THE EFFECTS OF FAMILY THERAPIES FOR ADOLESCENT DELINQUENCY AND SUBSTANCE ABUSE: A META-ANALYSIS

Scott A. Baldwin, Sarah Christian, and Arjan Berkeljon
Brigham Young University

William R. Shadish
University of California

with Roy Bean
Brigham Young University

This meta-analysis summarizes results from k = 24 studies comparing either Brief Strategic Family Therapy, Functional Family Therapy, Multidimensional Family Therapy, or Multisystemic Therapy to either treatment-as-usual, an alternative therapy, or a control group in the treatment of adolescent substance abuse and delinquency. Additionally, the authors reviewed and applied three advanced meta-analysis methods including influence analysis, multivariate meta-analysis, and publication bias analyses. The results suggested that as a group the four family therapies had statistically significant, but modest effects as compared to treatment-as-usual (d = 0.21; k = 11) and as compared to alternative therapies (d = 0.26; k = 11). The effect of family therapy compared to control was larger (d = 0.70; k = 4) but was not statistically significant probably because of low power. There was insufficient evidence to determine whether the various models differed in their effectiveness relative to each other. Influence analyses suggested that three studies had a large effect on aggregate effect sizes and heterogeneity statistics. Moderator and multivariate analyses were largely underpowered but will be useful as this literature grows.

Adolescent delinquency, substance abuse, and other conduct problems are a major public health concern (e.g., Johnston, O'Malley, Bachman, & Schulenberg, 2008; Letourneau, Resnick, Kilpatrick, Saunders, & Best, 1996). Further, the treatment and management of adolescent offenders and their victims is costly (e.g., Post, Mezey, Maxwell, & Wibert, 2002). Consequently, it is essential that policy makers and marriage and family therapists identify interventions that reduce the incidence of delinquency and substance use among adolescents. Meta-analysis plays an important role in the evaluation and synthesis of research evidence regarding the effects of interventions. In our previous meta-analyses and reviews of meta-analyses, we have found that family therapy is effective as compared to control for a host of outcomes, including delinquency and substance abuse (Shadish & Baldwin, 2003; Shadish, Ragsdale, Glaser, & Montgomery, 1995). Four specific family therapy approaches show considerable promise for treating delinquency and substance abuse among adolescence: (a) Brief Strategic Family Therapy (BSFT; Szapocznik, Hervis, & Schwartz, 2003), (b) Functional Family Therapy (FFT; Alexander & Parsons, 1982), (c) Multidimensional Family Therapy (MDFT; Liddle & Hogue, 2001), and (d) Multisystemic Therapy (MST; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998). The first aim of this article is to meta-analytically combine outcomes from randomized trials involving BSFT, FFT, MDFT, and MST for adolescent delinquency and substance abuse.

Scott A. Baldwin, PhD, is on the Faculty of Psychology at Brigham Young University, Sarah Christian, BS, Doctoral student in the Department of Psychology at Brigham Young University, and Arjan Berkeljon, BS, Doctoral student in the Department of Psychology at Brigham Young University, Department of Psychology, Brigham Young University; William R. Shadish, PhD, is on the Faculty in the School of Social Sciences, Humanities, and Arts at the University of California, Merced. School of Social Sciences, Humanities, and Arts.

Address correspondence to Scott A. Baldwin, 268 TLRB, Brigham Young University, Provo, Utah 84460; E-mail: scott_baldwin@byu.edu

The second aim of this article is to introduce and apply advances in meta-analysis methodology so that family therapy researchers and clinicians can better interpret the meta-analysis literature.

FAMILY THERAPY FOR DELINQUENCY AND SUBSTANCE ABUSE

Family variables such as parental discipline, parent–child communication, parental substance use, and quality of the parent–child relationship have been linked with adolescent behavior problems (e.g., Liddle & Hogue, 2001). Consequently, it is not surprising that family-based approaches to treatment of delinquency and substance abuse have been identified as effective treatments for these problems. The four family therapy approaches with the largest research base are BSFT, FFT, MDFT, and MST. All four treatments are system-oriented approaches that aim to change dysfunctional family patterns that contribute to the onset and maintenance of adolescent delinquency and substance use. To accomplish these changes, these treatments focus on such things as helping parents and adolescents better communicate with each other and reduce conflict, helping parents improve their parenting skills (e.g., limit setting), and helping adolescents become better integrated with their extrafamilial environment (e.g., school; Liddle & Hogue, 2001). Detailed descriptions of the theory and implementation of BSFT, FFT, MDFT, and MST can be found in Szapocznik et al. (2003), Alexander and Parsons (1982), Liddle and Hogue (2001), and Henggeler et al. (1998), respectively.

Previous meta-analyses have shown that family therapy generally is effective for treating many kinds of problems, including alcoholism, schizophrenia, drug abuse, and conduct problems (Shadish & Baldwin, 2003). Each of the four family therapies has been studied in randomized trials and has been identified in narrative reviews as effective treatments (Alexander, Pugh, & Parsons, 2000; Chorpita et al., 2002; Liddle & Hogue, 2001; Szapocznik & Williams, 2000). Meta-analyses that have focused specifically on adolescent delinquency and substance abuse problems have had mixed findings. Waldron and Turner (2008) combined within-condition baseline to posttest effect sizes for 17 family therapy studies aimed at treating adolescent substance abuse. The average baseline to posttest effect for family therapy was $\bar{d} = 0.50$, whereas the average effect for the control conditions was $\bar{d} = 0.19$, suggesting that family therapy was more effective than control. Although this difference was statistically significant, it is difficult to interpret because standard errors for baseline to posttest effect sizes require knowledge of the correlation between baseline and posttest scores. If these correlations are ignored or the wrong correlation is used, then the standard errors will be biased (cf., Borenstein, 2009), which will affect statistical tests and confidence intervals involving the baseline to posttest effect sizes. Further, although the review included BSFT, FFT, MDFT, and MST studies, they did not distinguish between the treatment type in the analysis. In our analysis, we calculated between-condition effect size (family therapy versus control, treatment-as-usual, and alternative therapy), so we do not face the same problem as Waldron and Turner (2008) and we included a specific test for differences among the therapy types.

A second meta-analysis reviewed eight MST trials (Littell, Popa, & Forsythe, 2005). Across a number of analyses and outcomes, Littell et al. concluded that MST was not consistently more effective than treatment-as-usual. This conclusion was at odds with a number of narrative reviews of MST's effects (cf., Henggeler, Schoenwald, Swenson, & Borduin, 2006; Littell et al., 2005). Henggeler et al. (2006) suggested that this difference was owing to, among other things, too much weight being given to a single, unpublished study (Leschied & Cunningham, 2002), although Littell (2006) had noted that no weighting in the analysis occurred beyond what is customary in meta-analysis. In our analysis, we implement influence analysis, which allows us to explore whether a given study is carrying more weight in the aggregate effect size. In any case, the debate about Littell et al.'s findings highlights two important aspects of meta-analysis that we attend to in our review: locating and including unpublished studies and identifying studies that highly influence aggregate effect sizes.

META-ANALYSIS METHODOLOGY

The process of meta-analysis is a complicated one. Indeed, entire books have been devoted to the subject (e.g., Cooper, Hedges, & Valentine, 2009). In this section, we briefly describe the

typical methods used in meta-analyses. We then discuss advances in meta-analysis methodology that we use in our review. We also note that considerable attention has been devoted to reporting standards in meta-analysis (APA Publications and Communications Board Working Group on Journal Article Reporting, 2008; Liberati et al., 2009). These standards increase the transparency and clarity in the reporting of meta-analysis, which helps with replicability of results. To the extent possible, we adhere to these reporting standards throughout this article.

Standard Methods

The process of meta-analysis consists of five general steps each of which consists of several substeps. First, the researcher decides on a research question, such as what are the effects of BSFT, FFT, MDFT, and MST for adolescent deliquency and substance abuse? Second, the researcher uses computer and manual searches to locate and obtain all relevant studies (or a sample of studies), including unpublished studies that meet the inclusion criteria. Third, the researcher collects the data. In meta-analysis, this involves creating a coding manual, establishing interrater reliability, and coding effect sizes and study characteristics (e.g., published versus unpublished studies). Fourth, the researcher combines the effect sizes. The current standard in meta-analysis is to use random effect models to combine effect sizes as these models allow for generalizations beyond the studies included in the model (Hedges & Vevea, 1998). Fifth, the researcher uses meta-regression to examine whether study characteristics account for variability in effect sizes (i.e., moderator analyses). For example, the researcher can compare whether published and unpublished studies have different effect sizes. Meta-regression is conceptually similar to standard least-squares regression. It differs from least-squares regression in that the estimation techniques are specifically tailored to the unique aspects of meta-analytic data (Harbord & Higgins, 2008).

Advanced Methods

Given the widespread use of meta-analysis in the social sciences and medicine, methodologists have devoted considerable attention to improving the meta-analytic methodology and thereby improving inferences drawn from meta-analyses. We introduce and implement three advanced methods: (a) influence analysis, (b) multivariate meta-analysis, and (c) publication bias analyses.

Influence analysis. The primary purpose of an influence analysis is to determine whether there are studies that have a large impact on aggregate results. An influence analysis begins with a "leave-one-out" analysis (Viechtbauer, 2010), where the meta-analysis is repeated k times—k is the number of studies. At each iteration, one study is left out of the analysis. Influential studies will have a large impact on aggregate effect sizes ($\bar{d}$) and heterogeneity statistics (i.e., indices of how much studies differ beyond sampling error) such as the between-study variance component (τ^2) and I^2. Once influential studies have been identified, researchers can determine whether there are characteristics of the studies that may explain why the studies are outliers.

Multivariate meta-analysis. A limitation of standard meta-analytic methods is that they only allow one effect size per study, which prohibits within-study comparisons. For example, effects may be larger for primary outcomes versus secondary outcomes within studies. We could compare primary and secondary outcomes by (a) creating a variable corresponding to primary and secondary outcomes within a study and using meta-regression to compare the differences between outcome types and (b) constructing separate datasets for primary and secondary outcomes and conducting univariate analyses for each dataset. However, the first approach is problematic because a given study would contribute multiple effect sizes to the analysis, which violates the assumption of independence of observations. Ignoring nonindependence of observations will produce standard errors that are too small, which means p-values will be too small and confidence intervals too narrow. The second approach is problematic because it assumes that there is no correlation between primary and secondary outcomes within and between studies.

A third approach is to use a multivariate meta-analysis (Kalaian & Raudenbush, 1996; Nam, Mengersen, & Garthwaite, 2003; Raudenbush, Becker, & Kalaian, 1988; Riley, Abrams,

Lambert, Sutton, & Thompson, 2007; White, 2009), which allows the analyst to include multiple outcomes from a single study in an analysis. Multivariate meta-analysis explicitly models the relationship between primary and secondary outcomes within studies, so it produces appropriate standard errors, p-values, and confidence intervals (Raudenbush et al., 1988). Additionally, a multivariate analysis estimates the relationship between treatment effects on primary and secondary outcomes between studies (Riley, Abrams, Lambert et al., 2007). The primary challenge with multivariate meta-analysis is that it requires estimates of the within-study correlations among outcomes, which are rarely reported. This can be overcome using external estimates of the within-study correlations or using a sensitivity approach where the magnitude of the within-study correlation is varied to determine a range of plausible outcomes.

Publication bias. Publication bias occurs when the publication of a study depends upon the statistical significance of the results (Greenwald, 1975; Rosenthal, 1979). Editors and reviewers may be less willing to publish null or negative findings or researchers may not even submit studies with null or negative findings. Consequently, the published literature may be skewed toward positive results, which will inflate aggregate effect sizes. The best method for dealing with publication bias is to include unpublished studies (we were able to locate one unpublished study that met our inclusion criteria). Additionally, researchers can assess for publication bias with funnel plots (Peters, Sutton, Jones, Abrams, & Ruston, 2008; Sterne & Harbord, 2004), Begg's rank correlation test (Begg & Mazumdar, 1994), and Egger's regression test (Egger, Smith, Schnieder, & Minder, 1997), trim-and-fill analysis (Duval & Tweedie, 2000a,b), and selection models (Vevea & Woods, 2005).

Funnel plots are scatterplots with effect sizes on the x-axis and standard errors on the y-axis. The standard errors are plotted from small on the top of the y-axis to large on the bottom. If there is no publication bias, the plot results in a symmetrical inverted funnel-shaped graph, narrow at the top and wide at the bottom because studies with small standard errors are less subject to sampling variability than studies with large standard errors. Thus, at the top of the graph, studies with small standard errors will be close together, and as standard errors get larger, the effect sizes will be more variable. If null or negative findings are less likely to be published, then the lower left part of the distribution would be missing. Therefore, an asymmetrical funnel plot can provide evidence of publication bias.

To help with the assessment of asymmetry, pseudo-95% confidence intervals around the aggregate effect size are added to the plot. If there is no publication bias, then 95% of the studies should fall within these limits. Contour-enhanced plots can also help researchers identify funnel plot asymmetry (Peters et al., 2008). In contour-enhanced plots, regions of statistical significance are added to the funnel plot. If all studies fall in the areas of statistical significance or if the truncated part of the funnel is only in areas of nonsignificance, then we have some evidence that publication bias is present. Unlike typical funnel plots, contour-enhanced funnel plots include a vertical reference line draw at zero because the null hypothesis for the significance tests is that the overall effect is zero.

Although funnel plots are conceptually simple and easy to implement, they have two main disadvantages. First, judgments of asymmetry are difficult when the number of studies is small. Second, asymmetry can occur for a number of reasons in addition to publication bias, including true heterogeneity owing to sample size, data irregularities (e.g., small studies may have poor or even superior methodology), artifacts, and chance (Sterne & Harbord, 2004). Contour-enhanced funnel plots can help researchers determine whether asymmetry is associated with statistical significance but does not necessarily rule out the other explanations. Consequently, we recommend that researchers supplement funnel plots with other analyses.

Begg and Mazumdar's (1994) rank correlation test and Egger et al.'s (1997) regression test provide statistical tests of asymmetry. Both tests assess the relationship between effect sizes and standard errors, and statistically significant relationships suggest the possibility of publication bias. Like the funnel plot, both the rank correlation and regression tests assume homogeneity of effect sizes.

One disadvantage of the previous methods for assessing publication bias is that they do not provide an estimate of what the aggregate effect size would be without publication bias. A commonly used method for obtaining an adjusted effect size is trim-and-fill analysis (Duval &

Tweedie, 2000a,b). The trim-and-fill method estimates how many studies are needed to make the funnel plot symmetrical and generates effect size estimates for the missing studies. Then, the aggregate effect size is estimated again using both the original and filled effect sizes to provide an overall effect size absent publication bias. When interpreting the results of a trim-and-fill analysis, it is important to keep in mind that it makes a strong assumption that missing studies are those with the most negative effects and, because it assumes homogeneity of effect sizes, it can falsely detect missing studies when there is heterogeneity in the effect sizes (Terrin, Schmid, Lau, & Olkin, 2003; Vevea & Woods, 2005).

An alternative method for obtaining effect sizes adjusted for publication bias are selection models (Vevea & Hedges, 1995; Vevea & Woods, 2005), which model the probability that a particular study is published. If publication bias is present, the larger the p-value for a given study, the lower the probability of publication. Furthermore, there should be a sharp decrease in probability of publication at or around $p = .05$. In selection bias models, the weight is the probability of studies with $p > .05$ surviving the selection process relative to studies with $p < .05$; selection process refers to publication bias based on $p < .05$. For example, a weight of .40 indicates that a study with $p > .05$ is 40% as likely as a study with $p < .05$ to survive the selection process. Like trim-and-fill, selection bias models produce an aggregate effect size adjusted for publication bias. The difficulty with most selection bias methods is that estimating weights for various p-values requires more studies (greater than 100; Vevea & Woods, 2005).

To overcome this requirement, Vevea and Woods (2005) adapted the selection bias models for meta-analyses with small numbers of studies. In this new model, weights are set by the researcher rather than estimated by the model. Vevea and Woods recommend that multiple sets of weights, each representing a different selection process, be used in a sensitivity analysis to allow the researcher to see how different assumptions about the selection process affect the aggregate estimate. An example of a possible selection process is a one-tailed moderate selection, where studies with small p-values have a high probability ($> .90$) of being observed and studies with large p-values have a moderate probability ($\approx .50$) of being observed. In a one-tailed severe selection, studies with large p-values have a low probability of being observed ($< .20$). An advantage of selection models over the other publication bias methods we have reviewed is that it can include moderator variables. Thus, if it appears that funnel plot asymmetry may be due to heterogeneity among the studies, we can include a variable that accounts for that heterogeneity and determine whether there is evidence of publication bias above and beyond the heterogeneity (Vevea & Woods, 2005).

When interpreting results of publication bias analyses, it is important to remember that the analyses provide indirect evidence of publication bias. That is, they do not prove that publication bias is a problem. Further, given the limitations of any one publication bias method, it is wise to use multiple methods for assessing publication bias and determine whether the different methods each suggest publication bias. In the end, however, the only sure way to rule out publication bias is to locate and include unpublished studies in the meta-analysis.

The Present Meta-analysis

In this meta-analysis, we evaluate the posttreatment effects of BSFT, FFT, MDFT, and MST on adolescent delinquency and substance abuse as compared to treatment-as-usual (TAU), alternative therapy, and control. We extend the previous work in by (a) updating the meta-analytic evidence to include recent studies, (b) focusing on between-condition comparisons, (c) including all four family therapy types and exploring differences in effects by treatment type, (d) identifying highly influential studies, (e) using multivariate models to explore differences between primary and secondary outcomes, and (f) comprehensively assessing for publication bias.

METHODS

Identification of Studies

Studies included in this meta-analysis are randomized trials that compared MST, FFT, MDFT, or BSFT to a comparison condition in the treatment of adolescent (age 11–19) delinquency, conduct problems, or substance use. The treatment had to be delivered on its own (i.e.,

not combined with other interventions). We excluded prevention studies. We included both published and unpublished studies that met our inclusion criteria.

We identified studies by performing an electronic search on PsycINFO, Medline, and Dissertation Abstracts International for randomized trials published by February 2009. We used the following search terms: Multisystemic Therapy or Functional Family Therapy or Multidimensional Therapy or Brief Strategic Family Therapy or Family Therapy AND delinquency or delinquent or substance use or conduct disorder or externalizing. Additionally, we searched reference lists of previous reviews or studies of the four family therapies and the websites of the developers of the four family therapies. Eligible studies were examined by two reviewers. Any disagreements were settled by consensus. Figure 1 describes the selection of studies.

Data Collection

Effect size coding. The effect size measure used in this meta-analysis is standardized mean difference statistic. The statistic was computed directly as:

$$d = \frac{\bar{X}_{\mathrm{T}} - \bar{X}_{\mathrm{C}}}{s_{\mathrm{p}}} \tag{1}$$

where $\bar{X}_{\mathrm{T}}$ is the mean of the family therapy group, $\bar{X}_{\mathrm{C}}$ is the mean of the comparison group, and s_{p} is the pooled standard deviation. When the means and standard deviations were not available, we used the methods described in Shadish, Robinson, and Lu (1999). Results reported only as nonsignificant were coded as $d = 0.00$. All effect sizes were corrected for small sample bias (Hedges & Olkin, 1985). We only used effect sizes calculated at the first assessment point for a given measure following treatment. Effect sizes were computed by two independent reviewers, and discrepancies were resolved via consensus.

Moderators and study characteristics. We used a coding manual to extract the following information for each study. Year of publication, published/unpublished, family therapy type (MST, FFT, MDFT, or BSFT), presenting problem (delinquency/conduct disorder, substance abuse, or mixed), mean age of participants, proportion of participants female, proportion of participants who are from minority ethnic groups, method of referral (court or other), outcome measure, and time since end of treatment that a measure was administered (in months). We also coded comparison type (no treatment control, treatment-as-usual, or alternative therapy). Treatment-as-usual was defined as the standard care for the participants, such as a court referral to a juvenile delinquency treatment center. Alternative therapy was defined as treatments that were provided or overseen by the researchers but that were not family therapy.
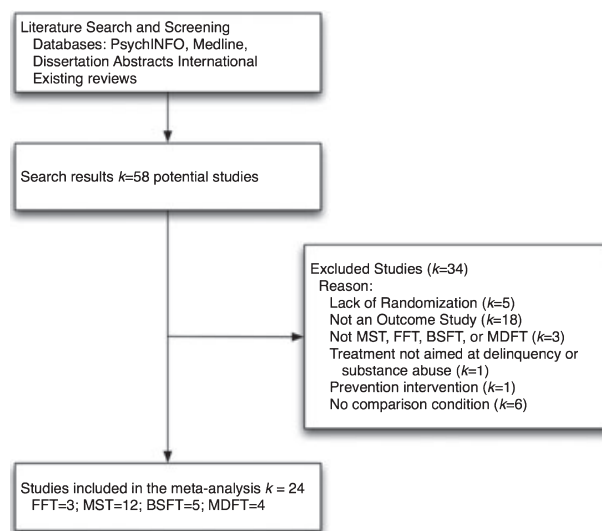

Figure 1. Flow chart describing the identification and selection of studies.

To assess interrater reliability, two reviewers independently extracted the above information for one treatment/comparison contrast and one outcome measure for 10 studies. Codes with unacceptable reliability were rewritten and piloted on 10 new studies. For categorical variables, final reliability ranged from $\kappa = .68$ to 1.0. For continuous variables, final reliability included intraclass correlations that ranged between .98 and 1.0. With regard to κ, Fleiss (1981) stated that "for most purposes, values $> .75$ or so may be taken to represent excellent agreement beyond change, values below .40 or so may be taken to represent poor agreement beyond change, and values between .40 and .75 may be taken to represent fair to good agreement beyond chance" (p. 218). Thus, all variables were sufficiently reliable to proceed with coding. All studies were coded by two reviewers. As before, disagreements were settled via consensus.

Data Analysis

For univariate analyses, we used the *metafor* package (version 1.0-1; Viechtbauer, 2010) in R (version 2.10; R Development Core Team, 2009) to estimate the random effect models and influence analyses and publication bias analyses. In all univariate analyses, data were collapsed so that each study only contributed one effect size to the analysis. For the mulitvariate analyses, we used the *mvmeta* program (White, 2009) for Stata (version 11; StataCorp, 2009) to estimate the random effect models.

RESULTS

We first present results combining all studies regardless of type of comparison group. We do not report the aggregated effect size from this analysis, however. Instead, we use these analyses to explore the distribution of effect sizes across studies, in particular the heterogeneity of effect sizes. We also searched for outlier studies that had a large influence on the results. Next, we present univariate and multivariate results stratified by comparison type. These analyses include influence analyses, moderator analyses, and multivariate analyses. Additionally, we explore whether publication bias influences the results of the meta-analysis.

Overall Analysis

We aggregated across all $k = 24$ studies regardless of comparison type to explore heterogeneity in effect sizes and to identify influential studies. For studies that included two comparison groups (Alexander & Parsons, 1973; Szapocznik et al., 1989), we retained the control condition for the overall analysis. We do not report the overall aggregated effect size because the varying comparison type makes it difficult to interpret. Homogeneity of effect sizes was rejected, $Q(23) = 49.26$, $p < .05$, indicating more between-study variability than what would be expected from sampling error. The between-study variance component was $\tau^2 = 0.05$, and the degree of heterogeneity between studies was moderate, $I^2 = 52.9\%$.

We focused only on the heterogeneity statistics in the influence analysis involving all studies in the review. The "leave-one-out" analysis indicated that three studies were influential with respect to heterogeneity (Dennis et al., 2004; Nickel, Luley et al., 2006; Timmons-Mitchell, Bender, Kishna, & Mitchell, 2006). Removing the studies one at a time reduced I^2 from 52% to between 38% and 44%. Nickel, Luley, et al. (2006) and Timmons-Mitchell et al. (2006) are the two largest effect sizes ($d = 1.43$ and 0.96, respectively) in the meta-analysis, whereas the study by Dennis et al. (2004) is the smallest ($d = -0.16$). We reviewed these three studies to determine whether they had any shared characteristics that set them apart from the other studies besides extreme effect sizes, but no other complicating factors were clearly evident. The three outliers involved three treatment types (MST, MDFT, BFST), all three comparison group types, and two problem types (substance abuse and delinquency).

We repeated the overall analysis, leaving out the three influential studies. In the second analysis, homogeneity of effect sizes was not rejected, $Q(20) = 20.59$, $p = .42$. The between-study variance component was reduced to $\tau^2 = 0.01$ (an 80% reduction), and I^2 was reduced to 11.18%. Together, these results suggest that once the outliers are removed, between-study variability is small and is what we would expect to find by chance alone.

Not surprisingly, effect sizes that compared family therapy to either TAU or an alternative treatment were smaller than effect sizes that compared family therapy to control. Both TAU and alternative therapy differed from control by $d = -0.45$. However, comparison type was not a significant moderator of effect size, $F(2, 21) = 2.6$, $p = .09$. Two problems complicate the interpretation of this moderator analysis. First, the analysis includes a moderate amount of studies ($k = 24$), and some comparison types have few studies, such as family therapy versus control where $k = 4$. Power is 52%, given a difference of $d = -0.45$, $k = 24$, and considerable variability in the number of studies within comparison type. Second, the moderator analyses included the three outlier studies, which we have shown introduce considerable heterogeneity into the analysis. We repeated the analysis removing the outlier studies, and comparison type was a significant moderator of effect size, $F(2, 18) = 3.6$, $p < .05$. Together, these results suggest that there may be large differences in effect size owing to type of comparison group used. Future meta-analyses should continue to test for this difference when computing aggregated effect sizes.

Family Therapy versus Treatment-as-Usual

Univariate analyses. Eleven studies randomized participants to family therapy or treatment-as-usual (TAU; MST = 10, BSFT = 1). The left panel of Figure 2 displays a forest plot of the study-level effect sizes (aggregated across measures within a study) and 95% confidence intervals for each study, along with the citation, total sample size for the family therapy versus TAU comparison, and the aggregate effect size and confidence interval. Study-level effect sizes are represented by squares, where the size of the square represents the weight of the study in the analysis. The random effect weighted-average effect size comparing family therapy to TAU was $\bar{d} = 0.21$ ($p = .03$, 95% CI [0.02–0.40]). The between-study variance component was $\tau^2 = 0.04$ and $I^2 = 47.23\%$, indicating that 47.23% of the variance in effect sizes was between studies. Homogeneity of effect size was not rejected, $Q(10) = 17.91$, $p = .06$, although it is close and power is low with only 11 studies. The average effect size for MST was $\bar{d} = 0.22$ ($p = .04$, 95% CI [0.01–0.43]) and for BSFT was $\bar{d} = 0.09$ ($p = .82$, 95% CI [−0.83, 1.02]). This difference was not statistically significant, although again power is quite low.

As with the overall analysis, we used a "leave-one-out" analysis to identify highly influential studies. The left-hand panel of Figure 3 displays aggregate effect sizes and confidence intervals when a given study is removed from the family therapy versus TAU analysis. The aggregate effect sizes in the influence analysis ranged from 0.12 to 0.24. Removing Timmons-
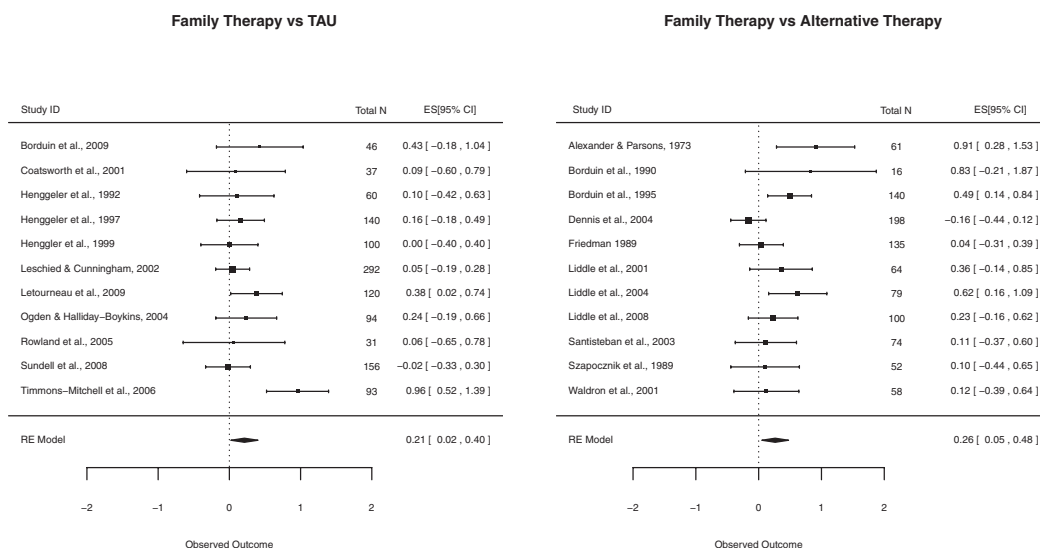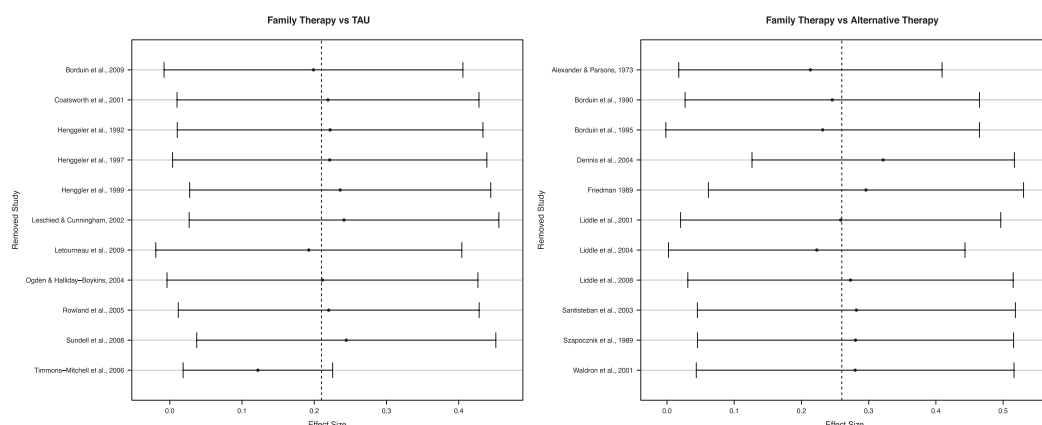

Figure 2. A forest plot of effect sizes and 95% confidence intervals for studies in the family therapy versus treatment-as-usual (TAU) and the family therapy versus alternative therapy meta-analyses.

Figure 3. An influence plot from a "leave-one-out" analysis. The aggregate effect size from all studies is represented by a dashed line. The dots and confidence bars represent the aggregate effect size, and 95% confidence interval after the study on the *y*-axis was removed.

Mitchell et al. (2006) reduced the overall effect size from $\bar{d} = 0.21$ to $\bar{d} = 0.12$, which is a 42% reduction. Additionally, removing Timmons-Mitchell et al. reduced τ^2 and I^2 to zero, suggesting that all heterogeneity in the family therapy versus TAU comparisons is because of the Timmons-Mitchell et al. study. Removing Borduin, Schaeffer, and Heiblum (2009), Letourneau et al. (2009), or Ogden and Halliday-Boykins (2004) slightly reduced the aggregate effect size and resulted in a nonsignificant effect size—*p*-values ranged from .053 to .07. However, the changes in both the aggregate effect size and *p*-values were slight and are only perceived as important changes because the overall *p*-value is close to .05. Consequently, the study by Timmons-Mitchell et al. was the only study treated as an outlier in subsequent analyses. Excluding Timmons-Mitchell et al., a MST study, the average effect size for MST was $\bar{d} = 0.13$ ($p = .04$, 95% CI [0.01–0.24]).

Moderator analysis. We used meta-regression to explore whether the following study characteristics moderated effect size: sample size, average age of participants, proportion of participants who were female, proportion of participants who were part of an ethnic minority group,

Table 1
Results of Moderator Analyses

Moderator	Comparison type		
	TAU	Alternative treatment	Control
Sample size	−0.001	−0.003	−0.01
Mean age	−0.02	−0.03	0.15
Proportion female	−0.91	−0.44	0.97
Proportion minority	−0.74	0.15	n/a
Time since treatment	−0.01	0.01	0.01
Presenting problem: substance abuse	−0.26	−0.49*	n/a
Presenting problem: other	−0.18	−0.54	−0.79
Court referral	0.07	0.50*	0.14

Note. n/a = could not estimate because there were no sufficient data; TAU = treatment-as-usual.

timing of posttreatment assessments, presenting problem, and referral source (court versus other). Table 1 presents the results of the moderator analysis. None of the study characteristics were significant predictors of effect size, although with only 11 studies power is a problem. When we reran the analyses excluding Timmons-Mitchell et al. (2006), only proportion female was significant ($\beta = -.86$, $p < .05$), indicating that the more women in the study, the smaller the effect size. However, given that 14 separate analyses were conducted, it is reasonable to expect that one of fourteen tests would be significant by chance alone. In fact, the proportion female was not significant once we applied a Bonferroni correction to the p-value. Thus, researchers should be cautious in interpreting this particular finding.

Multivariate analyses. The top rows of Table 2 present the results of a multivariate meta-analysis comparing family therapy to TAU. We compared delinquency/substance abuse measures to other measures (e.g., internalizing, noncriminal externalizing behaviors, family). Given that within-study correlations among these measures were not available, we repeated the multivariate analysis three times, setting the within-study correlation to .25, .5, or .75. The omnibus test for the multivariate analyses tests the null hypothesis that all average effect sizes are zero. The null hypothesis was rejected in all cases. Table 2 also presents univariate analyses for comparison purposes.

Across analyses, effect sizes were larger for delinquency and substance abuse measures compared to other measures. Only delinquency measures were statistically different from zero. In the multivariate model, we can also test whether the average effect sizes differ from one another. For example, assuming a within-study correlation of .5, the average effect size for delinquency measures ($\bar{d} = 0.31$) does not significantly differ from the average effect size for other measures ($\bar{d} = 0.17$; $\chi^2(1) = 1.01$, $p = .3$). As it turns out, delinquency and substance abuse measures do not differ from other measures, regardless of the size of the within-study correlation. Nevertheless, as with other analyses, power is low and it is probably wise to model these effect sizes separately.

We reran the multivariate analyses excluding Timmons-Mitchell et al. (2006). The aggregate effect size for delinquency/substance abuse measures was similar to the previous model, $\bar{d} = 0.30$, .30, .31 for within-study correlations of .25, .5, and .75, respectively. However, the aggregate effect size for other measures was approximately half of the previous model, $\bar{d} = 0.09$, 0.09, 0.08 for within-study correlations of .25, .5, and .75, respectively. Thus, there is clear evidence that the study by Timmons-Mitchell et al. was highly influential.

The univariate and three multivariate analyses produced similar average effect sizes. However, the multivariate analyses produced slightly larger estimates of τ^2. Although the difference in τ^2 between the models had little effect in the present meta-analysis, changes in the τ^2 affect significance tests and confidence intervals (White, 2009), making it important for future meta-analyses to consider a multivariate analysis. Additionally, the multivariate analyses estimated the between-study correlations among outcomes. These correlations are interpreted as the relationship between the effects of family therapy as compared to control on delinquency/substance abuse measures and on other measures. These correlations were small or even negative when the within-study correlation is assumed to be .75, which indicates that the effects of family therapy as compared to TAU on delinquency/substance abuse measures were relatively independent of its effects on other measures.

Family Therapy versus Alternative Treatments

Univariate analyses. Eleven studies randomized participants to either family therapy or an alternative treatment (MST = 2; FFT = 3; MDFT = 4; BSFT = 3). Alternative treatments included group therapy, psychodynamic family therapy, individual therapy, parent groups, and family education therapy. The right-hand panel of Figure 2 presents a forest plot of the results. The random effect weighted-average effect size comparing family therapy to alternative treatments was $\bar{d} = 0.26$ ($p < .05$, 95% CI [0.05–0.48]). Homogeneity was rejected, $Q(10) = 20.18$, $p < .05$, $\tau^2 = 0.05$, and $I^2 = 48.74\%$, indicating that 48.74% of the total variance in effect sizes was between studies. The average effect size for MST was $\bar{d} = 0.57$ ($p = .07$, 95% CI [−0.07 to 1.20]), for BSFT was $\bar{d} = 0.11$ ($p = .68$, 95% CI [−0.49 to 0.70]), for MDFT was $\bar{d} = 0.22$ ($p = .21$, 95% CI [−0.16 to 0.60]), and for FFT was $\bar{d} = 0.29$ ($p = .19$, 95% CI

Table 2
Results of Multivariate Meta-analyses Comparing Delinquency/Substance Abuse (D/SA) Measures and Other Measures

Outcome	Univariate		Mulitivariate ($\rho_W = .25$)			Mulitivariate ($\rho_W = .5$)			Mulitivariate ($\rho_W = .75$)		
	d	τ^2	d	τ^2	ρ_B	d	τ^2	ρ_B	d	τ^2	ρ_B
Family therapy versus TAU											
D/SA	0.31**	0.09	0.31**	0.09	.28	0.31**	0.09	.09	0.31**	0.10	−.08
Other	0.18	0.06	0.18	0.06		0.17	0.07		0.17	0.07	
Omnibus test			$\chi^2(2, k = 11) = 8.31, p < .05$			$\chi^2(2, k = 11) = 8.07, p < .05$			$\chi^2(2, k = 11) = 7.80, p < .05$		
Family therapy versus alternative therapy											
D/SA	0.43*	0.11	0.41*	0.23	1	0.40*	0.23	1	0.39*	0.23	1
Other	0.26**	0	0.24**	0.01		0.24**	0.01		0.24**	0.01	
Omnibus test			$\chi^2(2, k = 11) = 10.29, p < .05$			$\chi^2(2, k = 11) = 10.35, p < .05$			$\chi^2(2, k = 11) = 10.53, p < .05$		

Note. D/SA = delinquency and substance abuse outcomes; ρ_W = within-study correlation; ρ_B = between-study correlation. *$p < .05$; **$p < .01$.

[−0.18 to 0.76]). These differences among treatment type were not statistically significant, although statistical power remains a limitation in making this determination.

The right-hand panel of Figure 3 displays the aggregate effect sizes and confidence intervals for the ''leave-one-out'' analysis. The aggregate effect sizes in the influence analysis ranged from 0.21 to 0.32, whereas τ^2 ranged from 0.01 to 0.06. No study had as dramatic an effect in this analysis as Timmons-Mitchell et al. (2006) did in the family therapy versus TAU analysis. Removing Dennis et al. (2004) increased $\bar{d}$ to .32 (a 23% increase) and decreased τ^2 to 0.01 (an 80% reduction). The study by Dennis et al. is the largest study ($N = 198$) and has the smallest effect size ($d = -0.16$). No other study appeared to have a significant effect on the results. Consequently, we only considered Dennis et al. (2004) an outlier in subsequent analyses. Excluding Dennis et al., a MDFT study, the average effect size for MDFT was $\bar{d} = 0.39$ ($p = .04$, 95% CI [0.01–0.77]).

Moderator analysis. The middle columns of Table 1 present the results of the moderator analysis for family therapy versus alternative treatment studies. Studies focused on treating substance abuse problems had significantly worse outcomes than studies that focused on treating delinquency, $\beta = -.49$, $p < .05$, 95% CI [−0.95 to 0.04]. Additionally, studies where participants were recruited via court referrals had significantly better outcomes than studies that used other means of recruitment, $\beta = .49$, $p < .05$, 95% CI [−0.09 to 0.91]. However, court referral and presenting problem are completely confounded—the only studies that used court referrals as their sole recruitment method were focused on treating delinquency. We repeated the analyses removing Dennis et al. (2004). The pattern of results was similar, except that the *p*-value for the coefficient comparing substance abuse and delinquency as presenting problems slightly exceeded .05.

Multivariate analysis. The bottom rows of Table 2 present the results of the multivariate meta-analysis comparing family therapy to alternative therapy. As before, we compared delinquency/substance abuse measure to other measures and repeated the multivariate analysis three times, setting the within-study correlation to .25, .5, or .75. Across multivariate analyses, the omnibus test was significant, indicating that at least one aggregate effect size was statistically significant. Aggregate effect sizes were statistically different from zero for both delinquency/substance abuse measures and other measures. Effect sizes were larger for delinquency/substance abuse measures than for other measures, but these differences were not statistically significant, regardless of the size of the within-study correlation. Repeating these analyses after removing Dennis et al. (2004), we found that aggregate effect sizes for other measures remained approximately the same. Aggregate effect sizes for delinquency/substance abuse measures increased to $\bar{d} = 0.49$, 0.48, and 0.46 assuming a within-study correlation of .25, .5, and .75, respectively.

The between-study correlation was estimated as one across all analyses. This should not be interpreted as a perfect correlation between changes in delinquency/substance abuse measures and other measures. Instead, the perfect correlation is a consequence of the estimation procedures, which require that correlations do not exceed ± 1 (Riley, Abrams, Sutton, Lambert, & Thompson, 2007). That is, to ensure that between-study correlations are not estimated beyond the theoretical boundary of a correlation, the estimation procedure truncates the correlation at ± 1. This truncation is common when the number of studies is small (Riley, Abrams, Sutton et al., 2007), as is the case in this analysis. The truncation of correlations is similar to nonnegativity constraints on variance components in multilevel modeling software. Commenting on correlations equal to ± 1 Riley, Abrams, Sutton et al., (2007) state: ''Practitioners should not, though, be overly concerned about this. We have shown it does not cause any systematic bias in the pooled estimates from BRMA [i.e., multivariate models], and it leads to conservative standard errors and mean-square errors'' (Between-study covariance parameters, para. 1).

The univariate and three multivariate analyses produced similar average effect sizes, but the multivariate analyses produced larger estimates of τ^2. Specifically, τ^2 increased from 0.11 in the univariate model to 0.23 in all three multivariate models (a 110% increase). Thus, some of the between-study variability in the delinquency/substance abuse effect sizes is masked by ignoring the correlation between outcomes within studies. Because τ^2 contributes to the standard error of the aggregate effect size, underestimating it, as occurred in the univariate models, leads to *p*-values that are too small and confidence intervals that are too narrow.
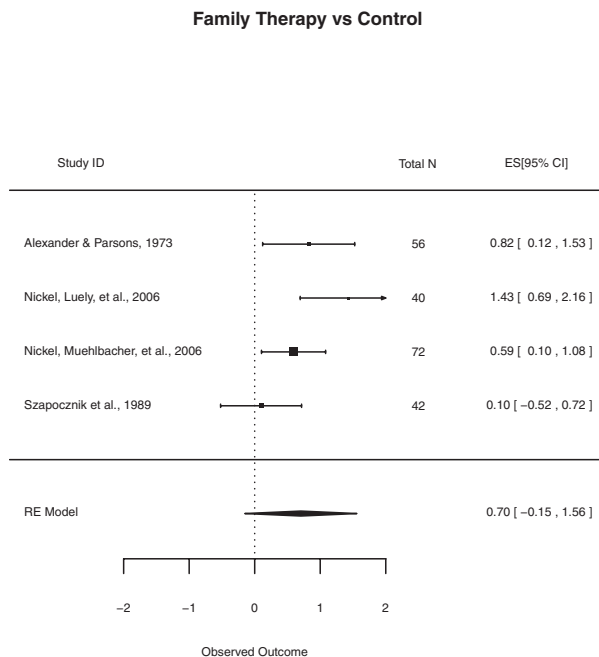
Family Therapy versus Control

Univariate analyses. Four studies randomized participants to either family therapy or a control condition (FFT = 1; BSFT = 3). Figure 4 presents a forest plot of the results. The random effect weighted-average effect size comparing family therapy to control was $\bar{d} = 0.70$ ($p = .08$, 95% CI [−0.15 to 1.56]). Homogeneity was rejected, $Q(3) = 7.67$, $p = .05$, $\tau^2 = 0.17$, and $I^2 = 62.68\%$, indicating that 62.68% of the total variance in effect sizes was between studies. The average effect size for FFT was $\bar{d} = 0.82$ ($p < .05$, 95% CI [0.12–1.53]) and for BSFT was $\bar{d} = 0.68$ ($p = .21$, 95% CI [−0.94 to 2.30]). This difference was not statistically significant, although power remains a problem.

Figure 5 displays the aggregate effect sizes and confidence intervals for the "leave-one-out" influence analysis. Influence analyses are difficult to interpret when there are only four studies and when there is a large of amount of heterogeneity among those studies. Consequently, these analyses should be interpreted with caution. The aggregate effect sizes in the influence analysis ranged from 0.50 to 0.89 (none were statistically significant) whereas τ^2 ranged from 0.02 to 0.33. Removing Nickel, Luely et al. (2006) reduced $\bar{d}$ to .50 (a 29% reduction) and removing Szapocznik et al. (1989) increased $\bar{d}$ to .90 (a 29% increase). Further, removing Nickel, Luely et al. (2006) reduced τ^2 to 0.02 (an 88% reduction) and removing Szapocznik et al. (1989) reduced τ^2 to 0.08 (a 53% reduction). On the other hand, removing Alexander and Parsons (1973) and Nickel, Muehlbacher, et al. (2006) increased τ^2 to 0.31 (an 82% increase) and 0.32 (an 88% increase), respectively. Because removing any study within this analysis had a large impact on the results, we only considered Nickel, Luely et al. (2006) to be an outlier in subsequent analyses because it was also an outlier in the overall analysis. Excluding Nickel, Luely et al., a BSFT study, the average effect size for BSFT was $\bar{d} = 0.38$ ($p = .36$, 95% CI [−2.72 to 3.5]).

Moderator analysis. The right-hand columns of Table 1 present the results of the moderator analysis for family therapy versus control. Given the number of studies contributing to these analyses, it is not surprising that none of the moderators were significant predictors of effect size. Removing Nickel, Luely et al. (2006) did not have any influence on the results.

Multivariate analysis. We do not report results for multivariate analyses for family therapy versus control because with so few studies the results were too unstable.

Figure 4. A forest plot of effect sizes and 95% confidence intervals for studies in the family therapy versus control meta-analysis.
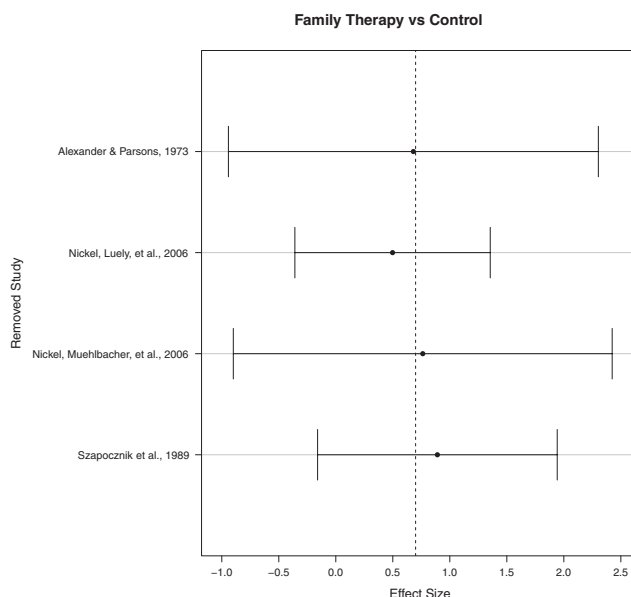
Figure 5. An influence plot from a "leave-one-out" analysis. The aggregate effect size from all studies is represented by a dashed line. The dots and confidence bars represent the aggregate effect size, and 95% confidence interval after the study on the *y*-axis was removed.

Publication Bias

Figure 6 presents the funnel plots for family therapy versus TAU, alternative therapy, and control. The left-hand column presents traditional funnel plots with pseudo-95% confidence intervals, and the right-hand column presents contour-enhanced funnel plots. In the contour-enhanced plots, the white area represents the 90% confidence and below area (low statistical significance), the light gray area represents the 90–95% confidence bands (moderate statistical significance), and the dark gray area represents the 95–99% confidence bands (high statistical significance). In the traditional funnel plots, most studies fall within the 95% confidence limits regardless of the particular comparison group, suggesting no publication bias. The contour-enhanced plots were not as consistent across comparison groups. In the family therapy versus TAU data and the family therapy versus alternative therapy data, most effect sizes are above zero and/or are in the low statistical significance area, which argues against publication bias. In the family therapy versus control data, all of the effect sizes are above zero and three of four are in the moderate or high statistical significance area, which may indicate some publication bias. However, as noted earlier, judgments of asymmetry are difficult when the number of studies is small.

Table 3 presents the results of the rank correlation test, regression test, trim-and-fill analysis, and selection bias models for all comparison types. Both the rank correlation test and the regression test were not statistically significant for any comparison, although the regression test was marginally significant ($p < .10$) in the family therapy versus alternative therapy analysis. The results of the trim-and-fill analysis varied by comparison type. In the family therapy versus TAU analysis, no additional studies were added and thus there was no adjustment to the aggregate effect size. In the family therapy versus alternative therapy, the trim-and-fill analysis suggested that two additional studies would need to be added to the data. These additional studies reduced $\bar{d}$ to .20 (a 23% reduction). In the family therapy versus control, the trim-and-fill analysis suggested that one additional study would need to be added to the data. This additional study reduced $\bar{d}$ to .50 (a 28% reduction). Given that the family therapy versus control analysis only involved four studies, one additional study has the potential to be highly influential so the large change should be interpreted with caution.

For the selection models, we used the four selection processes outlined in Vevea and Woods (2005)—one-tailed moderate, one-tailed severe, two-tailed moderate, and two-tailed
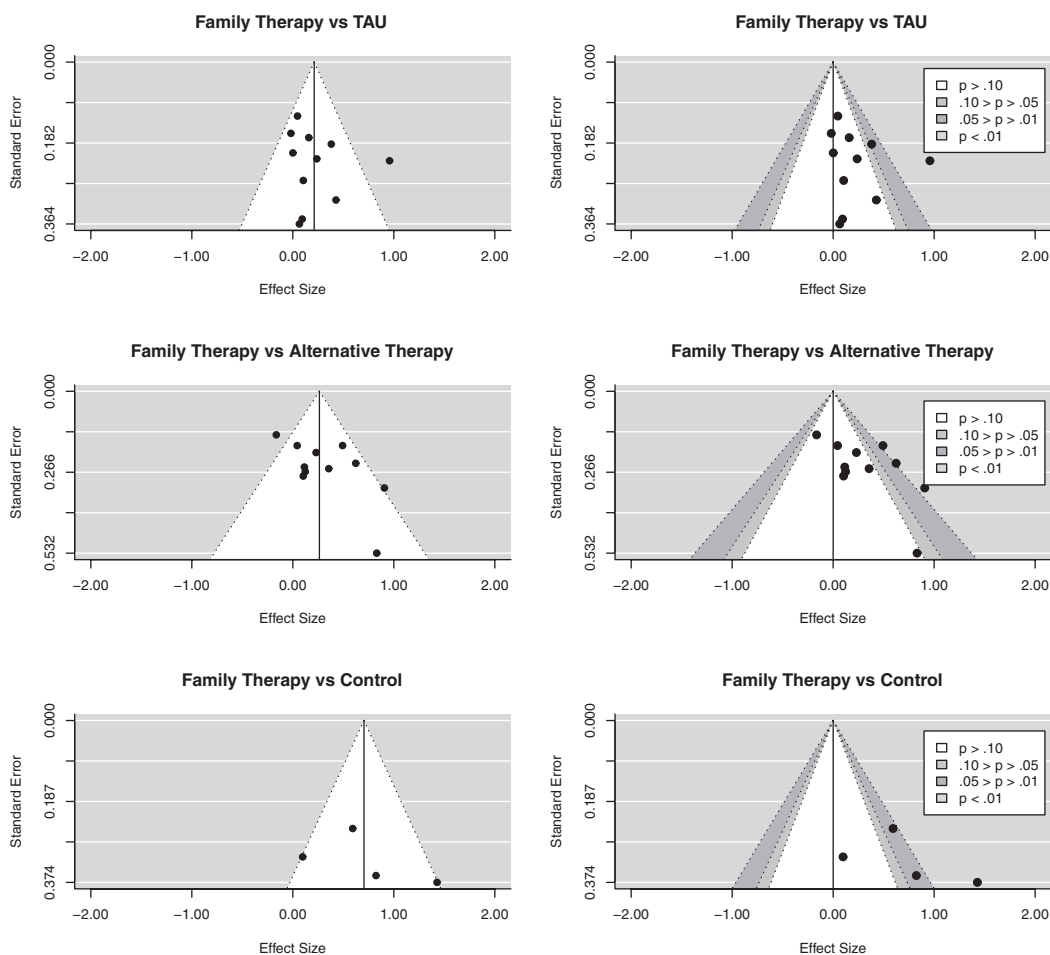
Figure 6. Funnel plots for all three comparison group types.

severe selection. As can be seen in Table 3, all four selection processes result in reduced effect sizes for each of three comparison types. Across comparison groups, the adjusted effect size was similar for one-tailed moderate, two-tailed moderate, and two-tailed severe selection. Further, the one-tailed severe selection process produced the biggest change in the effect size across comparison groups. In the family therapy versus TAU analysis, the change in the one-tailed severe model was implausibly large (the adjusted $\bar{d} = -0.60$). Finally, the adjusted effect sizes in the selection bias models were not consistently in agreement with adjusted effect sizes in the trim-and-fill analysis or were consistently above or below the adjusted effect sizes in the trim-and-fill analysis.

In sum, the publication bias analyses were inconsistent and did not present clear evidence of publication bias. Across analyses, the funnel plots appeared symmetric and the rank correlation and regression tests were not statistically significant. The adjusted effect sizes in both the trim-and-fill analysis and the selection bias models were smaller than the original effect sizes. However, in most cases, the change in effect size was not substantial and when it was substantial, the number of studies was small, which likely exaggerated the adjustment.

DISCUSSION

The results of this meta-analysis suggest that participants with delinquency or substance abuse problems receiving BSFT, FFT, MDFT, or MST fared better than participants receiving either TAU or an alternative therapy. Although these differences were statistically significant,

Table 3
Results of Publication Bias Analyses

Comparison condition	Rank correlation (z)	Regression test (bias)	Trim and fill		Selection bias ($\bar{d}'$)				$\bar{d}$
			Trimmed studies	$\bar{d}'$	One-tailed moderate	One-tailed severe	Two-tailed moderate	Two-tailed severe	
TAU	0.20	0.53	0	0.21	0.15	−0.60	0.18	0.15	0.21
Alternative treatment	0.30	2.12*	2	0.20	0.20	0.10	0.23	0.20	0.26
Control	0.67	0.94	1	0.50	0.64	0.57	0.66	0.62	0.70

Note. $\bar{d}$ = unadjusted aggregate effect size; $\bar{d}'$ = adjusted aggregate effect size. *p < .10.

they were relatively small ($d = 0.21$ for family therapy vs. TAU and $d = 0.26$ for family therapy versus alternative therapy). The difference between family therapy and control was $d = 0.70$ but was not statistically significant. We suspect that the lack of significance is owing to the fact that only four studies used a control condition. Consequently, we do not discuss these results further except to say more research of this type is needed. We limited our analyses to effect sizes calculated at the end of treatment, and future reviews should include effect sizes at follow-up timepoints. Finally, there was no consistent evidence that publication bias threatens the validity of these results.

The results for the family therapy versus TAU and alternative therapy are consistent with Waldron and Turner's (2008) findings regarding the effects of family therapies for adolescent substance abuse but larger than Littell et al.'s (2005) findings regarding MST's effectiveness for delinquency. The differences among the meta-analyses can likely be accounted for by the fact that our meta-analysis included studies published since the previous reviews, studies aimed at substance abuse and delinquency, and studies involving all four specific types of family therapy. Nevertheless, taken together, the current meta-analytic evidence suggests that across outcomes, family therapy has a modest added benefit beyond TAU and alternative treatments.

These general conclusions come with two important caveats brought to light by two of the advanced meta-analysis methods. First, this literature is not sufficiently large to reliably test moderator effects, probably owing to the difficulty and expense involved in conducting studies with intensive treatments and difficult populations (Hedges & Pigott, 2001, 2004). One important moderator was treatment type. Although differences among treatment types were present, no differences were statistically significant. Similar problems were present for all moderator analyses.

Power was also an issue in the multivariate meta-analyses. We compared effect sizes for primary outcomes (delinquency/substance abuse) to secondary outcomes (other measures) to see whether family therapies were most effective in dealing with the outcomes they were designed to treat. Family therapy had bigger effects on delinquency/substance abuse measures but none of the differences was significant. As the research literature grows, future meta-analyses will be better suited to explore these moderator questions in both univariate and multivariate models.

The second caveat regarding the general conclusions is that several studies had a large influence on aggregate effect sizes and estimates of between-study heterogeneity. The influential studies were Timmons-Mitchell et al. (2006) in the TAU analysis, Dennis et al. (2004) in the alternative therapy analysis, and Nickel, Luely et al. (2006) in the control analysis. Influential studies are important for a number of reasons. First, influential studies can distort aggregate effect sizes and estimates of heterogeneity, as we saw in the present meta-analysis. Second, if influential studies introduce additional heterogeneity, they can make it difficult to detect moderators. This occurred in the overall analysis when examining differences among comparison types because these differences were only significant when the influential studies were removed from the analysis. Third, understanding why a study has a particularly large (or small) effect size could provide insight into what makes treatment effective (or ineffective).

Interestingly, Leschied and Cunningham (2002) was not identified as an influential study, as has been argued previously (Henggeler et al., 2006). Henggeler et al. argued that the implementation of MST in Leschied and Cunningham (2002) was weak and that this should have been more carefully considered for inclusion in Littell et al.'s (2005) review. In our view, this is a question of generalization (Shadish, Cook, & Campbell, 2002). That is, what is the expected level of implementation in the setting to which we want to generalize? If the expected level of implementation is high, as is seen in studies of MST conducted by the developers of MST, then excluding studies with relatively poorer implementation is warranted. If such a high level of implementation is not realistic, then including studies with relatively poorer implementation is not only appropriate but also needed. We suspect that quality of implementation in typical clinical settings will not match the quality seen in studies conducted by the developers of MST. Thus, we retained Leschied and Cunningham (2002).

Comments on Meta-analysis

Meta-analysis is important to the evaluation of interventions because it affords researchers a level of precision unmatched by any single study. The forest plots highlight this critical feature of meta-analysis—namely, the borrowing of strength across studies. As can be seen in Figure 2, only two of the 11 studies in the family therapy versus TAU meta-analysis have an overall effect size that is statistically significant (i.e., nine of 11 confidence intervals overlap with zero). However, when the studies are combined, precision and thereby power are increased, and the overall effect size is statistically significant. A similar pattern is observed in family therapy versus alternative therapy meta-analysis. Thus, bringing multiple studies together increases the precision of our estimates.

Thankfully, this benefit as well as other benefits of meta-analysis is more accessible than ever. Coding of moderator variables and effect size can often be performed in Excel or other spreadsheet software. Our research team used database software to code moderator variables and the ES program (Shadish et al., 1999) to calculate effect sizes. We conducted all analyses, with the exception of the multivariate meta-analyses, and created all graphs in this paper with the open-source, free statistical software package R (R Development Core Team, 2009). All analyses and graphs in this paper could also be created with Stata (StataCorp, 2009), except for the selection bias models. An alternative to using multiple programs is the Comprehensive Meta-analysis software package (Comprehensive Meta-analysis, 2006), which integrates coding and analysis into a single program. As these software packages mature, more researchers will be able to use both basic and advanced meta-analysis methodologies to answer pressing questions in family therapy research.

CONCLUSIONS

Delinquency and substance abuse among adolescents are costly problems, both from a public health and economic standpoint. Family therapy—specifically BSFT, FFT, MDFT, and MST—appears to modestly exceed the effects of TAU and alternative therapies. Actually, the situation is likely a bit more complicated than that but the literature in this area is not yet sufficiently large to answer critical questions, such as is one treatment more effective than the others and on what outcomes do the family therapies have the biggest effect? Policy makers, researchers, clinicians, and patients all want answers to these more nuanced questions. This paper has given us a peek at the answers to these questions. As the literature grows, we can use meta-analysis and related techniques to gain a full view.

CLINICIAN RESPONSE: THE GOOD, BAD, AND GOOD NEWS

Roy A. Bean

In contrast to other articles in this issue, Baldwin, Christian, Berkeljon, and Shadish (in press) is a focused meta-analysis (not a narrative review) of the available studies on a particular topic (i.e., adolescent problem behaviors). As such, it examines the comparable viability of the four primary family-based treatments, rather than a detailed review of each approach. Therefore, my response will be focused on the models' overall benefits and limitations as they relate to clinical practice and training (graduate/postgraduate). As before, the treatment approaches in question are Brief Strategic Family Therapy (BSFT; Szapocznik et al., 2003), Functional Family Therapy (FFT; Alexander & Parsons, 1982), Multidimensional Family Therapy (MDFT; Liddle & Hogue, 2001), and Multisystemic Therapy (MST; Henggeler et al., 1998).

The Good News

The best news, based on these analyses, is that researchers have provided the consumer (whether that is the client family or the therapist-in-training) with a valuable and effective product. On average, families and their troubled adolescents get better when treated with one of these four approaches than if treated using treatment-as-usual (TAU, often an individual cogni-

tive-behavioral therapy) or alternative therapy such as group therapy or psychoeducation. The advantages over TAU and alternative therapies, although deemed statistically "modest," can be considered clinically significant because of associated cost savings to the juvenile justice system and other social/welfare services (Henggeler & Sheidow, 2003). Additionally, these findings provide reliable evidence for the value of family-based treatments over individual-only therapy approaches. In addition, when trained and supervised in relation to these models, clinicians are equipped with a set of skills and interventions that have been found to be effective in treating these difficult clinical issues. The mental health fields, the many families aided, and the assisting clinicians are all the richer for the effectiveness of these therapy approaches.

Furthermore, all the models have been tested, albeit to varying degrees, and found to be effective across various levels of delinquency severity and in relation to a number of specific behavior problems (e.g., sexual offenses, serious drug use, bullying). This is particularly relevant for the clinicians working in the "real world" where delinquent behaviors often occur comorbidly with substance use and adolescent clients rarely use one drug/substance only. It is also important to note that all the models, and MST and BFST particularly, have been examined for application to populations-of-color and some international samples so they can be viewed as generalizable beyond the white, European American majority.

The Bad News

As noted earlier, studies are limited in comparing the models side by side, so there is no clear answer to the empirically necessary and aficionado-incentivized question of which model works best or which model works best under a certain set of circumstances. Nor is there a clear answer to the question of how the models will perform when implemented outside the direct supervision of program developers. In fact, of the limited number of studies available for this meta-analysis, only Leschied and Cunningham (2002) examined therapy in "real-world conditions" without tightly controlled limits and it was an unpublished study. There is hope, however, that research addressing these concerns may be slowly accumulating, given recent attention to effectiveness (Henggeler & Sheidow; this issue), transportability (Schoenwald, 2010), and replications of model efficacy by researchers acting independently of model originators (e.g., Friedman, 1989; Ogden & Hagen, 2006; Timmons-Mitchell et al., 2006).

Nevertheless, the most significant limitation is that training in these models is not readily accessible for most practicing clinicians and interested trainees. Reasons are multi-faceted and complicated but are briefly considered here. First, as currently constructed, the models are not easily transportable to typical clinical settings. Training for individual clinicians is available only in case of MDFT, while training in the other models is oriented toward groups working for the same treatment setting. Based on personal inquiries and reviews of model-specific online materials and the SAMHSA website (http://www.nrepp.samhsa.gov/), *initial* training costs are estimated conservatively at between $4,500 (for MDFT, per individual) and $24,000 (for MST, for the entire team). Accordingly, very few practitioners are able to afford trainings, and most find it difficult to take time from their clinical practice to receive the necessary education and supervision. Universities and other training programs are similarly cash-strapped and/or are limited in meeting the inclusion criteria to become an approved training site. Treatment manuals are available directly from the researchers/developers; however, these models are not easily implemented and require additional training and oversight because of their family-based (and in some cases, extrafamilial) focus. In addition, for all models and for MDT and MST particularly, therapists are purposefully limited to small case loads where they work intensely and frequently with the adolescent, their family, and the larger system (e.g., juvenile court administrators, social service providers, school teachers and officials, neighborhood, and church leaders). This necessitates a contract for services because third-party insurance providers (at present) are not interested in reimbursing providers for allied services not involving direct therapeutic contact with the identified patient. Consequently, model trainings usually take place in the context of county, state, or federal government-funded agencies where training/supervision contracts have been established with specific programs. Unfortunately, these public-funded mental health service organizations are plagued by regular clinician turnover and are particularly vulnerable to funding cutbacks when economic downturns occur.

Second, as noted by numerous authors (e.g., Northey & Hodgson, 2008; Salloum, Sulkowski, Sirrine, & Storch, 2009), access to these and other empirically supported therapies is hampered by significant dissemination difficulties. Along with the limiting factors noted earlier, a primary and very justified hurdle to dissemination is the developer concern over treatment fidelity. To ensure clinical competence, the approaches devote considerable focus to training, supervision, and regular re-trainings so that participating interventionists continue to practice the model as described in the treatment manuals and other materials.

Third, as an emerging and unintended consequence of outcome-based competency testing, training programs currently have little incentive to train students in these approaches because the statistical majority of their graduating students will not work for agencies that utilize these modalities. Additionally, in high-turnover clinical settings, employers are hesitant to invest in expensive therapist trainings when they are unlikely to recoup their costs. This is unfortunate because for a field (in this case, marriage and family therapy) that is anxious for measureable outcomes and competency-based training markers, the level of specificity found in these models would be ideal if adopted by training programs.

Finally, there is the looming issue of privatization or private proprietorship of the treatment models. Three of the four models (MST, FFT, BSFT) have been privatized, and the trainings are now being conducted from a "for-profit" basis, with MDFT as the lone exception. This allows retention and ownership of intellectual property by program developers and helps facilitate treatment adherence and fidelity. However, the primary issue here is that millions of taxpayer dollars have been used directly and indirectly in the discovery, development, and refinement of these approaches, but the results of these investments are not readily accessible to clinicians and educators in the public arena. In essence, we have already purchased the treatment manuals, videotapes, and resources created for training MST, FFT, BSFT, and MDFT therapists but when we get home, we can only take a small portion of the materials out of the bag to use.

In the interest of full disclosure, I do not fault these programs for their decision. In privatizing, they maintain a larger measure of control over their "brainchild," which helps assure better quality practice and implementation. Privatization also provides a direct financial benefit to the researchers, and assumedly, some portions of the resultant funds are turned back into the purveyor organization to support further research efforts. Nevertheless, while hard work and intellectual property were exchanged for grant support—a fair and appropriate free-market transaction—the ultimate purpose of the grants was only partially met because the full product is not really available to clinicians and educators.

I imagine that proponents of privatization would argue that the complete approach (with all interventions, curriculum, and therapeutic nuances) is beyond the capability of most clinicians, except where they are trained and supervised for treatment adherence. I do not doubt this because good therapy is often complex and strenuous and difficult to provide reliably. This justifiable position notwithstanding, I argue that there must be more gains than risks involved in making the complete product available to clinicians (and their distressed families).

Surprisingly, there has been little or no published attention given to the topic of privatization—not only as it relates to the models discussed here—but to treatment approaches in general. This is disappointing and somewhat alarming; given that the issue has been in play in terms of delinquency treatments because the first approach was incorporated over 14 years ago (MST Services was licensed for MST dissemination in 1996). Furthermore, while I am not convinced that privatization is wrong, I do not view it as right either and I am unequivocally distressed by the fact that an open public discussion of this matter has been absent from our deliberations as mental health disciplines. At best, this suggests a series of administrative oversights on the part of federal funding agencies, which is a tolerable option only when compared to the other possibilities (i.e., complete incompetence or administrative wrongdoing). We can and should do better as a field in "watching the watchers" and inviting or demanding greater accountability in terms of the use of public funds for private product development.

The Good News, Again

Accompanying the bad news is a number of possibilities for continued development in the mental health fields for future generations of clinicians and researchers. Among the opportuni-

ties is the need for a detailed examination of the common factors associated with these four treatment models (e.g., Blow & Distelberg, 2006). These investigations, conducted by external reviewers, would help bridge the clinician–researcher gap and make the materials more accessible to therapists interested in treating this challenging population.

Another positive element is that we can plan on additional publications and further refinements to the approaches. All four models enjoy continuous research activity and expansion by therapy developers and colleagues. Of particular importance are the next generation of studies that will be focused on identifying the primary mechanisms responsible for individual- and family-level change. As these are isolated and shared in sufficient detail, it will be possible for individual therapists and educational programs to focus their training on the key skills and interventions that facilitate change. Concerns about privatization aside, there is hope that researchers will continue to be forthcoming about what makes their approaches work so that treatments can be more easily adapted to settings typical of mental health practice such as out-patient and day-treatments.

Finally, as the most resounding positive, there is no clear evidence to favor one program over another, so there is no wrong choice if an educational program or agency was to select a model for implementation. Consequently, this decision can be made on the basis of more practical marketplace realities such as cost or which model helps student trainees find jobs in the treatment facilities in the area.

In closing, let me suggest that most of us would exchange our problems for the problems that remain for these treatment approaches and their developers. They have been immensely successful and have made huge contributions to the literature and to the lives of countless adolescents and their families. Is there more work to be performed? Yes, however, it is only easier to note the areas where additional attention is needed because so much good, even great, work has already been performed. If there is anything overly negative in my response here, it is meant to be interpreted as constructively critical, derived from two major sentiments. First, I have to admit a genuine level of admiration for what these researchers have created, not only have they created an effective and empirically supported intervention program but many of them have found ways to successfully turn the programs into revenue-generating entities. Second, I want to ensure that quality students and quality clinicians everywhere (not just those affiliated with a particular treatment franchise or training site) are getting the opportunity to learn empirically supported treatments so that more families can be benefitted and more lives improved.

REFERENCES

References marked with an asterisk (*) are studies included in the meta-analysis.

*Alexander, J. F., & Parsons, B. V. (1973). Short-term behavioral intervention with delinquent families: Impact on family process and recidivism. *Journal of Abnormal Psychology*, *81*, 219–225.

Alexander, J. F., & Parsons, B. V. (1982). *Functional family therapy: Principles and procedures*. Carmel, CA: Brooks/Cole.

Alexander, J. F., Pugh, C., & Parsons, B. V. (2000). *Functional family therapy* (Vol. 3). Golden, CO: Venture.

APA Publications and Communications Board Working Group on Journal Article Reporting. (2008). Reporting standards for research in psychology. Why do we need them? What might they be? *American Psychologist*, *63*, 839–851.

Begg, C. D., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088–1101.

Blow, A. J., & Distelberg, B. (2006). *Common factors in four evidence-based family therapy approaches*. Workshop presented at the National Council on Family Relations Annual Conference. Minneapolis, MN.

*Borduin, C. M., Henggeler, S. W., Blaske, D. M., & Stein, R. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, *35*, 105–114.

*Borduin, C. M., Mann, B. J., Cone, L. T., Henggeler, S. W., Fucci, B. R., Blaske, D. M., et al. (1995). Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology*, *63*, 569–578.

*Borduin, C. M., Schaeffer, C. M., & Heiblum, N. (2009). A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: Effects on youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology*, *77*, 26–37. doi:10.1037/a0013035.

Borenstein, M. (2009). Effect size for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221–235). New York: Russell Sage Foundation.

*Brown, T. L., Henggeler, S. W., Schoenwald, S. K., Brondino, M. J., & Pickrel, S. G. (1999). Multisystemic treatment of substance abusing and dependent juvenile delinquents: Effects on school attendance at post-treatment and 6-month follow-up. *Children's Services: Social Policy, Research And Practice*, 2, 81–93.

Chorpita, B. F., Yim, L. M., Donkervoet, J. C., Arensdorf, A., Amundsen, M. J., McGee, C., et al. (2002). Toward large-scale implementation of empirically supported treatments for children: A review and observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology: Science and Practice*, 9, 165–190.

*Coatsworth, J. D., Santisteban, D. A., McBride, C. K., & Szapocznik, J. (2001). Brief strategic family therapy versus community control: Engagement, retention, and an exploration of the moderating role of adolescent symptom severity. *Family Process*, 40, 313–332.

Comprehensive Meta-analysis. (2006) *[Computer Program]*. Englewood, NJ: Biostat.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

*Dennis, M., Godley, S. H., Diamond, G., Tims, F. M., Babor, T., Donaldson, J., et al. (2004). The Cannabis Youth Treatment (CYT) Study: main findings from two randomized trials. *Journal of Substance Abuse Treatment*, 27, 197–213. doi:S0740-5472(04)00087-X [pii]10.1016/j.jsat.2003.09.005.

Duval, S., & Tweedie, R. (2000a). A nonparametric trim and fill method for accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–99.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel plot based method for testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463.

Egger, M., Smith, G. D., Schnieder, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, 315, 629–634.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

*Friedman, A. (1989). Family therapy vs. parent groups: Effects on adolescent drug abusers. *American Journal of Family Therapy*, 17, 335–347.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1–20.

Harbord, R. M., & Higgins, J. P. (2008). Meta-regression in Stata. *The Stata Journal*, 8, 493–519.

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

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203–217. doi:10.1037//1082-989X.6.3.203.

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–445. doi:10.1037/1082-989X.9.4.426.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 4, 486–504.

*Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 868–874.

*Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65, 821–833.

*Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology*, 60, 953–961.

*Henggeler, S. W., Melton, G. B., Smith, L. A., Schoenwald, S. K., & Hanley, J. H. (1993). Family preservation using multisystemic treatment: Long-term follow-up to a clinical trial with serious juvenile offenders. *Journal of child and family studies*, 2, 283–293.

*Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance abusing and dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research*, 1, 171–184.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford.

Henggeler, S. W., Schoenwald, S. K., Swenson, C. C., & Borduin, C. M. (2006). Methodological critique and meta-analysis as Trojan horse. *Children and Youth Services Review*, 28, 447–457.

Henggeler, S. W., & Sheidow, A. J. (2003). Conduct disorder and delinquency. *Journal of Marital and Family Therapy*, 29, 505–522.

Henggeler, S. W., & Sheidow, A. J. Empirically supported family-based treatments for conduct disorders and delinquency. *Journal of Marital and Family Therapy*, doi: 10.1111/j.1752-0606.2011.00244.x

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2008). *Monitoring the Future national results on adolescent drug use: Overview of key findings, 2007*. (NIH Publication No. 09-6418). Bethesda, MD: National Institute on Drug Abuse.

Kalaian, H., & Raudenbush, S. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Bulletin*, *1*, 227–235.

*Leschied, A. W., & Cunningham, A. (2002). *Seeking effective interventions for young offenders: Interim results of a four-year randomized study of multisystemic therapy in Ontario, Canada*. London, Ontario: Centre for Children and Families in the Justice System. Unpublished manuscript.

*Letourneau, E. J., Henggeler, S. W., Borduin, C. M., Schewe, P. A., McCart, M. R., Chapman, J. E., et al. (2009). Multisystemic therapy for juvenile sexual offenders: 1-year results from a randomized effectiveness trial. *Journal of Family Psychology*, *23*, 89–102. doi:10.1037/a0014352.

Letourneau, E. J., Resnick, H. S., Kilpatrick, D. G., Saunders, B. E., & Best, C. L. (1996). Comorbidity of sexual problems and posttrumatic stress disorder in female crime victims. *Behavior Therapy*, *27*, 321–336.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*, b2700.

*Liddle, H. A., Dakof, G. A., Parker, K., Diamond, G., Barrett, K., & Tejeda, M. (2001). Multidimensional family therapy for adolescent drug abuse: Results of a randomized clinical trial. *American Journal of Drug and Alcohol Abuse*, *27*, 651–688.

*Liddle, H. A., Dakof, G. A., Turner, R. M., Henderson, C. E., & Greenbaum, P. E. (2008). Treating adolescent drug abuse: A randomized trial comparing multidimensional family therapy and cognitive behavior therapy. *Addiction*, *103*, 1660–1670.

Liddle, H. A., & Hogue, A. (2001). Multidimensional family therapy for adolescent substance abuse. In E. F. Wagner & H. B. Waldron (Eds.), *Innovations in adolescent substance abuse interventions* (pp. 229–261). Kidlington, Oxford: Elsevier.

*Liddle, H. A., Rowe, C., Dakof, G. A., Henderson, C. E., & Greenbaum, P. E. (2009). Multidimensional family therapy for young adolescent substance abuse: Twelve-month outcomes of a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *77*, 12–25.

*Liddle, H. A., Rowe, C., Dakof, G. A., Ungaro, R. A., & Henderson, C. E. (2004). Early intervention for adolescent substance abuse: Pretreatment to posttreatment outcome of a randomized clinical trial comparing multidimensional family therapy and peer group treatment. *Journal of Psychoactive Drugs*, *36*, 49–63.

Littell, J. H. (2006). The case for multisystemic therapy: Evidence or orthodoxy? *Children and Youth Services Review*, *28*, 458–472.

Littell, J. H., Popa, M., & Forsythe, B. (2005). Multisystemic Therapy for social, emotion, and behavior problems in youth age 10–17. *Cochrane Database of Systematic Reviews*, (3), CD004797.

Nam, I. S., Mengersen, K., & Garthwaite, P. (2003). Multivariate meta-analysis. *Statistics in Medicine*, *22*, 2309–2333. doi:10.1002/sim.1410.

*Nickel, M., Luley, J., Krawczyk, J., Nickel, C., Widermann, C., Lahmann, C., et al. (2006). Bullying girls—Changes after brief strategic family therapy: A randomized, prospective, controlled trial with one-year follow-up. *Psychotherapy and Psychosomatics*, *75*, 47–55. doi:PPS2006075001047 [pii]10.1159/000089226.

*Nickel, M., Muehlbacher, M., Kaplan, P., Krawczyk, J., Buschmann, W., Kettler, C., et al. (2006). Influence of family therapy on bullying behavior, cortisol secretion, anger, and quality of life in bullying male adolescents: A randomized, prospective, controlled study. *Canadian Journal of Psychiatry*, *51*, 355–362.

Northey, W. F. Jr, & Hodgson, J. (2008). Empirically supported therapies: Training opportunities and evidence of success. *Journal of Family Psychotherapy*, *19*, 50–84.

Ogden, T., & Hagen, K. A. (2006). Multisystemic therapy of serious behaviour problems in youth: Sustainability of therapy effectiveness two years after intake. *Journal of Child and Adolescent Mental Health*, *11*, 142–149.

*Ogden, T., & Halliday-Boykins, C. A. (2004). Multisystemic treatment of antisocial adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and Adolescent Mental Health Volume*, *9*, 77–83.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Ruston, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*, 991–996.

Post, L. A., Mezey, N. J., Maxwell, C., & Wibert, W. N. (2002). The rape tax: Tangible and intangible costs of sexual violence. *Journal of Interpersonal Violence*, *17*, 773–782.

R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved February 12, 2010, from http://www.r-project.com.

Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, *103*, 111–120.

Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J., & Thompson, J. R. (2007). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine*, 26, 78–97. doi:10.1002/sim.2524.

Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C., & Thompson, J. R. (2007). Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology*, 7, 3. doi:10.1186/1471-2288-7-3.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 85, 185–193.

*Rowland, M. D., Halliday-Boykins, C. A., Henggeler, S. W., Cunningham, P. B., Lee, T. G., Kruesi, J. P., et al. (2005). A randomized trial of multisystemic therapy with Hawaii's felix class youths. *Journal of Emotional and Behaviorial Disorders*, 13, 13–23.

Salloum, A., Sulkowski, M. L., Sirrine, E., & Storch, E. A. (2009). Overcoming barriers to using empirically supported therapies to treat childhood anxiety disorders in social work practice. *Child and Adolescent Social Work Journal*, 26, 259–273.

*Santisteban, D. A., Coatsworth, J. D., Perez-Vidal, A., Kurtines, W. M., Schwartz, S. J., LaPerriere, A., et al. (2003). Efficacy of brief strategic family therapy in modifying hispanic adolescent behavior problems and substance use. *Journal of Family Psychology*, 17, 121–133.

*Schaeffer, C. M., & Borduin, C. M. (2005). Long-term follow-up to a randomized clinical trial of multisystemic therapy with serious and violent juvenile offenders. *Journal of Consulting and Clinical Psychology*, 73, 445–453.

Schoenwald, S. K. (2010). From policy pinball to purposeful partnership: The policy contexts of multisystemic therapy transport and dissemination. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed.) (pp. 538–553). New York: Guilford.

Shadish, W. R., & Baldwin, S. A. (2003). Meta-analysis of MFT interventions. *Journal of Marital and Family Therapy*, 29, 547–570.

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

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*. Minneapolis, MN: Assessment Systems.

StataCorp. (2009). *Stata statistical software: Release 11*. College Station, TX: StataCorp LP.

Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, 4, 127–141.

*Sundell, K., Hansson, K., Lofholm, C. A., Olsson, T., Gustle, L.-H., & Kadesjo, C. (2008). The transportability of multisystemic therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology*, 22, 550–560.

Szapocznik, J., Hervis, O., & Schwartz, S. (2003). *Brief Strategic Family Therapy for adolescent drug abuse*. (NIH Pub. No. 03-4751). Bethesda, MD: National Institute on Drug Abuse.

*Szapocznik, J., Rio, A., Murray, E., Cohen, R., Scopetta, M., Rivas-Vazquez, A., et al. (1989). Structural family versus psychodynamic child therapy for problematic Hispanic boys. *Journal of Consulting and Clinical Psychology*, 57, 571–578.

Szapocznik, J., & Williams, R. A. (2000). Brief Strategic Family Therapy: Twenty-five years of interplay among theory, research and practice in adolescent behavior problems and drug abuse. *Clinical Child and Family Psychology Review*, 3, 117–134.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publicating bias in the presence of heterogeneity. *Statistics in Medicine*, 22, 2113–2126.

*Timmons-Mitchell, J., Bender, M. B., Kishna, M. A., & Mitchell, C. C. (2006). An independent effectiveness trial of multisystemic therapy with juvenile justice youth. *Journal of Clinical Child and Adolescent Psychology*, 35, 227–236. doi:10.1207/s15374424jccp3502_6

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60, 419–435.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, 10, 428–443. doi:10.1037/1082-989X.10.4.428.

Viechtbauer, W. (2010). Meta-analysis package for R (Version 1.0-1) [Computer software]. Retrieved February 12, 2010, from http://www.r-project.org/.

*Waldron, H. B., Slesnick, N., Brody, J. L., Turner, C. W., & Peterson, T. R. (2001). Treatment outcome from adolescent substance abuse at 4- and 7- month assessments. *Journal of Consulting and Clinical Psychology*, 69, 802–813.

Waldron, H. B., & Turner, C. W. (2008). Evidence-based psychosocial treatments for adolescent substance abuse. *Journal of Clinical Child and Adolescent Psychology*, 37, 238–261.

White, I. R. (2009). Multivariate random-effects meta-analysis. *The Stata Journal*, 9, 40–56.