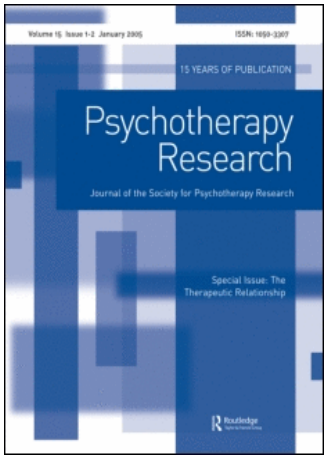

Psychotherapy Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663589

Direct comparisons of treatment modalities for youth disorders: a meta-analysis

Scott Miller [a]; Bruce Wampold [b]; Katelyn Varhely
[a] Institute for the Study of Therapeutic Change,
[b] University of Wisconsin - Madison,

Online Publication Date: 01 January 2008
To cite this Article: Miller, Scott, Wampold, Bruce and Varhely, Katelyn (2008) 'Direct comparisons of treatment modalities for youth disorders: a meta-analysis', Psychotherapy Research, 18:1, 5 - 14
To link to this article: DOI: 10.1080/10503300701472131
URL: http://dx.doi.org/10.1080/10503300701472131

Psychotherapy Research, January 2008; 18(1): 5–14

Direct comparisons of treatment modalities for youth disorders: a meta-analysis

SCOTT MILLER[1], BRUCE WAMPOLD[2], & KATELYN VARHELY[3]

[1]*Institute for the Study of Therapeutic Change*, [2]*University of Wisconsin—Madison and* [3]*Chicago, Illinois*

(Received 7 November 2006; revised 21 May 2007; accepted 23 May 2007)

Abstract
A meta-analysis was conducted to determine whether differences in efficacy exist among treatment approaches applied to youth. Included were all studies published between 1980 and 2005 involving participants 18 years of age or younger with diagnoses of depression, anxiety, conduct disorder, and attention-deficit/hyperactivity disorder that contained direct comparisons among two or more treatment methods intended to be therapeutic. Effect sizes were found to vary significantly, providing some evidence that differences in efficacy exist among treatments for these disorders in youth. However, the upper bound of the true difference in effects among treatments was small. Furthermore, researcher allegiance was found to be strongly associated with the difference in effect sizes so that when allegiance was controlled there was no evidence of any differences among treatments.

A number of studies and scholarly reviews published over the last two decades have provided strong empirical support for the general efficacy of psychological treatments as applied to youth (Kazdin, 2000, 2004; Weisz & Weiss, 1993), with effect sizes that are largely equivalent to those reported in the adult literature (Kazdin, 2004; Weisz, Weiss, Han, Granger, & Morton, 1995). At the same time, however, the question of differential efficacy continues to be debated. For example, in a meta-analysis of 108 "well-designed" outcome studies, Weisz, Weiss, Alicke, and Klotz (1987) found behavioral interventions to be associated with significantly larger effect sizes than nonbehavioral interventions, whereas Casey and Berman (1985) found no such difference in effect once a confound between type of treatment and outcome measures used was corrected. In a meta-analysis of treatments for depression in children, Weisz, McCarty, and Valeri (2006) found that, although various treatments were more effective than no treatment (but with smaller effects than the adult literature), no difference in outcome was found between cognitive and noncognitive approaches. Importantly, none of the approaches currently listed by the American Academy of Child and Adolescent Psychiatry (1997, 1998) Task Force on the Promotion and Dissemination of Psychological Procedures (1995) have been shown to be demonstrably superior to other treatments intended to be therapeutic for the disorder treated.

Wampold, Mondin, Moody, Stich, et al. (1997) and Shadish and Sweeney (1991) have argued that determining whether one approach is superior to another is most validly tested when two or more therapies intended to be therapeutic are compared within the same study. Analyses of the adult literature limited to direct comparisons of bona fide treatments have provided strong support for the equivalence of outcome, what has long been referred to as the "dodo bird verdict" (Wampold, Mondin, Moody, & Ahn, 1997; Wampold, Mondin, Moody, Stich, et al., 1997). The only meta-analysis published to date in the youth literature of studies involving direct comparisons produced largely similar results. Spielmans, Pasek, and McFall (in press) found no difference in outcome between cognitive and noncognitive approaches for the treatment of anxiety and depression in children when directly compared.

One factor not considered in any prior meta-analysis of the youth literature is the impact of researcher allegiance on outcome. In 1992, Shirk and Russell argued that allegiance effects and other methodological confounds might account for the superiority of behavioral approaches reported by Weisz et al. (1987), although Weiss and Weisz

Correspondence: Scott Miller. E-mail: Scottdmiller@talkingcure.com

ISSN 1050-3307 print/ISSN 1468-4381 online
DOI: 10.1080/10503300701472131

(1995) claimed that no empirical support existed for these conjectures. There is considerable evidence in the adult psychotherapy literature that a researcher's belief in or commitment to a particular method of treatment has a considerable influence on treatment outcome (Berman, Miller, & Massman, 1985; Devilly, 2001; Dush, Hirt, & Schroeder, 1983; Hoag & Burlingame, 1997; Lambert, 1999; Luborsky et al., 1999, 2002; Paley & Shapiro, 2002; Robinson, Berman, & Neimeyer, 1990; Shapiro & Shapiro, 1982; Smith, Glass, & Miller, 1980). Allegiance is often characterized by the researcher's development of the treatment or commitment to and identification with a particular theoretical approach and involves, for example, the researcher's training and supervising of the therapists in the study. Allegiance effects appear to be greater than the effects produced by comparisons of treatments, as much as three times greater when the most liberal estimates are used (Wampold, 2001).

The purpose of the present study was to determine whether differences in effectiveness exist among treatment approaches applied to the youth by (a) conducting a meta-analysis of studies involving direct comparisons of bona fide psychological therapies and (b) examining the effects of researcher allegiance and whether allegiance explains any difference between treatments that might be found. Studies involving either children or adolescents or both were combined in the meta-analysis for several reasons: (a) no generally agreed-on or reliable demarcation exists in the literature between childhood and adolescence; (b) many studies either failed to distinguish between children and adolescents or applied different definitions; (c) there were insufficient numbers of studies that definitively treated one age group or the other to make reliable tests of the hypotheses of this study; (d) and, most importantly, the design of the present study was aimed to provide an omnibus test for differences in effectiveness among bona fide treatments rather than first examining subpopulations and specific diagnoses.

The hypotheses of the present meta-analysis were that the true effect sizes for all comparisons between treatments would not vary significantly from zero (i.e., would be homogeneously distributed about zero) and that, if there were treatment effects, researcher allegiance would account for the variation among the effect sizes. If the primary hypotheses were not supported (e.g., the effects sizes were not homogenously distributed about zero after accounting for allegiance), post hoc analyses would be conducted to explore the potential impact of other variables (e.g., age, gender, diagnosis, severity, treatment type, outcome measure used) on outcome variability.

One can examine the null hypothesis that there are no differences among treatments in two ways, each of which provides information on the relative effects of treatments. First, one can examine the relative efficacy of treatments for particular disorders. Typically, however, there are an insufficient number of studies in which comparisons among bona fide treatments for particular disorders are made in order to test the hypothesis. Second, one can examine relative efficacy of treatments across all disorders, which gives an omnibus test of the relative efficacy issue; this is valuable but could lead to a conclusion that may not apply to every disorder. We chose an intermediate strategy and limited this study to the treatment of the most prevalent disorders of children and adolescents: depression, anxiety, conduct disorders, and attention-deficit disorder (Kazdin, 2004).

The design, methods, and procedures used in the current study were modeled closely after those of Wampold, Mondin, Moody, Stich, et al. (1997). As in that study, the research synthesized in the present meta-analysis was limited to studies that directly compared two or more treatments, which were fully intended to be therapeutic. Treatments were not classified into categories of treatments (e.g., behavioral, cognitive–behavioral, problem solving) because such a strategy results in comparisons of categories rather than treatments, which results in multiple statistical tests, and because classification rules are notoriously ambiguous (see Wampold, 2001).

Method

Study Selection

To be included in this meta-analysis, studies had to (a) appear in a professional, peer-reviewed English language journal between January 1980 and January 2005; (b) provide statistical data sufficient for calculating effect sizes (e.g., means, standard deviations, and samples sizes of comparison groups); (c) directly compare at least two bona fide psychological treatments (excluding pharmacologic, educational, and prevention approaches); (d) use a treatment manual or equivalently clear guidelines for conducting each treatment approach; (e) include participants who were 18 years of age or younger and who were being treated for depression, anxiety, conduct, or attention-deficit disorders; (f) use qualified mental health professionals (psychologists, social workers, and therapists-in-training) to administer treatments; and (g) deliver treatment in face-to-face sessions.

The present study used the criteria outlined by Wampold, Mondin, Moody, Stich, et al. (1997) to classify treatments as bona fide. As such, in addition to conditions d and e just noted, the treatments contained in the study had to meet at least one of the following two conditions: (a) contain a citation to an established therapeutic approach (e.g., a reference to Rogers's, 1951, client-centered therapy) or (b) specify the active ingredients of the treatments being compared. Comparisons with treatments designed to control for common or nonspecific factors (e.g., placebo control groups, nonspecific therapies, treatment-as-usual conditions, or alternative therapies) were excluded as were those that merely added a common, nonspecific, educational, or prevention approach to the bona fide therapy under study. Finally, in keeping with the goal of the present study to test the differential efficacy of two or more treatments, no dismantling studies or studies testing different doses of a single treatment were included.

Considerable time and effort were devoted to finding all studies meeting inclusion criteria in the research literature. First, books and chapters summarizing research on the treatment of children and adolescents were consulted (Christophersen & Mortweet, 2001; Fonagy, Target, Cottrell, Phillips, & Kurtz, 2002; Kazdin & Weisz, 2003; Lambert, 2004; Roth & Fonagy, 2005). From these sources, potentially relevant research and meta-analytic studies were identified and the references contained in each reviewed. Based on this review, the journals *Behavior Therapy, Journal of Consulting and Clinical Psychology, Journal of Counseling Psychology, Psychological Bulletin,* and *Archives of General Psychiatry* were determined to most likely include studies involving direct comparisons, and each issue was hand-searched. Following these efforts, a list of descriptive terms was generated and used to conduct an electronic search of the literature. The terms used in the search included *depression and meta-analysis and children*; *depression and meta-analysis and adolescents*; *anxiety and meta-analysis and children*; *anxiety and meta-analysis and adolescents*; *conduct disorder and meta-analysis and children*; *conduct disorder and meta-analysis and adolescents*; *ADHD and meta-analysis and children*; *ADHD depression and meta-analysis and adolescents*; *depression and clinical trials*; *anxiety and clinical trials*; *conduct disorder and clinical trials*; *ADHD and clinical trials*; *phobia and clinical trials*; *depression and empirically validated/supported*; *anxiety and empirically validated/supported*; *conduct disorder and empirically validated/supported*; *phobia and empirically validated/supported*; *ADHD and empirically validated/supported*; *depression and randomized clinical trials*; *anxiety and randomized clinical trials*; *conduct disorder and randomized clinical trials*; *ADHD and randomized clinical trials*; *phobia and randomized clinical trials*; *depression and controlled children/adolescents*; *anxiety and controlled children/ adolescents*; *conduct disorder and controlled children/adolescents*; *ADHD and controlled children/adolescents*; *phobia and controlled children/adolescents*; *depression and meta-analysis*; *anxiety and meta-analysis*; *conduct disorder and meta-analysis*; *ADHD and meta-analysis*; and *phobia and meta-analysis.* Finally, meta-analyses were examined to find primary studies that compared treatments for these disorders.

Katelyn Varhely retrieved all studies identified during the search process. Scott Miller and Katelyn Varhely then read and independently rated each of the studies according to the criteria outlined previously based solely on the descriptions in the Method section (i.e., all other parts of the manuscript were masked). If both agreed that two or more treatments in a given study were bona fide, the study was retained. Studies were rejected when both raters agreed that at least two of the treatments were not bona fide. In the four instances in which the two raters disagreed, Bruce Wampold settled the tie (unaware of the ratings of the first two raters). Thus, to be included, two of three authors had to agree independently that the treatments being compared were bona fide therapeutic approaches. Although more than 1,000 studies were reviewed and rated, only 23 met the stringent inclusion criteria (Table I).

Meta-Analysis Method

Effect size calculation. To test the hypothesis that the true effect size for all comparisons between treatments does not vary significantly from zero, it was necessary to compute an estimate of the overall effect size for each study to be included in the meta-analysis. Determining the effect size for a given study was accomplished by first calculating the difference between the means of treatments for each dependent variable and then dividing that value by the pooled standard deviation of the two treatments. Subsequently, the resulting value was adjusted to yield an unbiased estimate of the population effect size (Hedges & Olkin, 1985). Finally, the unbiased estimates were aggregated across all dependent measures within a given study, resulting in a single estimate of the effect size (d_i) for each study (i) as well as the standard error of that estimate s(di), assuming the correlation among the dependent variables was .50 (see Hedges & Olkin, 1985; Wampold, Mondin, Moody, Stich, et al., 1997). A criticism of past meta-analyses has raised an issue with regard to aggregating measures of symptoms of the disorder (i.e., targeted measures) with other

Table I. Studies Included in the Meta-Analyses.

Study	*N*	Age group	Disorder	Comparison
Brent et al. (1997)	65	13–18 years	Depression	Systemic behavioral family vs. individual cognitive–behavioral
Butler et al. (1980)	28	5th–6th grade	Depression	Role play vs. cognitive restructuring
Denkowski & Denkowski (1984)	28	3rd–5th grade	Hyperactivity	Group relaxation vs. EMG biofeedback
Dubey et al. (1983)	30	6–10 years	Hyperactivity	Behavioral modification group vs. parent effectiveness training
Horn et al. (1987)	12	7–11 years	ADHD	Behavioral parent training vs. cognitive–behavioral self-control
Hughes & Wilson (1988)	6	6–15 years	Behavioral	Contingency management-child present vs. communication skills training-child present
	8			Contingency management-child absent vs. communication skills training-child absent
	7			Contingency management-child present vs. communication skills training-child absent
	7			Contingency management-child absent vs. communication skills training-child present
Kahn & Kehle (1990)	34	10–14 years	Depression	CBT vs. relaxation
	34			CBT vs. self-modeling
	34			Relaxation vs. self-modeling
Kazdin (1987)	31	7–13 years	Antisocial	Problem-solving skills training vs. relationship therapy
Kazdin et al. (1989)	75	3–8 years	Antisocial	Problem-solving skills training vs. problem-solving skills training- practice
	75			Problem-solving skills training- practice vs. relationship therapy
	74			Problem-solving skills training vs. relationship therapy
Kazdin et al. (1992)	60	7–13 years	Antisocial	Problem-solving skills training vs. parent management training
	66			Problem-solving skills training vs. problem-solving skills training+parent management training
	67			Parent management training vs. problem-solving skills training+parent management training
Leal et al. (1981)	20	10th grade	Anxiety	Cognitive modification vs. systemic desensitization
Luk et al. (1998)	19	Elem. school	Conduct	Modified CBT vs. conjoint family
Murnis et al. (1998)	18	8–17 years	Phobia	EMDR vs. exposure
Reynolds & Coats (1986)	14	High school	Depression	CBT vs. relaxation
Rossello & Bernal (1999)	36	13–18 years	Depression	CBT vs. interpersonal
Schneider (1991)	41	7–13 years	Aggressiveness	Skill building vs. desensitization
Silverman et al. (1999)	65	6–16 years	Phobia	Contingency management vs. cognitive self-control and educational support
Sonuga-Barke et al. (2001)	58	3 years	ADHD	Parent training vs. parent counseling and support
Spence et al. (2000)	36	7–14 years	Social phobia	Child focused CBT vs. CBT+ parent involvement
Stark et al. (1987)	19	5–13 years	Depression	Self-control vs. behavioral problem solving
Szapocznik et al. (1989)	52	6–12 years	Behavioral/ emotional	Structural family vs. psychodynamic child
Webster-Stratton & Hammond (1997)	70	4–8 years	Conduct	Parent training vs. child training
	91			Combined group vs. child training
	75			Parent training vs. combined group
Webster-Stratton et al.	97	3–8 years	Conduct problems	Individual videotape vs. groupvideotape
	96			Individual video tape vs. group discussion
	95			Group videotape vs. group discussion

Note. EMG = electromyography; CBT = cognitive–behavioral therapy; ADHD = attention-deficit/hyperactivity disorder; EMDR = eye movement desensitization and reprocessing.

measures of psychological functioning (e.g., comorbid symptoms or well-being; i.e., nontargeted measures; e.g., Crits-Christoph, 1997). Because (a) outcome measures, whether targeted or not, typically are highly correlated (e.g., anxiety and depression; Tanaka-Matsumi & Kameoka, 1986), (b) authors interpret nontargeted measures as indicators of the superiority of a particular treatment, (c) not infrequently is it difficult to classify measures as targeted (e.g., they are not so classified in the primary studies typically), (d) nontargeted measures often assess important aspects of the patients' lives, such as well-being and role functioning, and (e) past meta-analyses have found no differences between targeted and nontargeted variables (Wampold, Mondin, Moody, & Ahn, 1997), all outcomes variables used by the primary authors to examine relative efficacy were aggregated.

Of the 23 studies meeting inclusion criteria, six contained more than two bona fide therapies, thereby creating more than one comparison within the same study. As such, a study that contained three bona fide treatments (e.g., Treatments A, B, and C) generated three comparisons (viz., A vs. B, A vs. C, and B vs. C at termination). Next, the estimates of the effect sizes (d_i) for all studies (i) were aggregated by weighting the effect size d_i of each study i, in standard fashion, by the inverse of its variance to yield an estimate of the aggregate effect size of all comparisons d_+ (Hedges & Olkin, 1985).

Measurement of allegiance. The second hypothesis in this meta-analysis was that the allegiance of the researcher would account, to a significant degree, in the variation of effect sizes produced by the comparisons between treatments. To test this hypothesis, the allegiance for each treatment was rated on a scale of 0 (*no evidence of allegiance to treatment demonstrated*) to 4 (*treatment developed by one of the authors and author trained and/or supervised the therapists*[1]) by two raters. To eliminate any potential for bias, the raters were unaware of the results of the studies being rated (only the introduction and Method sections were provided) and had no information about or connection to the meta-analysis. If the ratings differed by one point, the two raters discussed and reached agreement about the rating to be made. If the raters differed by more than one point, the ratings were declared a "disagreement" and the average of the ratings was used. In the present case, no disagreements were obtained (i.e., the agreement was 100%).

For the comparison of Treatments A and B, the allegiance rating for Treatment B was subtracted from the allegiance rating for Treatment A. The association of allegiance and effects size was assessed; it was expected that the greater the magnitude of the allegiance to one treatment over another, the greater would be the effect size for the favored treatment vis-à-vis the less favored treatment. As well, the effect of allegiance on heterogeneity was assessed.

Test of effect sizes. As noted by Wampold, Mondin, Moody, Stich, et al. (1997), when comparing two treatments of equal standing, the arithmetic sign of the effect is ambiguous as the treatment designated as first is arbitrary (i.e., should the mean of Treatment A be subtracted from the mean of Treatment B or vice versa?). Two strategies are used here following the suggestions of Wampold, Mondin, Moody, Stich, et al. (1997). First, the sign of the effect is randomly assigned, providing an aggregated effect size of zero. The test of whether there are differences among bona fide treatments is achieved by examining the homogeneity of the effects about the estimate of zero. If there are differences among treatments, there will be relatively many large effects resulting in a rejection of the null hypothesis of homogeneity. On the other hand, if there are no true differences among treatments, the distribution of effects about zero will be homogeneous (i.e., the number of effects deviating from zero is what would be expected by chance). Thus, the homogeneity of effects (with random signs) around zero is the primary test of the hypothesis that there are not differences among treatments (see Wampold, Mondin, Moody, Stich, et al., 1997).

A secondary way to determine the sign is to take the absolute value of the effect for each study, the goal of which is solely to provide an estimate of the upper bound of the true difference. If the true difference among treatments was zero, then, because of sampling error, some obtained effects would be positive and some negative. Thus, taking the absolute value of each effect and averaging will necessarily yield a positive aggregate effect size when the null hypothesis of no differences is true. Nevertheless, it provides an estimate of the upper bound of the true effect. Occasionally, this upper bound is erroneously interpreted as the estimate of the differences among treatments (e.g., Howard, Krause, Saunders, & Kopta, 1997); it should be realized that this is simply the upper bound and clearly an overestimate of true differences.

Statistical analysis. For the primary analysis involving the effects with random signs, we assumed that the studies in this meta-analysis were sampled from a population of studies and consequently a random-effects model was used (Hedges & Olkin, 1985). The analysis was conducted using a multilevel model

where variances are known (Raudenbush & Bryk, 2001, Chapter 7) using HLM6 (Raudenbush, Bryk, & Congdon, 2004). The first model was an unconditional model (i.e., not conditioned on study level variable, in this case allegiance). At Level 1,

$$d_j = \delta_j + e_{j,}$$

where d_j is the estimate of the effect size for study j, δ_j is the true effect for study j, and the variance of the errors e_j are known. At Level 2,

$$\delta_j = \gamma_o + u_{j,}$$

where γ_o is the grand mean of the effects and u_j is the Level 2 error. When random signs are assigned to effect sizes, the grand mean γ_o will be close to zero; if there are no true differences among treatments, the variance of u_j will be small (i.e., will not be significantly greater than would be expected by chance and thus the effects are distributed homogenously about zero). Homogeneity is tested with the statistic H, which indexes the deviations of the sampled effects from the grand mean, weighted by the inverse of the variance (Hedges & Olkin, 1985; Raudenbush & Bryk, 2001). H has a chi-square distribution with k–1 degrees of freedom, where k is the number of studies aggregated.

The effects of allegiance were examined by a model conditioned on allegiance, and thus the Level 2 equation becomes

$$\delta_j = \gamma_o + \gamma_1(\text{allegiance}) + u_{j,}$$

where γ_o is the expected effect for a study with equal allegiances to the two treatments and γ_1 is the expected difference in effect size between two studies whose allegiance differs by one point. If allegiance accounts for the effects detected, the fixed effect γ_1 will be significantly greater than zero and the variance of the Level 2 error u_j will be reduced.

For the absolute value of the effect sizes, which is an upper bound of the true effect sizes, using a random-effects analysis the grand mean was estimated using the unconditional model.

Results

The results of the unconditional and conditional models using random signs of 23 studies containing 1,060 participants (mean per treatment = 23; median = 31) and 36 effects is found in Table II. As expected, the grand mean was close to zero. In the unconditional model, the variance of the true effect sizes was .037, yielding $H = 56.11$, which, compared with a chi-square distribution with 35 degrees of freedom, was sufficiently large to reject the null hypothesis that the effects were homogenously distributed around zero ($p = .013$). Thus, some differences between treatments were larger than would be expected if all the treatments compared were equally effective. At first glance, it appears that, in contrast to the adult literature (cf. Wampold, Mondin, Moody, Stich, et al., 1997), there is some evidence that there are true differences among the effects produced by bona fide treatments of youth disorders, because the effects are more widely distributed about zero than is expected by chance. However, the mean of the absolute value of the effects, which is the upper bound of the true effects, was .22, which is small and similar in size to that found for adults (cf. Wampold, Mondin, Moody, Stich, et al., 1997).

Table II. Tests of Homogeneity for Unconditional and Allegiance Conditioned Models (Random Signs).

Fixed effect	Coefficient	*SE*	*t*	*p*	
	Unconditional model				
Grand mean λ_o	−.006	.0533	−.119	.907	
	Model conditioned on allegiance				
Intercept λ_o	−.013	.038	−.344	.733	
Allegiance λ_1	.131	.026	4.99	.000	
Random effect	**Variance component**	***df***	***H***	***p***	I^2
	Unconditional model				
True effect size δ_j	.037	35	56.11	.013	.38
	Model conditioned on allegiance				
True effect size δ_j	.000	34	28.06	>.500	.000

Note. The grand mean λ_o in the unconditional model is the aggregate effect size. In the model conditioned on allegiance, the intercept λ_o is the aggregate effect size when the allegiance to each treatment is equal. The variance component related to the true effect size δ_j is an estimate of the true variability for differences among treatments. The fixed effect for allegiance (viz., λ_1) is expected effect size change for every unit increase in allegiance to a treatment. In each model, the H statistic provides a test of whether the variability among studies is greater than would be expected by chance (i.e., effects are heterogeneous); H has an approximate chi-square distribution with k–1 degrees of freedom, where k is the number of effects. I^2 indexes the proportion of variability in effects as a result of true differences among effects.

One way of understanding the heterogeneity of effect sizes is by decomposing the variability among the obtained effects into two sources. The first source of variability is sampling error (i.e., the variance of the e_js). In the present example, as discussed previously, if the true difference among treatments is zero, studies investigating the difference would, by chance, show some effects in one direction or the other; this is sampling error. The second source is the variability of the true effect sizes, that is, var(δ_j), which suggests some systematic influence on the effects. The present meta-analysis suggests that there is variability among the true effects. To make sense of this variability, the descriptive statistic I^2 can be used: I^2, which is equal to $[Q - (k - 1)]/Q$ (and set to zero, if negative), is the estimate of the proportion of observed variability that is due to variability of the true effect sizes (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). In the present case, $I^2 = .38$; that is, it is estimated that 38% of the observed variation in effects was due to variability between true effects (see Table II).

There are two possibilities for the variability of the true effects found in the unconditional model. The first is that some treatments are superior to others (i.e., the dodo bird conjecture is false). The alternative hypothesis tested in this study is that the variation of true effects was due to the allegiance of the researcher. This hypothesis is tested with the conditional model, as described previously, and the results are presented in Table II.

The fixed coefficient for allegiance was .131, which is significantly greater than zero. An increase in one point for allegiance to a treatment over the comparison treatments results is an increase in the true effect size of .131. For example, if the allegiance to one treatment was 4 (treatment developed by author and author-supervised or author-trained therapists) and the allegiance to the other treatment was 2 (treatment developed by authors but authors did not supervise or train therapists), then an effect for the first treatment of .262 would be expected ($2 \times .131$) entirely as a result of allegiance. The effect of allegiance is also shown by the random effect in the conditional model, because there is no variation among true effects when allegiance to treatment is modeled. That is, allegiance accounts for 100% of the variation of true effects detected in the unconditional model. Because there was no residual variability in effects sizes, no moderator variables were tested to explain the remaining variability; thus, post hoc tests of age, severity, and diagnosis were not conducted.

Discussion

The present meta-analysis of psychological treatments for youth found that effect sizes were not homogeneously distributed around zero, thereby suggesting the possibility of differential effects of the various treatments. At the same time, however, the upper bound of the true effect size of the difference between treatments was only .22, a small value that is consistent with the one reported for adult treatments (cf. Wampold, Mondin, Moody, Stich, et al., 1997). Furthermore, controlling for allegiance of the researcher to the treatment approach under investigation removed all variability among the effects. In other words, allegiance explained all the observed systematic differences among treatments.

Given that the current investigation avoided confounds known to have affected prior meta-analyses of treatments of children and adolescents (i.e., different dependent measures, diagnostic assessments, treatment doses) by including only those studies in which two or more treatments were compared within the same study, the results are generally consistent with the dodo bird verdict, when allegiance is controlled. Where treatments for the youth population are concerned, apparently "all have won and all deserve prizes" (Rosenzweig, 1936).

Also consistent with findings from the adult literature (Berman et al., 1985; Devilly, 2001; Dush et al., 1983; Hoag & Burlingame, 1997; Lambert, 1999; Luborsky et al., 1999, 2002; Paley & Shapiro, 2002; Robinson et al., 1990; Shapiro & Shapiro, 1982; Smith et al., 1980), the allegiance of the researcher to the treatment being studied was clearly related to the effect produced. It appears that allegiance is robustly related to the results of clinical trials; in the current study, the superiority of any treatment to another was due to the researcher's allegiance to the superior treatment. Interestingly, this result is consistent with the finding that evidence-based treatments for youth are superior to usual care only if the evidence-based treatment was developed by the researcher (Weisz, Jensen-Doss, & Hawley, 2006). As Luborsky et al. (1999) discussed, the manner in which allegiance affects the outcomes of clinical trials is not determinable from a meta-analysis of this sort, because the effects may be that the therapists in the study are more skilled in or have greater belief in the preferred treatment or that the design of the experiment is biased in a way unrelated to the delivery of the treatment. Wampold (2001) has hypothesized that allegiance effects are due to therapists, which is supported by fact that the coding system in this study focused on aspects of the design

related to the therapists' relationship to the researcher. We suggest that therapists trained and supervised by a researcher who developed a particular treatment or who has advocated for a particular treatment have greater belief in the efficacy of that treatment, are better trained, and receive more attention vis-à-vis comparison treatments. A fairer comparison is created when the therapists for each treatment are selected for their expertise and belief in the treatment and are trained and supervised by experts in the respective treatments, as was the case in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (see Elkin, Parloff, Hadley, & Autry, 1985).

A number of potential limitations of the present meta-analysis should be noted. First and foremost, the number of studies included in the analysis is small ($n = 23$). Although a thorough search of the literature was conducted and more than 1,000 research articles were identified and reviewed, it is entirely possible that some studies were missed. Additionally, unpublished studies (i.e., dissertations) were purposefully excluded. With regard to the latter, however, publication bias suggests that studies finding differences would be more likely to be published than studies finding no differences (Atkinson, Furlong, & Wampold, 1982; Rotton, Foos, Van Meek, & Levitt, 1995), so omission of unpublished reports would likely inflate differences among treatments. Although it is entirely possible that missing or excluded studies could lead to different results, similar findings from the research on adults cited earlier, in combination with the meta-analyses of the youth literature (Spielmans et al., in press; Weisz, McCarty, & Valeri, 2006), raise strong doubts about the likelihood of a substantially different outcome.

A second limitation, related to this first, regards the number of disorders treated and the variety of approaches compared. For example, of the many disorders said to affect youth, the current analysis included only four: depression, anxiety, conduct disorder, and attention-deficit/hyperactivity disorder (ADHD). At the same time, however, it is important to note that the scope of the study was purposefully limited to depression, anxiety, conduct disorder, and ADHD because these diagnoses had been identified in prior reviews as "key problem domains" for which evidence-based treatments for children and adolescents exist (Kazdin, 2000, 2004; Kazdin & Weisz, 2003; Lonigan & Elbert, 1998; Nathan & Gorman, 2002; Task Force on the Promotion and Dissemination of Psychological Problems, 1995). At least where the key problem domains are concerned, the results undermine the claim of "compelling evidence that some techniques are clearly the treatment of choice for . . . children and adolescents" (Kazdin, 2004, p. 580).

Another limitation of this meta-analysis is that only a handful of the 550 documented psychotherapies applied to children and adolescents were included in the analysis (Kazdin, 2000). Moreover, of those tested, an argument could be made that the models were more similar than different, with most sharing a common cognitive or behavioral basis (see Table I). As such, at a minimum, it would be inappropriate to conclude from the present study that all therapies are equally effective for all disorders affecting children and adolescents. However, it is worth noting that, based on the primary studies available, it is equally inappropriate to conclude that some treatments are superior to others; that is, if the set of studies restricts one conclusion, it similarly restricts all conclusions of the same type. It remains to be seen whether future studies comparing more diverse treatment approaches would result in substantially different findings. Given the significant effect that researcher allegiance was shown to have on outcome in this and other studies, considerable care will need to be exercised to ensure evenly balanced allegiance between the treatments being compared.

Limitations aside, the findings from the present study have important implications for research, policy, and practice. For example, in light of these and other findings cited, current attempts aimed at identifying and codifying a list of best practices for the treatment of children and adolescents can at best be viewed as premature and at worst misleading. At a minimum, much more research comparing two or more bona fide treatments needs to be done before professional organizations, payers, and regulators deem specific approaches as "best."

References

References marked with an asterisk indicate studies included in the meta-analysis.

American Academy of Child and Adolescent Psychiatry. (1997). Practice parameters [Entire issue]. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(10, Suppl.).

American Academy of Child and Adolescent Psychiatry. (1998). Practice parameters [Entire issue]. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*(10, Suppl.).

Atkinson, D. R., Furlong, M. J., & Wampold, B. E. (1982). Statistical significance, reviewer evaluations, and the scientific process: Is there a (statistically) significant relationship? *Journal of Counseling Psychology, 29*, 189–194.

Berman, J. S., Miller, C., & Massman, P. J. (1985). Cognitive therapy versus systematic desensitization: Is one treatment superior? *Psychological Bulletin, 97*, 451–461.

*Brent, D. A., Holder, D., & Kolko, D. (1997). A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Archives of General Psychiatry, 54*, 877–885.

*Butler, L., Miezitis, S., Friedman, R., & Cole, E. (1980). The effect of two school-based intervention programs on depressive symptoms in preadolescents. *American Educational Research*, *17*(1), 111–119.

Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin*, *98*, 388–400.

Christophersen, E. R., & Mortweet, S. L. (2001). *Treatments that work with children: Empirically supported strategies for managing childhood problems*. Washington, DC: American Psychological Association.

Crits-Christoph, P. (1997). Limitations of the dodo bird verdict and the role of clinical trials in psychotherapy research: Comment on Wampold et al. (1997). *Psychological Bulletin*, *122*, 216–220.

*Denkowski, K. M., & Denkowski, G. C. (1984). Is group progressive relaxation training as effective with hyperactive children as individual EMG biofeedback treatment? *Biofeedback and Self-Regulation*, *9*, 353–365.

Devilly, G. J. (2001). The influence of distraction during exposure and researcher allegiance during outcome trials. *Behavior Therapist*, *24*, 18–21.

*Dubey, D. R., O'Leary, S. G., & Kaufman, K. F. (1983). Training parents of hyperactive children in child management: A comparative outcome study. *Journal of Abnormal Child Psychology*, *11*, 229–246.

Dush, D. M., Hirt, M. L., & Schroeder, H. E. (1983). Self-statement modification with adults: A meta-analysis. *Psychological Bulletin*, *94*, 408–422.

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

Fonagy, P., Target, M., Cottrell, D., Phillips, J., & Kurtz, Z. (2002). *What works for whom? A critical review of treatments for children and adolescents*. New York: Guilford Press.

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

Hoag, M. J., & Burlingame, G. M. (1997). Evaluating the effectiveness of child and adolescent group treatment: A meta-analytic review. *Journal of Clinical Child Psychology*, *26*, 234–246.

*Horn, W. F., Ialongo, N., Popovich, S., & Peradotto, D. (1987). Behavioral parent training and cognitive-behavioral self-control therapy with ADD-H children: Comparative and combined effects. *Journal of Clinical Child Psychology*, *16*, 57–68.

Howard, K. I., Krause, M. S., Saunders, S. M., & Kopta, S. M. (1997). Trials and tribulations in the meta-analysis of treatment differences: Comment on Wampold et al. (1997). *Psychological Bulletin*, *122*, 221–225.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*, 193–206.

*Hughes, R. C., & Wilson, P. H. (1988). Behavioral parent training: Contingency management versus communication skills training with or without the participation of the child. *Child and Family Behavior Therapy*, *10*, 11–23.

*Kazdin, A. E. (1987). Treatment of antisocial behavior in children: Current status and future directions. *Psychological Bulletin*, *102*, 187–203.

Kazdin, A. E. (2000). *Psychotherapy for children and adolescents: Directions for research and practice*. New York: Oxford University Press.

Kazdin, A. E. (2004). Psychotherapy for children and adolescents. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed) (pp. 543–589). Oxford, England: Wiley.

*Kazdin, A. E., Bass, D., Siegel, T., & Thomas, C. (1989). Cognitive-behavioral therapy and relationship therapy in the treatment of children referred for antisocial behavior. *Journal of Consulting and Clinical Psychology*, *57*, 522–535.

*Kazdin, A. E., Bass, D., Siegel, T., & Thomas, C. (1992). Cognitive problem solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting and Clinical Psychology*, *57*, 522–535.

Kazdin, A. E., & Weisz, J. R. (2003). *Evidence-based psychotherapies for children and adolescents*. New York: Guilford Press.

Lambert, M. J. (1999). Are differential treatment effects inflated by researcher therapy allegiance? Could clever Hans count? *Clinical Psychology: Science and Practice*, *6*, 127–130.

Lambert, M. J. (2004). *Handbook of psychotherapy and behavior change* (5th ed). New York: Wiley.

*Leal, L. L., Baxter, E. G., Martin, J., & Marx, R. W. (1981). Cognitive modification and systemic desensitization with test anxious high school students. *Journal of Consulting and Clinical Psychology*, *28*, 525–528.

Lonigan, C. J., & Elbert, J. C. (Eds). (1998). Special issue on empirically supported psychosocial interventions for children. *Journal of Consulting and Clinical Psychology, 27,* 138–226.

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

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice*, *9*, 2–12.

*Luk, E. S. L., Staiger, P., Mathai, J., Field, D., & Adler, R. (1998). Comparison of treatments of persistent conduct problems in primary school children: A preliminary evaluation of a modified cognitive-behavioral approach. *Australian and New Zealand Journal of Psychiatry*, *32*, 379–386.

*Murnis, P., Merckelbach, H., Holdrinet, I., & Sijsenaae, M. (1998). Treating phobic children: Effects of EMDR versus exposure. *Journal of Consulting and Clinical Psychology*, *66*, 193–198.

Nathan, P. E., & Gorman, J. M. (Eds). (2002). *Treatments that work* (2nd ed.). New York: Oxford University Press.

*Nock, M. K., & Kazdin, A. E. (2005). Randomized controlled trial of a brief intervention for increasing participation in parent management training. *Journal of Consulting and Clinical Psychology*, *73*, 872–879.

Paley, G., & Shapiro, D. A. (2002). Lessons from psychotherapy research for psychological interventions for people with schizophrenia. *Psychology and Psychotherapy: Theory, Research and Practice*, *75*, 5–17.

Raudenbush, A., & Bryk, S. (2001). *Hierarchical linear models: Applications and data analysis methods (advanced quantitative techniques in the social sciences)*. New York: Sage.

Raudenbush, A., Bryk, S., & Congdon, Y. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

*Reynolds, W. M., & Coats, K. (1986). A comparison of cognitive–behavioral therapy and relaxation training for the treatment of depression in adolescents. *Journal of Consulting and Clinical Psychology*, *54*, 653–660.

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

Rogers, C. R. (1951). *Client-centered counseling*. Boston: Houghton-Mifflin.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods in psychotherapy. *Journal of Orthopsychiatry*, *6*, 412–415.

*Rossello, J., & Bernal, G. (1999). The efficacy of cognitive–behavioral and interpersonal treatments for depression in Puerto Rican adolescents. *Journal of Consulting and Clinical Psychology*, *67*, 734–745.

Roth, A., & Fonagy, P. (2005). *What works for whom: A critical review of psychotherapy research* (2nd edn). New York: Guilford Press.

Rotton, L., Foos, P. W., Van Meek, L., & Levitt, M. (1995). Publication practices and the file drawer problem: A survey of published authors. *Journal of Social Behavior and Personality*, *10*, 1–13.

*Schneider, B. (1991). Comparison of skill-building and desensitization strategies for intervention with aggressive children. *Aggressive Behavior*, *17*, 301–311.

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

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

Shirk, S., & Russell, R. (1992). A reevaluation of child therapy effectiveness. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*, 703–709.

*Silverman, W. K., Kurtines, W. M., Weems, C. F., Ginsburg, G. S., Rabian, B., & Serafini, L. (1999). Contingency management, self control, and education support in the treatment of childhood phobic disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *67*, 675–687.

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

*Sonuga-Barke, E. J. S., Daley, D., Thompson, M., Laver-Bradbury, C., & Weeks, A. (2001). Parent-based therapies for preschool attention-deficit/hyperactivity disorder: A randomized, controlled trial with a community sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 402–409.

*Spence, S. H., Donovan, C., & Brechman-Toussaint, M. (2000). The treatment of childhood social phobia: The effectiveness of a social skills training-based, cognitive-behavioral intervention, with and without parental involvement. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 713–727.

Spielmans, G., Pasek, L., & McFall, J. (in press). What are the active ingredients in cognitive and behavioral therapy for anxious and depressed children? A meta-analysis. *Clinical Psychology Review.*

*Stark, K. D., Reynolds, W. M., & Kaslow, N. J. (1987). A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression in children. *Journal of Abnormal Child Psychology*, *15*, 91–113.

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

Tanaka-Matsumi, J., & Kameoka, V. A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. *Journal of Consulting and Clinical Psychology*, *54*, 328–333.

Task Force on the Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatment: Report and recommendations. *The Clinical Psychologist, 48*, 2–23.

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

Wampold, B. E., Mondin, G. W., Moody, M., & Ahn, H. (1997). The flat earth as a metaphor of the evidence for uniform efficacy of bona fide psychotherapies: Reply to Howard et al. (1997) and Crits-Christoph (1997). *Psychological Bulletin*, *122*, 226–230.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes. *Psychological Bulletin*, *122*, 203–215.

*Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology*, *65*, 93–109.

*Webster-Stratton, C., Kolpacoff, M., & Hollinsworth, T. (1988). Self-administered videotape therapy for families with conduct-problem children: Comparison with two cost-effective treatments and a control group. *Journal of Consulting and Clinical Psychology*, *56*, 558–566.

Weiss, B., & Weisz, J. R. (1995). Effectiveness of psychotherapy. *Journal of the American Academy of Child & Adolescent Psychiatry*, *34*, 971–972.

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus clinical care: A meta-analysis of direct comparisons. *American Psychologist*, *61*, 671–689.

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin*, *132*, 132–149.

Weisz, J. R., & Weiss, B. (1993). *Effects of psychotherapy with children and adolescents*. Newbury Park, CA: Sage.

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

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117*, 450–468.