pretest or because that is all that the author of the primary study reported (e.g., Cedar & Levant, 1990; Giblin, 1985). However, using gain scores without adjusting for the pretest-posttest correlation results in an effect size that is not on the same metric as those computed only on posttest means, and that can be as much as 300% different from the posttest metric (Zhang, 1996). Effect sizes can be calculated in a very large number of ways, many computational errors can occur, and sophisticated judgments are often necessary to decide exactly which effect size index is best given available data. So researchers will benefit from computer programs that not only help clarify these issues but also automate the process to reduce error (e.g., Johnson, 1993; Shadish et al., 1999).

Miscellaneous Methodological and Statistical Issues

The 20 meta-analyses reviewed in this chapter performed unevenly on a host of methodological and statistical issues in meta-analytic work. First, one quarter of the meta-analyses treated effect sizes within studies as the unit of analysis, failing to aggregate those effect sizes to the study level (Butler & Wampler, 1999; Cedar & Levant, 1990; Edwards & Steinglass, 1995; Giblin, 1985; Plattor, 1991). This allows studies with more outcome measures to unduly influence meta-analytic results.[12] Second, a quarter of the meta-analyses did not report weighting effect sizes by a function of sample size (Butler & Wampler, 1999; Cedar & Levant, 1990; Edwards & Steinglass, 1995; Giblin, 1985; Hahlweg & Markman, 1988), a practice that is standard in meta-analysis. Third, only one meta-analysis used random rather than fixed effects models (Hight, 2000). The difference between fixed and random effects models is in the inferences they allow (Shadish et al., 2000). Fixed effects models infer confidence in conclusions that would be reached if the set of studies being meta-analyzed were to be repeated identically except with new participants. Random effects models infer confidence in conclusions that would be reached if these studies were repeated in ways that varied not just participants but also any

other feature of the study such as the length of treatment, the kinds of outcomes measured, or the clinical representativeness of conditions. Since we are usually interested in generalizations of the latter kind, random effects models have become the accepted state-of-the-art for most meta-analyses. Fourth, when most meta-analyses tested for the effects of moderators, they did so using univariate categorical tests. Only a few (e.g., Hight, 2000; Shadish et al., 1993) used regression equations to take into account redundancy among predictors. In part, this lacuna is due to the small number of studies in some meta-analyses, making regression statistically impractical. But in many other cases regression could and should have been done.

Methodological Recommendations

The criticisms in the previous sections suggest the kinds of practices that meta-analysts ought to be using in their work. To make these practices explicit, Table 7 lists a set of methodological recommendations for meta-analysts to follow.

THE COSTS OF META-ANALYSIS

Remember Moses Herzog's quote at the start of this chapter about the need for a good five-cent synthesis? What are the costs of meta-analysis? Olkin and Allen (1999) estimated that the number of hours it takes to produce a quality meta-analysis in medicine varies with the number of citations that the initial literature search discovers. In 37 meta-analyses that were conducted by a private company specializing in meta-analysis, the mean total number of hours was 1,139 (median = 1,110), with a wide range from 216 to 2,518 hours. They developed a prediction equation from this database of: Total time = $721 + 0.243x - 0.0000123x^2$, where x is the number of citations before applying exclusion criteria. Of course, this equation might not accurately predict the amount of time it would take to do a meta-analysis by someone who did not specialize in the task, or who did not have the same support systems that this private company might have had. Still, it is likely to provide a reasonable lower bound to the number of hours it might take to do a meta-analysis.

The costs of meta-analysis are substantially a function of these hours. Assuming, for example, that

---

Table 7
Recommendations for Doing Better Meta-Analyses

1. Test for differences between randomized and nonrandomized studies, and if they do differ, do not combine results in subsequent analyses.

2. Do not rely on scales that purport to measure study quality with a single number, but do include individual items that measure quality and test for their relationship to effect size.

3. Test for the effects of attrition on effect size, and if possible, explore the effects of different assumptions about attrition on the overall meta-analytic results.

4. Include both published and unpublished studies.

5. If the number of studies permits, test differences between posttest and follow-up effect sizes only from studies that include both time points.

6. Be careful to use the most up-to-date methods for computing effect sizes.

7. Aggregate effect sizes to the study level before computing average effects.

8. Weight effect sizes by a function of sample size when computing average effects.

9. Use random effects models in most cases to do the analyses.

10. If the number of studies permits, use regression to take into account redundancy among predictors of effect size.

---

graduate student time is reimbursed at $10–20 per hour, costs might range from $2,000 for a very small meta-analysis to over $50,000 for a large one. If we use professionals who are paid more, costs will rise proportionately. To this we must add the material costs of obtaining and copying studies, printing coding manuals, buying computers and programs if needed, and paying administrative overhead. In response to an email to a meta-analytic listserve in which we asked about total costs estimates, we received two responses from researchers in organizations doing meta-analyses professionally, giving estimates of $115,000 in one case, and of $250,000 in the second. However, the author of the second email pointed out that one might distinguish between the costs of a full review that included a meta-analysis versus the costs of only adding meta-analytic methods to that review. After all, even a narrative review has to do many of the same tasks as a meta-analytic review, including searching for and obtaining studies, reading them, writing results, and paying administrative overhead. The author of the second email said that the incremental costs of adding meta-analytic methods to a narrative review was only about $15,000 or so.

Who reimburses these costs? We contacted the authors of most of the meta-analyses in this chapter to ask about funding sources. Only 4 of the 20 marriage and family therapy meta-analyses reported having grant funding (Hahlweg & Markman, 1988; Montgomery, 1991; Shadish et al., 1993; Wilson, 1989). In the latter three cases, the same grants supported the work, providing total costs of $144,000 from 1986–1990; and the principal investigator's Psychology Department provided considerable extra support in the form of copying, computer access, secretarial help, research assistant time, and the like. The grants came from the National Institutes of Mental Health and from the Russell Sage Foundation. The latter used to have a program on meta-analysis, but it no longer does. The remainder of the meta-analyses were either supported as part of the research mission of a university, or used student theses and dissertations as a vehicle for meta-analytic work. However, we know of meta-analyses in other areas that also were funded from federal sources, especially research labs that specialize in meta-analytic work; and informal conversations with representatives of many federal agencies suggest that they welcome meta-analytic proposals.

In summary, then, meta-analytic work is not cheap. In current dollars, the costs of a good synthesis are likely to range from $10,000 to $250,000. As Herzog said, we still need a good five-cent synthesis.

CONCLUSIONS

The present review differs from the one we did in 1995 (Shadish et al., 1995) in that the 1995 review reported detailed results from a single meta-analysis of the effects of marriage and family therapies, while the present study reviewed 20 meta-analyses. Nonetheless, the overall conclusions about the marriage and family research remain about the same. First, marriage and family interventions are clearly efficacious compared to no treatment. Second, those interventions are at least as efficacious as other modalities such as individual therapy, and perhaps more effective in at least some cases. Third, there is little evidence for differential efficacy among the various approaches to marriage and family interventions, particularly if mediating and moderating variables are controlled. Fourth, evidence that marriage and family interventions are effective in clinically representative conditions remains sparse, although there are some clear exceptions to that.

On the other hand, meta-analytic reviews of marriage and family therapy have provided some new and interesting directions. One is the increased evidence for the clinical significance of the effects of marriage and family interventions, shown across several quite different methodologies that all converge on about the same answer. Another is the increased attention to variables that moderate the effects of marriage and family interventions, although this area is woefully underdeveloped at a conceptual level. And finally, meta-analyses of marriage and family interventions are beginning to use more sophisticated meta-analytic methods, although again this progress has been uneven.

For those interested in pursuing meta-analytic work, Lipsey and Wilson (2000) provide a particularly useful introduction to practical meta-analysis. For those interested in seeing how meta-analysis can influence substantive debates about psychotherapy, we recommend Smith and Glass's (1977) original psychotherapy meta-analysis, Wampold's (2001) application of meta-analysis to the question of whether all psychotherapies have about the same effect, and Shadish et al.'s (2000) meta-analysis of

whether psychotherapy works under clinically representative conditions.

In summary, then, the work we reviewed in this chapter shows that marriage and family therapy is now an empirically supported therapy in the plain English sense of the phrase—it clearly works, both in general and for a variety of specific problems. Yet there is much more that can be done, not just in the specific ways we have outlined in this chapter, but also by applying new methodologies to meta-analysis, ranging from response surface modeling where we predict results from ideal studies that do not even exist in our sample of studies (Rubin, 1990; Shadish et al., 2000), to mediational models that try to explain how treatments work (Becker, 1992; Shadish, 1996), to new computer programs that make doing all of this much easier (Borenstein & Rothstein, 1999; Shadish et al., 1999). We hope this chapter stimulates marriage and family researchers to continue producing and improving this kind of meta-analytic work.

REFERENCES

References marked with an asterisk indicate meta-analyses included in this review.

Becker, B.J. (1992). Models of science achievement: Forces affecting male and female performance in school science. In T. D. Cook, H. M. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 209–281). New York: Russell Sage Foundation.

Bellow, S. (1964). *Herzog.* New York: Viking Press.

Borenstein, M., & Rothstein, H. (1999). *Comprehensive meta-analysis.* Englewood, NJ: Biostat.

*Butler, M. H., & Wampler, K. S. (1999). A meta-analytic update of research on the couple communication program. *The American Journal of Family Therapy, 27,* 223–237.

*Cedar, B., & Levant, R. F. (1990). A meta-analysis of the effects of parent effectiveness training. *The American Journal of Family Therapy, 18,* 373–384.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H., & Hedges, L.V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Dickerson, K., Higgins, K., & Meinert, C. L. (1990). Identification of meta-analyses: The need for standard terminology. *Controlled Clinical Trials, 11,* 52–66.

*Dunn, R. L. (1994). A meta-analytic review of marital therapy outcome research: General and comparative analyses (Doctoral dissertation, The Ohio State University, 1993). *Dissertation Abstracts International, 54,* 5940B.

*Dunn, R. L., & Schwebel, A. I. (1995). Meta-analytic review of marital therapy outcome research. *Journal of Family Psychology, 9,* 58–68.

*Dutcher, T. D. (2000). A meta-analytic study of marital therapy modalities and clients' presenting problems (Doctoral dissertation, The Union Institute, 1999). *Dissertation Abstracts International, 60,* 5223B.

*Edwards, M. E., & Steinglass, P. (1995). Family therapy treatment outcomes for alcoholism. *Journal of Marital and Family Therapy, 21,* 475–509.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

*Giblin, P. (1985). A meta-analysis of premarital, marital and family enrichment research (Doctoral dissertation, Purdue University, 1982). *Dissertation Abstracts International, 46,* 470B.

*Giblin, P. (1986). Research and assessment in marriage and family enrichment: A meta-analysis study. *Journal of Psychotherapy and the Family, 2,* 79–96.

*Giblin, P., Sprenkle, D. H., & Sheehan, R. (1985). Enrichment outcome research: A meta-analysis of premarital, marital and family interventions. *Journal of Marital and Family Therapy, 11,* 257–271.

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher, 5,* 3–8.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339–353.

Haddock, C. K., Shadish, W. R., Klesges, R. C., & Stein, R. J. (1994). Treatments for childhood and adolescent obesity: A meta-analysis. *Annals of Behavioral Medicine, 16,* 235–244.

*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.

*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.

*Hight, T. L. (2000). Do the rich get richer? A meta-analysis of methodological and substantive moderators of couple enrichment (Doctoral dissertation, Virginia Commonwealth University, 2000). *Dissertation Abstracts International, 61*, 3278B.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage Publications.

Jacobson, N. S., Follette, W. C., Revenstorf, D., Baucom, D. H., Hahlweg, K., & Margolin, G. (1984). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. *Journal of Consulting and Clinical Psychology, 52*, 497–504.

Johnson, B. T. (1993). *D-Stat: Software for the meta-analytic review of research literatures* (version 1.1). Hillsdale, NJ: Erlbaum.

*Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (1999). Emotional focused couples therapy: Status and challenges. *Clinical Psychology: Science and Practice, 6*, 67–79.

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282*, 1054–1060.

Li, Z., & Begg, C. B. (1994). Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. *Journal of the American Statistical Association, 89*, 1523–1527.

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

Lipsey, M. W., & Wilson, D. B. (2000). *Practical Meta-Analysis.* Newbury Park, CA: Sage Publications.

*Mari, J. J., & Streiner, D. L. (1994). An overview of family interventions and relapse of schizophrenia: Meta-analysis of research findings. *Psychological Medicine, 24*, 565–578.

Markus, E. (1988). *Het effect van gezinstherapie: Een wantitatief literatuuroversicht.* Internal Report. University of Amsterdam, Department of Clinical Psychology.

*Markus, E., Lange, A., & Pettigrew, T. F. (1990). Effectiveness of family therapy: A meta-analysis. *Journal of Family Therapy, 12*, 205–221.

Matt, G. E. (1993). Comparing classes of psychotherapeutic interventions: A review and reanalyses of English- and German-language meta-analyses. *Journal of Cross-Cultural Psychology, 24*, 5–25.

*Montgomery, L. M. (1991). The effects of family therapy for treatment of child identified problems: A meta-analysis (Doctoral dissertation, Memphis State University, 1990). *Dissertations Abstracts International, 51*, 6115B.

Olkin, I., & Allen, I. E. (1999). Estimating the time for producing a meta-analysis from number of citations retrieved. *Journal of the American Medical Association, 282*, 634–635.

Paul, G. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*, 109–118.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal, 2*, 1243–1246.

Phillips, G. (1998). *Moderators of outcomes in psychotherapy meta-analyses.* Unpublished major area paper, The University of Memphis, Department of Psychology, Memphis, TN.

Phillips, G. A., & Shadish, W. R. (2002). *Moderators of outcomes in psychotherapy meta-analyses.* Manuscript in preparation.

*Pitschel-Walz, G., Leucht, S., Bauml, J., Kissling, W., & Engel, R. R. (2001). The effect of family interventions on relapse and rehospitalization in schizophrenia: A meta-analysis. *Schizophrenia Bulletin, 27*, 73–92.

*Plattor, J. R. (1991). Evaluating the effectiveness of marital therapies: A meta-analytic investigation (Doctoral dissertation, University of Miami, 1990). *Dissertation Abstracts International, 51*, 4605B.

Rosen, G. M., & Davison, G. C. (in press). Psychology should identify empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification.*

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166–169.

Rubin, D. B. (1990). A new perspective. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta-analysis* (pp. 155–165). New York: Russell Sage Foundation.

*Shadish, W. R. (1992). Do family and marital psychotherapies change what people do? A meta-analysis of behaviorai outcomes. In T. D. Cook, H. M. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 129–208). New York: Russell Sage Foundation.

Shadish, W. R. (1996). Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods, and issues. *Psychological Methods, 1*, 47–65.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Shadish, W. R., Hu, X., Glaser, R. R., Kownacki, R. J., & Wong, T. (1998). A method for exploring the effects of attrition in randomized experiments with dichotomous outcomes. *Psychological Methods, 3,* 3–22.

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin, 126,* 512–529.

*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.

Shadish, W. R., Ragsdale, K., Glaser, R. R., & Montgomery, L. M. (1995). The efficacy and effectiveness of marital and family therapy: A perspective from meta-analysis. *Journal of Marital and Family Therapy, 21,* 343–358.

Shadish, W. R., Robinson, L., & Lu, C. (1999). *ES: A computer program and manual for effect size calculation.* St. Paul, MN: Assessment Systems Corporation.

Shadish, W. R., & Sweeney, R. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology, 59,* 883–893.

Shoham-Salomon, V., & Rosenthal, R. (1987). Paradoxical interventions: A meta-analysis. *Journal of Consulting and Clinical Psychology, 55*(1), 22–28.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760.

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

*Stanton, M. D., & Shadish, W. R. (1997). Outcome, attrition, and family-couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. *Psychological Bulletin, 122,* 170–191.

*Sweeney, R. (1991). *Interactions: A secondary analysis of marital/family therapy meta-analysis.* Unpublished master's thesis, University of Memphis, Memphis, Tennessee.

*Wampler, K. S. (1982). Bringing the review of literature into the age of quantification: Meta-analysis as a strategy for integrating research findings in family studies. *Journal of Marriage and the Family, 44,* 1009–1023.

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

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

Wilkinson, L. & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

*Wilson, P. G. (1989). *The effects of marital therapy for the treatment of distressed couples: A quantitative analysis.* Unpublished master's thesis, University of Memphis, Memphis, Tennessee.

*Wilson, P. G. (1994). Marital/family versus individual therapy: A meta-analysis (Doctoral dissertation, Memphis State University, 1993). *Dissertation Abstracts International, 54,* 4414B.

Zhang, T. (1996). *The comparison between exact and inexact effect sizes computed by change score methods.* Unpublished master's thesis, University of Memphis, Department of Psychology, Memphis, Tennessee.

NOTES

1 See http://willshadish.psyc.memphis.edu/psychotherapy.htm

2 These error bars must be interpreted with great caution given that we had to estimate standard errors a variety of different ways when standard errors were not reported in the original meta-analysis, and that even when standard errors were included, they may have been computed in very different ways in different meta-analyses.

3 http://www.apa.org/divisions/div12/rev_est/bpd.shtml

4 http://www.cochrane.org/

5 http://www.apa.org/divisions/div12/rev_est/index.shtml

6 http://www.apa.org/divisions/div12/rev_est/drugs.shtml

7 http://www.apa.org/divisions/div12/rev_est/marital.shtml

8   Wilson (1994) initially examined differences between randomized and nonrandomized experiments, but excluded the latter when initial analyses suggested a trend ($p < .07$) toward overestimating effects in nonrandomized experiments.

9   Dunn (1994) notes that the sample included one study "in which the sample of couples served as their own controls" (p. 107).

10 Dutcher's (2000) sample of studies is nearly identical to Dunn's (1994) sample, and includes this same study without a control group.

11 The web site referenced in that article for obtaining that program is no longer available. However, researchers can obtain the program by clicking on ftp://141.225.10.58/aa.zip, which will download a zip file. Unzip the file, click on Setup.exe, and the program will be installed.

12 This statement applies to overall summaries of effects. When different kinds of effects are compared to each other, the most commonly recommended practice is that each study should contribute only one effect size to each cell of an analysis. If, for example, one is comparing self-report measures to observational measures, and a study has one measure of each kind, then the only cost of allowing that study to contribute one effect size to each cell is that the test of significance of the difference between the cells will be conservative (it will be harder to find a significant difference). Many researchers view that cost as acceptable. However, if that study had, say, two self-report measures and two observational measures, the meta-analyst should average the first two and average the second two before doing the test for differences between categories.