Effects of Psychotherapy for Depression in Children and Adolescents: A Meta-Analysis

John R. Weisz,
Judge Baker Children's Center, Harvard University

Carolyn A. McCarty, and
University of Washington

Sylvia M. Valeri
Brown University

Abstract

Serious sequelae of youth depression, plus recent concerns over medication safety, prompt growing interest in the effects of youth psychotherapy. In previous meta-analyses, effect sizes (ESs) have averaged .99, well above conventional standards for a large effect and well above mean ES for other conditions. The authors applied rigorous analytic methods to the largest study sample to date and found a mean ES of .34, not superior but significantly inferior to mean ES for other conditions. Cognitive treatments (e.g., cognitive–behavioral therapy) fared no better than noncognitive approaches. Effects showed both generality (anxiety was reduced) and specificity (externalizing problems were not), plus short- but not long-term holding power. Youth depression treatments appear to produce effects that are significant but modest in their strength, breadth, and durability.

Keywords

depression; children; adolescents; psychotherapy; meta-analysis

Depression in children and adolescents (herein referred to collectively as *youths*) is a significant, persistent, and recurrent public health problem that undermines social and school functioning, generates severe family stress, and prompts significant use of mental health services (Angold et al., 1998; Clarke, DeBar, & Lewinsohn, 2003). Youth depression is also linked to increased risk of other psychiatric disorders (Angold & Costello, 1993) as well as drug use and suicide (Gould et al., 1998; Rohde, Lewinsohn, & Seeley, 1991), which is the third most common cause of death among adolescents (Arias, MacDorman, Strobino, & Guyer, 2003). Relapse rates have been reported at 12% within 1 year and 33% within 4 years (Lewinsohn, Clarke, Seeley, & Rohde, 1994), and by the age of 18 years, some 20% of youths will have met criteria for a diagnosis of major depressive disorder at least once (Birmaher et al., 1996). Prospective longitudinal research has shown substantial continuity of youth depression into adulthood, with impaired functioning in work, social, and family life, and markedly elevated risk of adult suicide attempts and completed suicide (see, e.g., Costello et al., 2002; Weissman et al., 1999). The extent, impact, and long-term sequelae of youth depression underscore the need for effective treatment. A primary purpose of the current article was to assess the effects of the most extensively tested genre of youth depression treatment: psychotherapy. In this article, we seek to answer the six questions listed below.

Correspondence concerning this article should be addressed to John R. Weisz, Judge Baker Children's Center, 53 Parker Hill Avenue, Boston, MA 02120-3225. E-mail: jweisz@jbcc.harvard.edu.
John R. Weisz, Judge Baker Children's Center, Harvard University; Carolyn A. McCarty, Departments of Pediatrics and Psychology, University of Washington; Sylvia M. Valeri, Department of Psychiatry and Human Behavior, Brown University.

What Is the Overall Effect of Psychotherapy on Youth Depression?

The need to examine psychotherapy effects is underscored not only by evidence on the extent, impact, and sequelae of youth depression but also by recent debate over medication risks. Selective serotonin reuptake inhibitors (SSRIs) have become a widely used treatment for depressed youths (Safer, 1997; Treatment for Adolescents with Depression Study [TADS] Team, 2004; Weisz & Jensen, 1999), but concerns over possible risks, including suicidal ideation and suicide attempts (Vitiello & Swedo, 2004; Whittington et al., 2004), have led regulatory agencies to hold hearings, issue safety warnings, and (in the United Kingdom) classify SSRIs as "contraindicated" for pediatric use (see Committee on Safety of Medicines, 2004; U.S. Food and Drug Administration [FDA], 2004). Most recently, the FDA (2004) issued a "black box" warning on all antidepressants, not just SSRIs, to underscore the possible risk and thus encourage clinicians and parents to consider alternatives to medication. Concerns about pharmacotherapy have thus refocused attention on the most prominent medication alternative, psychotherapy, and on the question of how effective psychotherapy is with youth depression.

Efforts to answer this question can be found in multiple tests of youth psychotherapy programs conducted over the past 2 decades, and each individual trial sheds some light on treatment impact. However, quantitative experts (e.g., Cohen, 1990; Rosenthal, 1990; Schmidt, 1992) caution that reports of p values from individual studies may often be misleading and that a better way to evaluate progress is to rely on effect size (ES) values and on meta-analyses that synthesize those values across multiple studies. Like any statistical tool, meta-analysis is most informative when applied to the most representative data possible (e.g., the most complete collection of studies), and successive meta-analyses are needed to ensure accuracy as the pool of relevant studies builds over time.

These points are particularly relevant to the domain of youth depression. Three meta-analyses have been published on this topic, with findings that convey an unusually positive picture of treatment success. In the earliest of the three, Reinecke, Ryan, and DuBois (1998a) focused on a sample of six depression treatment studies with adolescents, all involving cognitive–behavioral therapy (CBT); they reported a mean ES across studies of 1.02. In response to a commentary by Harrington, Campbell, Shoebridge, and Whittaker (1998), Reinecke, Ryan, and DuBois (1998b) used an alternate computational method, generating a mean ES of 0.97, still markedly above means in the youth treatment literature generally. In a second meta-analysis, focused on CBT with depressed adolescents, Lewinsohn and Clarke (1999) reported an even larger mean ES of 1.27, on the basis of what appear to be 12 treatment–control comparison studies. In a third meta-analysis, Michael and Crowley (2002) addressed treatment of child and adolescent depression, placing no limits on the kinds of psychotherapy used. Their study collection included 14 controlled trials (plus a larger number of prepost design studies that were analyzed separately), encompassing multiple forms of youth depression treatment, and they reported a mean ES of 0.72 for these controlled trials. Averaging across these three meta-analyses, the mean ES is 0.99, well above Cohen's widely used benchmark of 0.80 for a large effect and markedly higher than the mean value of 0.54, reported in the most recent broad-based meta-analysis of youth therapy outcome encompassing diverse treated problems and disorders (Weisz, Weiss, Han, Granger, & Morton, 1995).

The very high mean ES across these three youth depression meta-analyses might be read as an indication that the search for highly effective nonbiological treatment has already succeeded, and that resources for treatment development and testing should be focused on youth problems and disorders other than depression, which appear to show much smaller effects. Before reaching such a conclusion, however, it may be wise to examine the prior meta-analyses closely and to consider additional evidence that is now available.

First, it should be noted that although the previous meta-analyses were valuable in relation to their specific goals, all three may have provided a somewhat less than complete picture, and new studies have appeared in the years since 1999, the date of the most recent study that was included in these previous meta-analyses. We have now identified more than twice as many randomized trials as the most complete meta-analysis to date (Michael & Crowley, 2002). Second, it should be noted that two of the prior meta-analyses (Lewinsohn & Clarke, 1999; Reinecke et al., 1998a,1998b) included only articles published in peer-reviewed journals. Including only such peer-reviewed work may increase the risk of ES overestimation because the peer review process favors positive findings (see, e.g., McLeod & Weisz, 2004). Third, for those studies that the meta-analysts did identify, ES computation appears to have relied solely on data in the published reports, on estimates of effects for the numerous studies not reporting sufficient data, and on exclusion of studies (e.g., 50% of the studies identified, in one of the meta-analyses), rather than on contacting authors and persuading them to provide the information needed for precise ES values. Fourth, close examination suggests that the meta-analyses did not consistently require random assignment of study participants to treatment and control conditions or consistently compute ES on the basis of all depression outcome measures obtained in the studies.

Another important issue is that previous meta-analyses relied on fixed effects analyses. This approach is widely used in meta-analyses, but technically it is appropriate only if homogeneity analyses support the assumption that all ES values, across studies, are estimating the same population mean. Our examination (see below) suggests that the assumption is not valid in the case of youth depression psychotherapy studies and thus that random effects analyses are appropriate. Moreover, only random effects analyses permit generalization from the particular studies under review to other treatment studies that used different participants, methods, doses, and outcome measures.

All these issues suggest the need for a current examination of the outcome literature on youth depression treatment. In providing such an examination in the current article, we (a) obtained and calculated ES for a particularly complete pool of youth depression psychotherapy trials, including both peer-reviewed and non-peer-reviewed studies, (b) required random assignment of participants to study conditions to ensure fair tests of treatment effects, (c) contacted authors repeatedly to obtain the exact information needed to compute ES precisely for all studies in our pool, thus eliminating the need to make rough estimates of study parameters or drop studies altogether, (d) used all of the published depression outcome measures included in each study, omitting none, and (e) tested homogeneity of ES and used random effects analyses as indicated.

How Lasting Are Youth Depression Psychotherapy Effects?

In addition to providing the most complete assessment of youth depression treatment effects to date, we sought to gauge the holding power of those effects. We did so by quantifying the effects found in clinical trial follow-up assessments and comparing these with the effects found at immediate posttreatment. Such a comparison provides an indication of whether the changes occurring during treatment are internalized in a way that lasts, and thus whether booster sessions and *continuation treatment* (see, e.g., Clarke, Rohde, Lewinsohn, Hopps, & Seeley, 1999; Weissman, 1994) may be needed to sustain treatment benefit over time. In the two previous youth depression meta-analyses that reported ES at follow-up, effects were found to diminish over time; however, these reports were based on only six studies in Reinecke et al.'s (1998b) analyses and eight studies in Michael and Crowley's (2002) analyses. Our larger and more representative pool of studies appears to offer a more reliable picture of holding power, with 19 studies providing usable follow-up tests.

Are the Psychotherapy Effects Specific to Depression, or Do They Generalize to Other Conditions?

We sought to learn whether depression treatment effects were limited to depression symptoms or generalized to other conditions. Previous analyses of the effects of youth psychotherapy (in Weisz et al.'s, 1995, study) examined specificity grossly by comparing treatment effects on problems targeted by the intervention with effects on all other problems; however, in depression treatment, a more refined question arises: whether benefits of depression treatment spread to symptoms of more closely related and more distally related syndromes—that is, anxiety and conduct problems, respectively. We expected that depression effects might not carry over to symptoms so different as those in the externalizing domain; however, high correlations between youth depression and anxiety (see, e.g., Achenbach, 1990) and theories that posit common risk factors for anxiety and depression (e.g., the tripartitate model of emotion; Clark, Watson, & Mineka, 1994) suggest that treatments that reduce depression may have benevolent effects on anxiety as well. This possibility is directly relevant to recent debates about whether depression and anxiety require separate treatments or can be treated by combined intervention for emotional disorders (i.e., depression plus anxiety; see Barlow, Allen, & Choate, 2004), but no previous meta-analysis has addressed the issue. Thus, we assessed the specificity of depression treatment effects by assessing outcomes separately for (a) depression measures, (b) measures of anxiety, and (c) measures of externalizing problems (e.g., disruptive conduct).

Can Youth Psychotherapy Outperform Active Control Conditions?

A potential weakness of most psychotherapy outcome research, including youth depression research, is that it rests on comparisons of active treatments with inert conditions such as waitlist or no treatment (see Jensen, 2003; Weisz, 2004). Such inert conditions control only for the passage of time and the natural time course of problems and disorders, not for placebo or expectancy effects and not for such nonspecific effects as improvement due to attention or to the benefits of a therapeutic relationship. Extensive research (e.g., Baskin, Tierney, Minami, & Wampold, 2003; Kazdin, Bass, Ayers, & Rodgers, 1990) has demonstrated that comparisons of two active conditions, such as treatment versus placebo, generate lower ES values than comparisons of active versus inert conditions. Accordingly, a key question for treatment of any problem, including depression, is whether active treatments show significant effects when compared with active control groups. If treatment effects of youth depression are found to be significant only in comparison with inert control groups, then one implication could be that the apparent benefit of treatments designed specifically for youth depression is somewhat illusory, resembling placebo effects or resting on more generic benefits of nonspecific factors such as the therapeutic relationship (see, e.g., Horvath & Luborsky, 1993). To address this issue, we examined effects for studies in which active treatments were compared with active control groups and for studies in which active treatments were compared with passive control groups.

Are Treatments That Emphasize Changing Cognitions More Effective Than Treatments That Lack a Cognitive Emphasis?

The current literature in youth depression treatment heavily emphasizes approaches that stress altering unrealistic negative cognitions (e.g., CBT and cognitive restructuring treatments). However, research on adult depression treatment (e.g., Hollon, 2000; Jacobson et al., 1996; Jacobson, Martell, & Dimidjian, 2001) has raised questions about whether a cognitive emphasis is needed to generate improvement and even whether cognitive intervention adds significantly to such noncognitive approaches as behavioral activation. Promising results of some recent youth depression trials that used treatments without a cognitive emphasis suggest that this

question is also relevant to youth depression treatment. To address the question, we tested the magnitude of effects for treatments with a cognitive emphasis and for treatments that do not emphasize cognitive change.

Is Youth Psychotherapy Effective in Clinically Representative Conditions?

We focused an additional set of analyses on the question of whether treatment effects were evident in clinically representative conditions. Critiques of psychotherapy research have raised the concern that many of the studies are controlled efficacy trials with design features that do not resemble actual clinical practice, and that some treatments that succeed under efficacy trial conditions may not work so well under clinical practice conditions (see, e.g., Weisz, 2004; Westen, Novotny, & Thompson-Brenner, 2004). Of special interest in these critiques have been the research participants (i.e., whether participants were recruited vs. clinically referred; see, e.g., Hammen, Rudolph, Weisz, Burge, & Rao, 1999), the treating therapists (i.e., whether therapists were research employees vs. practicing clinicians; see, e.g., Michael, Huelsman, & Crowley, 2005), and the treatment setting (i.e., whether treatment took place in a research setting vs. a clinical service setting; see, e.g., Southam-Gerow, Weisz, & Kendall, 2003). We included all three elements in our analyses and examined the magnitude of effects in (a) studies that used clinically referred youths and those that used recruited youths, (b) studies that used practicing clinicians as therapists and those that used research staff therapists, and (c) studies of treatment conducted in clinical service settings and those with treatment conducted in nonclinical research settings. A finding that treatment effects were insignificant for referred youths, with practicing clinician therapists, or in service settings, would support concerns about the relevance of research results to clinical practice.

Finally, in harmony with recent recommendations in the youth treatment literature (e.g., Kazdin, 2000; Weisz, 2004), we carried out a series of secondary analyses that examined whether depression treatment showed measurable benefit across dimensions that have been identified as important potential predictors of outcome in the treatment literature. These dimensions included youth characteristics (age, gender, and whether youths were selected for study samples based on depressive disorder diagnoses or based on depression symptom measures; see, e.g., Weisz, Valeri, McCarty, & Moore, 1999); treatment intervention characteristics (group vs. individual modality, and treatment duration; see, e.g., McRoberts, Burlingame, & Hoag, 1998); and study characteristics (study attrition rates, whether studies were peer-reviewed, and whether outcomes were assessed via youth self-report vs. parent report; see, e.g., Hammen & Rudolph, 1996). These analyses, although not addressing our six primary aims, were useful in delineating the boundary conditions within which psychotherapy is beneficial.

Method

For the purposes of this review, *psychotherapy for depression* was defined as an intervention designed to alleviate depressive disorders or elevated levels of depressive symptomatology through structured or unstructured interaction or a training program, which was provided by one or more individuals trained to deliver the intervention or through a program designed to be self-administered.

Literature Review

Studies were obtained through (a) computer searches on PsycINFO (1887–2004), Dissertation Abstracts International (1861–2004), and MEDLINE (1994–2004); (b) examination of reference lists in relevant review articles and reference trails from outcome studies; (c) hand searching all issues from 1965–2004 of those journals in which at least five psychotherapy studies had been identified through our computer and reference searches; and (d) personal

communications with authors of relevant studies, asking whether they knew of any additional relevant studies. Keywords used in the computer searches included the following three diagnosis/problem terms (*depression*, *dysthymia*, *major depression*), with the search limited to child and adolescent populations (mean age less than 18 years) and publications that were classified as treatment outcome, clinical trial, single-blind design, or double-blind design.

Criteria for Study Inclusion, and Resulting Pool of Studies

To be included in the meta-analysis, a study had to meet the following criteria: (a) participants selected because of elevated levels of depressive symptoms, formal diagnosis of major depressive disorder or dysthymic disorder, or research diagnostic criteria diagnoses of minor or intermittent depression; (b) random assignment of participants to at least one active treatment group and at least one untreated, waitlist, minimally treated, or active placebo control group (one study was excluded because it compared two forms of group therapy that were described as active treatments, with no control group); (c) samples of mean age younger than 19 years; and (d) intervention intended by the investigators to target depressive symptoms or disorder. To reduce the possibility of a publication bias favoring positive results, which threatens the validity of meta-analyses (Cook et al., 1993; McLeod & Weisz, 2004; Sohn, 1996), we included non-peer-reviewed studies (e.g., book chapters) and doctoral dissertations. When separate articles were published from the same data set (e.g., articles presenting posttreatment and follow-up findings separately, or dissertations that were later published in journals), these were combined for analysis as a single study. Single-subject designs were excluded because they generate ES values that are not comparable to those for group-comparison designs, and because they pose some risk of idiosyncratic findings based on unusual characteristics of a small number of participants.

Our final sample consisted of 35 studies. Summary information about the individual studies, including their ES values, appears in Table 1. Study references appear, with asterisks, in the reference list. One of the 35 studies (Clarke et al., 1995) was labeled by its authors as "targeted prevention"; we included it because it met all our criteria for studies of psychotherapy, and its sample included only adolescents who showed elevated scores on a standardized depression measure.

Study Coding Procedures, Codes Used, and Reliability of Codings

Studies were coded for multiple subject, design, and method features. Two judges independently coded all the studies, with interjudge reliability assessed for continuous codes via intraclass correlation coefficients (ICCs) and for categorical codes via kappas. Kappas are conventionally categorized as slight if .01–.20, fair if .21–.40, moderate if .41–.60, substantial if .61–.80, and almost perfect if .81–1.00 (Landis & Koch, 1977). Discrepancies in coding were resolved through the coders' joint review of study details (see also the *ES Calculation Procedures and Reliability* section below).

Outcome measure content—We coded outcome measures in each study for content—that is, whether measures assessed depression, anxiety, or externalizing/conduct problems ($\kappa = .90$).

Outcome informant—We also coded outcome measures as to informant—that is, whether they were derived from youth report, parent report, or teacher report ($\kappa = .76$). Because teacher reports were rarely used, we were only able to structure tests focused on youth and parent reports.

Type of control group—The type of control group used varied across studies. Some used passive controls that were given no additional attention or intervention (e.g., no-treatment or

waitlist-control groups); others used active control groups that provided attention and/or nonspecific treatment elements similar in dose to what was provided for the intervention group ($\kappa = .90$).

Sample type—Turning to participant characteristics, we coded studies for whether a diagnostic or symptom measure was used to identify depressed youths. Diagnosed samples were those who were administered a diagnostic interview and met full diagnostic criteria (*Diagnostic and Statistical Manual of Mental Disorders* or research diagnostic criteria) for major depressive disorder, dysthymic disorder, minor depressive disorder, or intermittent depressive disorder. Symptom measure samples showed elevated depressive symptoms on a symptom measure ($\kappa = 1.00$).

Age group of participants—We also coded samples according to age group. As in prior meta-analyses (e.g., Michael & Crowley, 2002; Weisz et al., 1995), we classified participants under 13 as *children* and those 13 or older as *adolescents* ($\kappa = .94$). Studies that included both children and adolescents ($n = 13$) were not included in age group analyses.

Psychotherapy type—Limited variability in treatment methods used to date (i.e., most involved some form of behavioral intervention) left us with only two meaningful treatment method categories: *cognitive emphasis* (i.e., approaches that emphasize changes in beliefs or ways of thinking about events and conditions) versus *no cognitive emphasis*—treatments that do not emphasize changing beliefs or ways of thinking about events and conditions ($\kappa = 1.00$). A treatment was coded as having a cognitive emphasis if the treatment description gave any indication that treatment focused on changing cognitions, thoughts, beliefs, ways of thinking, or internal self-talk. Treatments that were coded as not having a cognitive emphasis included attachment-based family treatment, behavioral problem solving, group support, interpersonal psychotherapy, relaxation training, role-playing, self-modeling, social skills training, structured learning therapy, and systematic behavior family therapy.

Therapy modality—Treatment modality was coded as *group* versus *individual* treatment ($\kappa = .77$).

Treatment duration—We used a continuous measure of treatment duration, calculating the number of therapy hours provided (e.g., total time spent in parent, family, and youth sessions; ICC = .97).

Attrition—For each study, we coded the percentage of sample attrition that occurred between the point of randomization and the posttreatment assessment (ICC = .90).

Peer-reviewed versus non-peer-reviewed study—We distinguished between studies that had been published in a peer-reviewed journal and those that had not (including dissertations and book chapters; $\kappa = 1.00$).

Clinically referred versus recruited youths—We coded whether the majority of participants were already referred for mental health services independently of the study (clinically referred) or were recruited into the study (e.g., via advertisement; $\kappa = 1.00$).

Therapists' primary vocation—We differentiated studies in which a majority of therapists were used primarily as clinicians versus studies in which a majority of therapists were not primarily clinicians (e.g., researchers, graduate students, professors; $\kappa = .69$).

Treatment setting—We coded whether treatment was provided in a clinical service setting (e.g., outpatient, inpatient, or day treatment program) or nonclinical setting (e.g., university lab or lab clinic, primary or secondary school; $\kappa = .86$).

ES Calculation Procedures and Reliability

ES values were calculated separately for each outcome measure, informant (e.g., parent, youth), and time of assessment (e.g., posttreatment, follow-up), then averaged up to the level of the target comparison (as we describe later). Following Smith, Glass, and Miller's (1980) study, ES was the posttherapy difference between the control and treatment group means, divided by control group standard deviation. This resembles Cohen's (1988) *d*, except that the divisor is the control group standard deviation rather than the pooled standard deviation across both groups; pooling using the posttreatment standard deviation may be inappropriate for researchers conducting psychotherapy studies because treatment may increase variability (see Weisz et al., 1995, p. 455, for supporting evidence; see Michael & Crowley, 2002, for a useful alternative to our approach). All ESs were calculated such that positive values implied an advantage for treatment over control group. In addition, we adjusted all ES values using Hedges's small sample correction (Hedges & Olkin, 1985), which yields an unbiased estimator of ES. As sample size increases, *d* and ES approximate each other. However, with smaller sample sizes, variance estimates are larger, and the distribution of *d* values sampled tends to be skewed. To correct for this small sample bias, we used the following formula:

$$\text{ES} = \text{Unadjusted ES} \times [1 - (3 / [(4 \times N) - 9])],$$

where *N* is equal to the number of participants in the control group plus the number of participants in the treated group.

When means, standard deviations, or other information needed for our calculations or analyses were not reported in an article or manuscript, we contacted authors, sometimes repeatedly, until we obtained the information from them. Using Bollen's (1989) procedure of dropping ESs lying beyond the first gap of at least one standard deviation between adjacent ES values in a positive or negative direction, we conducted analyses that excluded ES outliers for individual measures. When different scales from the same measure were reported, scales were averaged for overall ES, but depression-specific scales were used in computing depression ES.

All ESs were independently calculated by two raters. Agreement, assessed via the ICC, was .98. The few interrater discrepancies were resolved by jointly reviewing the study methodology to ensure that correct information about means, standard deviations, and direction of the outcome measure (e.g., higher score indicating better outcome) had been used; when these steps were taken, no discrepancies in ES remained.

Overview of Data-Analytic and Reporting Procedures

Reporting ES means—We pooled ES values up to the most conservative level appropriate to each test. For example, in calculating ES for depression measures, we collapsed across treatment groups and averaged across all depression measures to produce a single ES mean for each study; however, in computing ES means for different forms of psychotherapy, we did not collapse across treatment groups.

Adjusting for heterogeneity of variance—We analyzed our data using a weighted least squares (WLS) approach; each ES was weighted by the inverse of its variance (Hedges & Olkin, 1985), thereby adjusting for heterogeneity of variance across individual observations. To facilitate comparison with previous meta-analyses that used unweighted least squares (ULS)

procedures (e.g., Michael & Crowley, 2002), we report ULS, as well as WLS values, for overall ES means and for *t* tests comparing these mean ES values with zero.

Tests of homogeneity—Homogeneity analyses (Hedges & Olkin, 1985) were conducted to test the assumption that all the ES values were estimating the same population mean and to inform a decision about random versus fixed effects analyses (see below). The tests were significant for depression measures ES, $Q(30) = 64.77$, $p < .01$, and other ES, $Q(29) = 43.72$, $p = .04$, suggesting that the psychotherapy studies may not estimate common ES parameters.

Random effects versus fixed effects analyses—The decision as to whether to use fixed or random effects models was undertaken by considering both the types of inference that we wished to make and the homogeneity of ES parameters (as discussed in Hedges & Vevea's, 1998, study). Fixed effects analysis only takes account of uncertainty due to the particular samples included in a specific meta-analytic set of studies and thus supports only conditional inference about that set of studies; such analysis is appropriate when homogeneity is not rejected. Random effects analysis is used to make inferences about a population of studies that is larger and more diverse (e.g., in samples, designs, treatment doses, and outcome measures) than a single observed study set used for a single meta-analysis; such analysis is appropriate when homogeneity is rejected. Because homogeneity was rejected in our analyses, and we wished to support inferences about psychotherapy outcomes for depression in the general population of children and adolescents, we used random effects analyses.

Analytic procedures—We used paired *t* tests to compare outcomes on different measures obtained for the same set of studies (e.g., outcomes based on parent vs. youth reports). For comparison of ES values with zero, we used SPSS macros that generate *z* tests based on the absolute value of the mean ES divided by the standard error of the mean ES (Wilson, 2003). To compare mutually exclusive categories of studies, a researcher should use *Q*-statistic analog to analysis of variance (Lipsey & Wilson, 2001). If the between-category variance is significant, then the mean ESs across groups differ by more than sampling error. The *Q* statistic is distributed as a chi-square with $k–j$ degrees of freedom, where k is the number of ESs (Hedges & Olkin, 1985), and j is the number of groups. We conducted all analyses using maximum likelihood, random effects models weighted by the inverse of the variance.

Inclusion of covariates—To control for the possibility that confounding characteristics might produce or obscure ES differences between theoretically important subgroups of studies, we also included multivariate meta-regression analyses. On the basis of prior meta-analyses and reviews (e.g., Kazdin et al., 1990; Weisz, 2004), we identified four variables with apparent potential to influence ES—mean age, percentage of boys, recruitment status (recruited vs. clinically referred youths), and type of control group (active vs. passive). For each instance in which we tested an ES difference between two subgroups of studies, we also assessed whether the two subgroups differed significantly on any of these four variables, and any of the variables found to differentiate the two groups were included as covariates in the metaregression analysis, as a complement to our *Q*-statistic comparison. For analyses of within-study variables (e.g., youth self-report measures vs. parent-report measures), confounding variable differences between different subgroups of studies were not at issue, so covariates were not used.

Estimated power and planned analyses—We limited our analyses to those with acceptable power rather than report numerous underpowered tests and risk unreliable findings due to chance. We followed Cohen's (1988) methods for estimating power on the basis of type of analysis, including those sets of analyses with mean power greater than 0.50 to detect a large effect. Following Cohen's (1988) guidelines, we used 0.80 as our cutoff for a large effect for

t tests and 0.40 as our cutoff for a large effect for *F* tests. These cutoffs permitted three types of planned analyses:

1. Tests of whether ES was greater than zero for the different groups of studies (mean power = 0.85);

2. Tests of some mutually exclusive study characteristics, for example, ES for interventions that used an active placebo versus those that used a passive placebo (mean power = 0.61);

3. Paired *t* tests comparing (a) outcomes on different kinds of outcome measures (e.g., depression vs. externalizing problems), (b) outcome based on youth versus parent reports, and (c) outcome at posttreatment versus follow-up assessments (mean power = 0.63).

Given the lack of literature on the topic, our estimates of power for the *z* tests (tests of ES greater than zero) and *Q* statistics (tests of mutually exclusive study characteristics) were based on Cohen's (1988) procedures for *t* tests and *F* tests (i.e., analogs for *z* and *Q*, respectively).

Results

Characteristics of the full set of studies are shown in Table 2. As the table shows, more studies focused on adolescents alone than children or mixed-age samples. Almost half of the studies included only youths with clinically diagnosed depression, and almost half of the studies used an active control condition. Fewer studies involved clinic-referred youths, provided treatment in a clinical setting, or used practicing therapists. On average, treatments involved 13 hr of psychotherapy, studies reported two depression-specific outcome measures and three nondepression measures (e.g., anxiety), and outcome information was typically available from two different informants (e.g., child and parent). Usable follow-up data were provided in slightly more than half of the studies. Three additional studies included follow-up assessments, but their waitlist-control group received treatment immediately after posttreatment, ruling out meaningful comparison between treatment and control groups.

In reporting our findings, we emphasize ES values for *depression* measures throughout this report, with *other* measures (e.g., anxiety) included in only two ways: (a) combined with depression measures in the next-to-last column of Table 1 to generate overall ES values for each of the studies and (b) in specificity analyses (see below) comparing ES for depression measures with ES for nondepression measures.

What Is the Overall Effect of Psychotherapy on Youth Depression?

First, we focus on mean ES values for the full collection of studies using different quantitative approaches.

Depression ES–WLS approach—Across the 35 studies, the mean WLS ES for depression measures was 0.34 ($SD = 0.40$; range was from −0.66 to 2.02), significantly different from zero ($z = 4.57$, $p < .01$). When we entered each of the treatment versus control group ES values separately, as some previous meta-analyses have done, the mean was 0.38 ($SD = 0.42$; range was from −0.66 to 2.02, reflecting 44 treatment–control group comparisons), also significantly different from zero ($p < .01$).

Depression ES–ULS approach—We also calculated the ULS mean for the 35 depression studies to permit comparison with previous meta-analyses that did not use weighting. Our mean ULS ES was 0.40, significantly different from zero, $t(34) = 4.26$, $p < .01$. The ULS mean when

treatment groups were used as the level of analysis was 0.46, also significantly different from zero, $t(43) = 5.45$, $p < .01$.

Comparison with nondepression studies in previous meta-analyses—To compare the mean ES for depression with the mean ES found for treatment of problems and disorders other than depression, we pooled data from two previous broad-based meta-analyses encompassing treatments for diverse child and adolescent conditions (Weisz et al., 1995; Weisz, Weiss, Alicke, & Klotz, 1987). We identified all studies in those two meta-analyses that tested treatments for conditions other than depression (e.g., aggression, disruptive behavior, attention-deficit/hyperactivity disorder, fears). Because the two previous meta-analyses had excluded non-peer-reviewed studies, we excluded the non-peer-reviewed studies in the current meta-analysis sample from this comparison. The same ES calculation methods were used for both the depression and nondepression study sets. For the peer-reviewed studies of problems other than depression ($n = 180$), the mean WLS ES (for the specific problems targeted in those studies) was 0.69, significantly higher than the WLS ES of 0.37 for the peer-reviewed depression studies of the current meta-analysis, $Q(1, 213) = 8.72$, $p = .01$. Regression analyses controlling for factors that prior literature suggests might influence ES—that is, mean age, percentage of boys, recruitment status (recruited vs. referred youths), and type of control group (active vs. passive)—still resulted in significantly higher ES values for the nondepression outcome studies ($B = 0.33$, $SE = 0.13$, $b = .19$, $p = .01$).

How Lasting Are Youth Depression Psychotherapy Effects?

To assess maintenance of treatment gains over time, we compared ES values obtained immediately after treatment with those obtained in follow-up assessments. In the 19 studies that used follow-up comparisons between treated and untreated groups, the mean lag between end of treatment and follow-up assessment was 37.5 weeks (range = 4–130 weeks). When follow-up assessments were conducted at multiple time points (e.g., 3 and 6 months), ES values were averaged across the time points. We found no overall difference between posttreatment and follow-up ES for depression outcomes (*M*s: 0.30 vs. 0.28, respectively, $p = .84$), suggesting at first blush that there was no fall off in effects. However, a closer look revealed that the correlation between follow-up time lag and follow-up ES was negative and significant ($r = -.50$, $p = .03$). Follow-up assessments conducted near the end of treatment (e.g., 2–3 months) showed relatively large effects, but follow-ups with lags of 1 year or more showed essentially no treatment effect.

Are the Psychotherapy Effects Specific to Depression, or Do They Generalize to Other Conditions?

Next we considered the specificity question, comparing ES for measures of depression with ES for two kinds of nondepression outcome measures: anxiety symptoms and externalizing behavior. In these analyses, we used only studies that had included both depression and anxiety outcomes ($n = 10$) and depression and externalizing behavior outcomes ($n = 11$), respectively. A matched sample *t* test comparing depression ES (0.57) and anxiety ES (0.39) was marginally significant, $t(9) = 2.05$, $p = .07$. By contrast, the *t* test comparing depression ES (0.31) and externalizing ES (0.05) was significant, $t(10) = 2.91$, $p = .02$. The WLS mean ES for anxiety measures was significantly different from zero ($z = 2.73$, $p = .01$), but the mean ES for externalizing problems was not ($z = -0.30$, $p = .77$).

Can Youth Psychotherapy Outperform Active Control Conditions?

The most common form of control group across the pool of studies was passive—that is, waitlist or no treatment. For the 20 studies that used such passive control groups, mean psychotherapy ES was 0.41, significantly different from zero ($z = 4.36$, $p < .01$). The remaining 15 studies

used active control groups; these studies showed a more modest ES mean of 0.24, also significantly different from zero ($z = 2.15$, $p = .03$). ESs for studies that used active versus passive control groups were not significantly different, $Q(1, 33) = 1.46$, $p = .23$. Studies that used active control groups had included clinically referred youths more often than studies that used passive control groups, $\chi^2(1, N = 35) = 9.65$, $p = .002$; when we controlled for this clinic referral variable, the ES differences between active and passive control groups remained insignificant ($B = 0.25$, $SE = 0.18$, $p = .16$).

Are Treatments That Emphasize Changing Cognitions More Effective Than Treatments That Do Not?

We computed mean ES separately for the 31 treatments that involved a cognitive change emphasis (e.g., CBT) and the 13 treatments that did not emphasize cognitive change (e.g., relaxation training). The ES mean for cognitive treatments was 0.35, significantly different from zero ($z = 4.54$, $p < .01$); the mean for noncognitive treatments was 0.47, also different from zero ($z = 3.57$, $p < .01$). The ES difference between cognitive and noncognitive treatments was not significant, $Q(1, 42) = 0.63$, $p = 42$. Studies that used cognitive treatments did not differ from studies that used noncognitive treatments in terms of mean age, gender, recruitment status, and control group, so analyses with covariates were not conducted.

Is Youth Psychotherapy Effective in Clinically Representative Conditions?

Next, we turned to the issue of clinical representativeness in study design, focusing on the youths, therapists, and settings used in the studies.

Clinically referred and recruited youths—Mean ES for studies that used primarily recruited participants was 0.34 ($n = 29$), significantly different from zero ($z = 4.24$, $p < .01$). Mean ES for studies that used clinically referred participants ($n = 6$) was 0.32, also different from zero ($z = 1.99$, $p < .05$). The ES difference between the two groups of studies was not significant, $Q(1, 33) = 0.02$, $p = .89$. Studies with referred youths used active control groups more often, $\chi^2(1, N = 35) = 9.65$, $p < .01$, but the difference between referred and recruited youths remained insignificant when we controlled for type of control group ($B = 0.16$, $SE = 0.22$, $p = .47$).

Therapists who were and were not practicing clinicians—The mean ES for studies in which the majority of therapists were research therapists rather than practicing clinicians was 0.52 ($n = 18$), significantly different from zero ($z = 3.92$, $p < .01$). For studies in which the majority of therapists were practicing clinicians ($n = 9$), the mean ES was 0.27, marginally different from zero ($z = 1.95$, $p = .05$ [$p < .06$]). The difference between the two groups of studies was not significant, $Q(1, 25) = 1.92$, $p = .17$. Studies that used practicing clinicians had higher rates of clinically referred youths, $\chi^2(1, N = 27) = 6.01$, $p = .01$, but the ES difference between practicing clinicians and research therapists remained statistically insignificant when we controlled for recruitment status of youths ($B = -0.26$, $SE = 0.20$, $p = .19$).

Treatment in clinical service settings and in research settings—Twenty-two studies were conducted in nonclinical settings, generating a mean ES of 0.41, significantly different from zero ($z = 4.56$, $p < .01$). Eleven studies were conducted in clinical service settings, yielding an ES of 0.24, also significantly different from zero ($z = 2.04$, $p = .04$). The ES difference between the two sets of studies was not significant, $Q(1, 31) = 1.26$, $p = .26$. Although studies conducted in clinical service settings used referred youths more often, $\chi^2(1, N = 33) = 8.25$, $p < .01$, differences between studies from clinical and nonclinical settings remained insignificant when we controlled for recruitment status ($B = -0.20$, $SE = 0.17$, $p = .24$).

Secondary Analyses of ES as a Function of Youth, Treatment, and Study Characteristics

Next we carried out a series of secondary analyses examining ES as a function of characteristics of the study participants, the treatments provided, and the design and procedural characteristics of the studies.

Youth age—The WLS mean for studies of children (seven studies with samples aged less than 13 years) was 0.41, significantly different from zero ($z = 2.19$, $p = .03$); the mean for adolescents ($n = 21$) was 0.33, also different from zero ($z = 3.64$, $p < .01$). The correlation between mean sample age and ES was 0.02 ($p = .91$). The Q test revealed no significant difference between child and adolescent ES, $Q(1, 264) = 0.16$, $p = .69$. Studies of children had higher rates of male participants than did studies of adolescents, $t(25) = 2.28$, $p = .03$. However, the youth age effect remained nonsignificant when we controlled for gender composition of the sample ($B = -0.17$, $SE = 0.23$, $p = .47$).

Youth gender—The bivariate correlation between ES and percentage of boys in the sample was $-.27$, $p = .12$.

Diagnosed samples and symptom measure samples—Fifteen studies used diagnostic procedures (e.g., assessing major depressive disorder and dysthymic disorder) to identify their samples. Mean ES for these studies was 0.35, significantly different from zero ($z = 3.39$, $p < .01$). Twenty studies used depression symptom measures to identify their samples. The mean ES for these studies was 0.32, also different from zero ($z = 3.25$, $p < .01$). The two study sets did not differ significantly in ES, $Q(1, 33) = 0.03$, $p = .86$. Studies that used diagnosed samples had older participants, $t(31) = 2.58$, $p = .02$, and used more clinically referred samples, $\chi^2(1, N = 35) = 4.84$, $p = .03$. Differences in ES between studies that used diagnosis and symptom measures remained nonsignificant when we controlled for these two covariates ($B = -0.08$, $SE = 0.18$, $p = .67$).

Treatment modality: Group versus individual We computed mean ES separately for the 31 group-based treatments and the 13 individually administered treatments. The WLS ES mean for group treatments was 0.38, significantly different from zero ($z = 4.63$, $p < .01$); the mean for individual treatments was 0.37, also different from zero ($z = 3.22$, $p < .01$). The ES difference between group- and individually administered treatments was not significant, $Q(1, 42) = 0.01$, $p < .93$. Studies that used group treatment also used passive control groups, $\chi^2(1, N = 44) = 7.05$, $p < .01$, and recruited youths, $\chi^2(1, N = 44) = 14.57$, $p < .01$, more frequently than studies that used individual treatment had. When we tested the effect of treatment modality controlling for type of control group and recruitment status, the effect remained nonsignificant ($B = -0.13$, $SE = 0.19$, $p = .49$).

Treatment duration—Treatment duration ranged from 4 to 32 hr, with a mean of 13.5 and a median of 12. The correlation between treatment duration and overall ES was -0.06 ($p = .78$). A direct test based on a median split comparing studies involving less than 12 treatment hr versus those involving 12 or more treatment hr revealed no significant difference, $Q(1, 28) = 1.22$, $p = .27$. Studies that used shorter treatment durations did not differ from studies that used longer treatment durations in terms of mean age, gender, recruitment status, or control group, so analyses with covariates were not conducted.

Study attrition rates—Attrition rates varied from 0% to 71%, with a mean of 12.8% and a median of 9.5%. The bivariate correlation between attrition rate and overall ES was nonsignificant ($r = -.03$, $p = .88$).

Peer-reviewed studies and non-peer-reviewed studies—Of the 35 studies, 27 were peer-reviewed, and these had a mean ES of 0.34, significantly different from zero ($z = 4.38$, $p < .01$). The 8 non-peer-reviewed studies had a mean ES of 0.32, only marginally different from zero (due in part to small sample size; $z = 1.66$, $p = .10$). A direct test of the ES difference between peer-reviewed and non-peer-reviewed studies was not significant, $Q(1, 33) = 0.01$, $p = .92$. Peer-reviewed studies included fewer boys in their samples than did non-peer-reviewed studies, $t(32) = 2.98$, $p < .01$, but when percentage of boys was included as a covariate in regression analyses, the effect of publication status remained non-significant ($B = 0.16$, $SE = 0.24$, $p = .50$).

Youth self-report versus parent-report measures—We compared posttreatment ES values for youth versus parent report using only studies that included both youth- and parent-report depression measures ($n = 6$). A paired-sample t test showed a significant youth–parent difference for depression measures, $t(5) = 2.56$, $p = .05$, Ms = 0.72 for youth report, 0.24 for parent report. The mean ES for youth report was significantly different from zero ($z = 4.70$, $p < .01$), but the mean ES for parent report was not ($z = 0.49$, $p = .62$).

Treatment Effects on Suicidality

Given current concerns about suicide risk in youth depression treatment, we focused on the six studies in our collection that included a measure of suicidality (e.g., the suicide symptom total from a diagnostic interview or a questionnaire designed to tap suicidal thinking and behavior). The average ES for these suicidality measures across the 6 studies was 0.18, marginally different from zero ($z = 1.80$, $p = .07$).

Why Was Mean ES Lower Than in Prior Youth Depression Meta-Analyses?

Finally, we turned to the question of why mean ES in our analyses (0.34) was so much lower than in the three prior meta-analyses of youth depression treatment trials, discussed in the introduction (0.99). It is not possible to discern all the reasons for this discrepancy, given the numerous differences that inevitably exist between any two meta-analyses (e.g., inclusion criteria, samples of studies used, procedures used to calculate ES). However, we sought to shed some light by examining the different collections of studies and considering what ES values might emerge with changes in procedure.

Comparing with Reinecke et al. (1998a, 1998b)—Our ES mean of 0.34 was markedly lower than the mean reported by Reinecke et al.: 1.02 (range = 0.40–1.85, 95% confidence interval [CI] = 0.81–1.23) in Reinecke et al.'s (1998a) study and 0.97 in Reinecke et al.'s (1998b) study. When we applied our data-analytic methods to the sample of six studies included in Reinecke et al.'s meta-analysis, we found a mean ES of 0.87 (WLS)/0.88 (ULS). When we shifted to the methods of ES calculation and data analysis described in the Reinecke et al.'s two articles (e.g., changing from our ES formula based on control group standard deviation to Reinecke's formula based on pooled treatment group plus control group standard deviation), we found a mean ES of 0.98 (ULS)/0.96 (WLS), almost identical to the figures reported by Reinecke et al. Thus, the difference between our mean ES of 0.40 and the much higher mean ES reported by Reinecke et al. appears to relate primarily to the difference between their small, select sample of six studies and our more broadly representative sample of 35 studies (e.g., including peer-reviewed and non-peer-reviewed trials, which used any form of psychotherapy).

Comparing with Lewinsohn and Clarke (1999)—Our ES mean of 0.34 differed most from that reported by Lewinsohn and Clarke, which was 1.27 (range and CI not reported). Using our methodology to calculate ES for the 12 studies cited by Lewinsohn and Clarke that included treatment versus control comparisons, we found a mean ES of 0.49 (WLS)/0.55 (ULS), closer to the ES in the current study. As in our comparison with Reinecke et al.

(1998a, 1998b), we sought to compute an ES for Lewinsohn and Clarke's study set on the basis of their exact methods of ES calculation and data analysis; however, information on their ES calculation and analytic procedures was not available from the article or from the authors. Nonetheless, our findings with their pool of studies do suggest that much of the difference between their mean ES of 1.27 and our much more modest mean of 0.34 was due to differences between their ES calculation methods and ours, with a smaller additional portion due to differences between the two samples of studies used.

Comparing with Michael and Crowley (2002)—Michael and Crowley reported a mean ES of 0.72 (range = 0.03–1.84, 95% CI = 0.48–0.94), on the basis of averaging 21 treatment effects reported in 14 articles and dissertations, using treatment group as the level of analysis. Their mean ES is larger than the unweighted ES values our meta-analysis generated (0.46 treatment level and 0.40 study level) and may differ both because of the different pool of studies used and differences in our methods of computing ES. Our methods differed in at least three ways from those of Michael and Crowley: (a) Our list of depression outcome measures differed slightly from theirs in that we included the standard Child Behavior Checklist anxious–depressed scale, whereas they used a depression-specific Child Behavior Checklist scale (see Clarke, Lewinsohn, Hops, & Seeley, 1992); (b) we computed ES using the posttest standard deviation of the control group only, whereas they pooled standard deviations of experimental and control groups at pretreatment and the control group at posttreatment; and (c) we used Hedges's correction for small sample size in computing each study's ES, whereas they did not. Calculating ES with their pool of studies and procedures but instead using each study as the level of analysis would have yielded an overall ES of 0.61 (Kurt Michael, personal communication, October 21, 2004). When we used the same set of studies as Michael and Crowley (2002) but applied our method of ES calculation (including using each study as an observation, rather than each treatment group), we found a mean ES of 0.48 (ULS)/0.38 (WLS), only modestly larger than the ES means we had calculated for our pool of 35 studies. Our calculated ES for Michael and Crowley's pool of studies with ESs averaged across treatments is 0.57 (WLS)/0.56 (ULS). Thus, it appears that some of the difference between Michael and Crowley's relatively larger mean ES and our smaller mean may reflect differences in the study collections used, but a larger portion of the difference is attributable to different data-analytic methods.

Discussion

We assembled the most comprehensive collection to date of youth depression treatment trials (both peer-reviewed and non-peer-reviewed), required random assignment for inclusion, contacted study authors repeatedly until we had obtained all data needed for the most precise ES calculation, and applied stringent meta-analytic methods (e.g., depression measures only, one mean ES per study, correction for small samples, weighted ES, and random effects analyses) to the data we obtained. The meta-analytic findings that emerged differed from those of previous reports in some very significant ways.

Perhaps the most striking difference between our findings and those of previous meta-analyses concerned the overall magnitude of treatment benefit. The mean effect of psychotherapy in our analyses was 0.34, falling between Cohen's (1988) benchmarks for a small (i.e., 0.20) and medium (i.e., 0.50) effect. Psychotherapy effects in previous youth depression meta-analyses (Lewinsohn & Clarke, 1999; Michael & Crowley, 2002; Reinecke et al., 1998a, 1998b) had averaged 0.99, comparing favorably with Cohen's benchmark of 0.80 for a large effect. The surprisingly modest treatment effect evident in our analyses suggests a new perspective on the success of youth depression psychotherapy. Our findings—including our direct comparison with previous meta-analytic findings with problems and disorders other than depression—

indicate that youth depression treatment does not surpass but instead may lag significantly behind treatments for other youth conditions.

Such an inference would need to be considered with caution. The years in which the depression and nondepression treatment studies were published overlapped substantially but not perfectly, which could complicate the comparison if year of publication were associated with ES; however, such an association is not evident in the youth treatment literature (see Weisz et al., 1995). The depression versus nondepression study comparison did control for multiple factors that previous literature suggests might explain differences between different collections of studies (mean age, gender distribution, recruited vs. referred youths, active vs. passive control group, methods of ES calculation, and peer-reviewed studies vs. non-peer-reviewed); these factors did not account for the difference between depression and nondepression studies. However, different collections of studies will inevitably differ in diverse ways, so it is possible that factors not identified and controlled for in our analysis might account for the ES difference between depression and nondepression studies.

In light of our finding that psychotherapies for youth depression have relatively modest mean effects, several potentially useful next steps might be considered. These could include (a) strengthening the substance or ramping up the dose of current treatments, (b) combining currently separate depression treatments into more potent multicomponent packages, and (c) developing and testing entirely new methods that produce more substantial benefit. That said, it is important to note that ES values showed a broad range across studies in our collection; indeed, five different treatment programs generated effects exceeding 1.0. Thus, some treatments in the current armamentarium may already have strong potential.

In this connection, it must be noted that the strongest potential may not attach to the most popular treatments. In the current zeitgeist, treatments that focus on altering unrealistic, negative cognitions have particularly prominent status. Indeed, 33 of the 44 treatments in our study set emphasized cognitive change (i.e., through CBT or other cognitive approaches). This broad approach is also popular in adult depression treatment; however, some of the most provocative adult research (see, e.g., Hollon, 2000; Jacobson et al., 1996, 2001) has highlighted the potential of noncognitive behavioral-activation strategies, providing evidence that the impact of such strategies is not improved on by treatment with a cognitive focus. Our analyses of youth treatment evidence indicated, similarly, that noncognitive treatments demonstrated effects that were easily as robust as the cognitive treatments, suggesting that beneficial treatment for youth depression may not require altering cognitions.

Although our overall mean ES for psychotherapy was much lower than in previous reports, it was significantly different from zero, suggesting reliable treatment effects across the group of studies as a whole. However, these effects proved durable only in the relative short-term. ES at follow-up periods of 1 year or longer showed no lasting treatment effect. This supports the potential value of booster sessions and continuation treatment (Clarke et al., 1999; Weissman, 1994) in extending treatment benefit over time. However, two caveats should be noted. First, more than one third of the studies reviewed did not include follow-up assessments with treatment versus control comparisons; we do not know how lasting effects were in those studies. Second, only five studies included follow-up at 1 year or beyond; this limits our ability to generalize, and it highlights the need for studies with longer term follow-up.

We also assessed the generality versus specificity of treatment effects, investigating the extent to which effects on depression-related outcome measures were replicated with nondepression outcome measures. Previous meta-analytic findings on generality–specificity across an array of youth treatments and treated problems (Weisz et al., 1995) had shown significant treatment effects on both targeted and nontargeted outcomes but with effects stronger for targeted than

nontargeted outcomes, suggesting specificity of benefit. However, the previous work had not focused on depression in particular or on the question of whether carryover effects might depend on conceptual similarity between depression and the outcomes being measured. When we addressed this question here, we found evidence of both generality and specificity in treatment effects. Depression treatment was associated with significant improvement in the conceptually similar domain of anxiety. In fact, following depression treatment, the reduction in anxiety symptoms was only marginally lower than the reduction in depressive symptoms. By contrast, we found that effects for the conceptually dissimilar domain of externalizing problems were significantly inferior to effects on depression measures, and that the mean effect on externalizing outcomes was not significantly different from zero. Our finding that depression treatment has beneficial effects on anxiety is consistent with growing evidence that youth depression and anxiety are closely associated empirically (see, e.g., Achenbach & Rescorla, 2001) and that they share a common core of negative affectivity (see, e.g., Cole, Peeke, Martin, Truglio, & Seroczynski, 1998; King, Ollendick, & Gullone, 1991). A useful question for future research is whether the effects of youth depression treatment on anxiety result from increased skill in addressing the negative affectivity that is apparently shared by the two syndromes. Whatever the answer to this question, the findings do offer some support for the possibility that youth depression and anxiety might be treated by a common intervention encompassing emotional disorders (see Barlow et al., 2004).

Taken together, these findings on the magnitude and specificity of effects may help inform the debate over alternatives to antidepressant medication, as discussed in the introduction (see Glass, 2004; Safer, 1997; TADS Team, 2004; Vitiello & Swedo, 2004; Weisz & Jensen, 1999; Whittington et al., 2004). Our results suggest that for those who seek an alternative to antidepressants, psychotherapy offers a reasonable option, generating a small to medium ES that generalizes to comorbid anxiety symptoms and shows substantial holding power for some months after treatment ends. Because recent concerns over SSRIs relate to elevated risk of suicidality, it may warrant attention that our study set included six investigations that assessed suicidality as an outcome, and that these studies averaged a small reduction in suicidality (mean ES = 0.18, marginally greater than zero).

As another perspective on these issues, one might construe psychotherapy as a potentially useful complement to, rather than a replacement for, antidepressants—that is, a form of intervention that may boost outcomes when combined with medication. This perspective is consistent with the findings of the most complete and sophisticated direct comparison, to date, of medication to psychotherapy in youth depression treatment—that is, the TADS (see TADS Team, 2004). In this study, adolescents treated with fluoxetine alone showed outcomes superior to those in a placebo condition, but adolescents treated with a combination of fluoxetine and a 12-week course of CBT showed the most positive treatment response, supporting the idea that psychotherapy may complement the effects of antidepressant medication. An important additional finding was that CBT alone did not significantly outperform the placebo condition, supporting concerns that psychotherapy alone (at least in its CBT form) may not be a very potent treatment force. If this finding were taken as definitive evidence on the potential of CBT, then the results could be quite discouraging to those who seek a psychotherapeutic alternative to medication. However, a close look at Table 1 indicates that the CBT ES generated in TADS is not characteristic of most CBT or psychotherapy effects on youth depression; 20 of the 23 other CBT programs in the table showed larger ES than the TADS version of CBT, and the mean ES value across the non-TADS CBT programs in the table was 0.48, markedly higher than the −0.07 ES associated with the TADS CBT intervention. What is not clear from the available data is whether this picture results from a low-potency version of CBT in TADS, from the unusual and challenging comparison of CBT with a medication placebo condition in TADS (see Baskin et al., 2003), from a combination of the two, or from other factors not identified.

A concern raised by some (e.g., Jensen, 2003; Weisz, 2004) about the evidence on youth treatment research in general is that so many of the trials have compared active treatment with passive control conditions, including no treatment and waitlist. This was evident in the current depression study set as well, with 20 of the 35 studies having used passive control conditions. Those studies showed a relatively strong treatment effect (mean ES = 0.41), significantly superior to zero at the .01 level. In contrast, the 14 studies comparing treatment with an active control group generated a mean ES of only 0.24, markedly lower but still superior to zero. Thus, our findings showed rather modest benefits of depression treatment when compared with the most rigorous comparison conditions (see Baskin et al., 2003). As Jensen (2003) stressed, we need studies that use "control groups comparable in intensity of exposure to the supposed active treatment. Such studies are critical, if we are to conclude that something about a given therapy is *specifically* effective, over and above simple compassion, friendliness, attention, and belief" (p. 37). The fact that passive control groups generate higher ES may also help explain why previous youth depression treatment meta-analyses have yielded higher mean ES than the current one, because the study collections in those prior meta-analyses involved somewhat heavier reliance than the current meta-analysis on no-treatment and waitlist comparison groups. Specifically, 40% of the studies in our meta-analysis used active control groups, in contrast to 26% averaging across the three previous youth depression meta-analyses —that is, 14% in Michael and Crowley's (2002) meta-analysis, 33% in Reinecke et al.'s (1998a, 1998b) meta-analysis, and 33% in Lewinsohn and Clarke's (1999) meta-analysis.

In the debate over empirically supported treatments, concern has been raised that the empirical support comes largely from efficacy studies in which experimental control is achieved at the cost of clinical representativeness (e.g., Weisz, 2004; Westen et al., 2004). One result, the argument goes, is that for many treatment programs we do not know whether the procedures actually work with clinically referred youths, treated by clinical practitioners, in clinical practice settings. Our findings may alleviate some of these concerns to some degree. Although we did find that ES values were somewhat larger for research therapists than for clinical practitioner therapists, and somewhat larger for treatments delivered in research settings than in service settings, neither difference was significant. Moreover, ES was reliably superior to zero for referred youths, practitioner therapists, and clinical service settings, suggesting that significant treatment benefit can be obtained across all three clinical representativeness dimensions.

Despite the modest overall treatment effects evident across the depression trials, we found that treatment benefit proved rather robust across some notable variations in person and treatment characteristics. For example, significant treatment effects were identified for (a) both child-majority and adolescent-majority samples considered separately, (b) samples identified as having depressive disorders and samples identified through depression symptom measures, (c) both group and individual treatments, and (d) treatments with and without a cognitive emphasis. In addition, treatment duration was not correlated with outcome, suggesting that some briefer treatments may have the potential to be as effective as lengthier ones. Thus, the benefits of psychotherapy, though modest on average, were evident across rather diverse characteristics of treated youths and across variations in the format, content, and duration of therapy.

We also found effects to be rather consistent across published and non-peer-reviewed studies, suggesting that publication bias may not be a major problem in the youth depression treatment literature thus far. In the youth psychotherapy literature generally, unpublished/non-peer-reviewed studies show significantly lower ES than published studies (see McLeod & Weisz, 2004). However, most youth depression psychotherapy research is relatively recent compared with treatment research with other youth conditions (see Weisz, Hawley, & Jensen Doss, 2004); more recent research may profit from an increased focus, in journal reviews, on the

quality of the research procedures rather than on the statistical significance of intervention effects.

Our findings shed light on another area of discussion among youth depression treatment experts: intervention outcome as perceived by different informants. We found that depression-related outcomes looked significantly better when the outcome information was provided by the youngsters themselves (e.g., via symptom self-report measures) than when their parents provided the information. Moreover, youth-report outcomes were significantly better than zero, whereas parent-report depression outcomes were not. Such a finding may reflect the fact that youths themselves have better access to information on their own internal state than do outside observers (see Hammen & Rudolph, 1996). Collateral reporters rather consistently report lower levels of depression than children themselves (Angold & Costello, 1993; Capaldi & Stoolmiller, 1999; Hammen & Rudolph, 1996). It is possible that parents' difficulty in evaluating their children's internal states may make them relatively insensitive to the changes that would need to be noted to detect improvement at the end of treatment. It may also be relevant that parents of depressed children are more likely than other parents to be depressed themselves (Kovacs & Devlin, 1998); some have argued that relatively depressed mothers may perceive their children's behavior in a more negative light than it actually warrants (Breslau, Davis, & Prabucki, 1988), but because maternal depression is in fact associated with more actual child disorder, it is not clear whether bias is involved (Boyle & Pickles, 1997; Richters, 1992). Whatever the reason for our findings, they do raise a concern that the evidence for beneficial effects of psychotherapy for youth depression rests almost entirely on reports by the youths themselves without confirmation from other more objective informants.

Our findings, together with the scrutiny of studies required for this meta-analysis, suggest several observations about the state of the evidence and ways to improve it. As noted previously, we need more studies designed to test whether specific depression treatments can outperform active conditions that control for attention and other nonspecific factors. In addition, the fact that more than one third of the studies in our collection generated no usable follow-up comparisons of treatment and control conditions is a reminder that we need more studies that include follow-up assessments, and in which the control condition remains unaltered throughout the follow-up period, so that treatment–control comparisons can be meaningful at the time of follow-up. The episodic nature of depression may also argue for extending the lag between posttreatment and follow-up. An important counter to this point is that maintaining control conditions for extended periods raises ethical concerns, if doing so exposes depressed youths to long periods without treatment. This point, in our view, underscores the potential value of the treatment-as-usual control condition (see, e.g., Weisz, 2004). Ethical concerns should not attach to procedures that provide youths with the intervention they would have received in the absence of the study. This suggestion and the previous one are quite compatible; no treatment and waitlist, currently the most commonly used control conditions, are not only the weakest experimentally but also the most difficult to sustain throughout a waitlist period, given ethical and humane concerns.

Another design limitation evident in the studies that we reviewed was the relative absence of intent-to-treat analyses (only 11 studies explicitly reported such analyses), a state of affairs that constrains interpretation of positive effects. Without such analyses, one cannot rule out the possibility that youths who dropped out of treatment (and were thus dropped from analyses) did so because they were not benefiting from treatment, and that the resulting ES values are overestimates.

Two final areas of concern involve the search for moderators and mediators of treatment impact (see discussion in the following books: Kazdin, 2000; Weisz, 2004). On the moderator front, many of the studies can be faulted for a failure to characterize the samples fully enough to

address the role of participant characteristics. As an example, only 13 of the 35 studies provided detailed information on the race/ethnicity of their samples, and only 6 of the 35 studies included any test of any potential moderator. Even more striking is the relative inattention to the question of what change processes underlie improvement. In only one of the studies was a candidate mediator or change process identified and its mediating role tested.

Taken together, our findings and our observations on the evidence suggest an agenda for future research on youth depression treatment. Clearly, a useful foundation has been laid, with evidence from 35 studies pointing to treatment effects that are significant, albeit markedly more modest than those reported in previous meta-analyses. Effects appear to be durable for the initial months following treatment but not when followed for 1 year or more; however, more evidence is needed regarding long-term holding power. Critical examination of the evidence suggests a need for increased use of active control conditions, meaningful follow-up assessment, intent-to-treat analysis, moderator assessment, and tests of proposed mechanisms of change. Much has been accomplished in 25 years of youth depression treatment research, but important work remains for the years ahead.

Acknowledgements

References

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

Achenbach T. "Comorbidity" in child and adolescent psychiatry: Categorical and quantitative perspectives. Journal of Child and Adolescent Psychopharmacology 1990;1:271–278.

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families; 2001.

*Ackerson J, Scogin F, McKendree-Smith N, Lyman RD. Cognitive bibliotherapy for mild and moderate adolescent depressive symptomatology. Journal of Consulting and Clinical Psychology 1998;66:685–690. [PubMed: 9735587]

Angold A, Costello EJ. Depressive comorbidity in children and adolescents: Empirical, theoretical, and methodological issues. American Journal of Psychiatry 1993;150:1779–1791. [PubMed: 8238631]

Angold A, Messer SC, Stangl D, Farmer EM, Costello EJ, Burns BJ. Perceived parental burden and service use for child and adolescent psychiatric disorders. American Journal of Public Health 1998;88:75–80. [PubMed: 9584037]

Arias E, MacDorman MF, Strobino DM, Guyer B. Annual summary of vital statistics—2002. Pediatrics 2003;112:1215–1230. [PubMed: 14654589]

*Asarnow JR, Scott CV, Mintz J. A combined cognitive–behavioral family education intervention for depression in children: A treatment development study. Cognitive Therapy and Research 2002;26:221–229.

Barlow DH, Allen LB, Choate ML. Toward a unified treatment for emotional disorders. Behavior Therapy 2004;35:205–230.

Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. Journal of Consulting and Clinical Psychology 2003;71:973–979. [PubMed: 14622072]

Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J, Dahl RE, et al. Childhood and adolescent depression: A review of the past 10 years, Part I. Journal of the American Academy of Child & Adolescent Psychiatry 1996;35:1427–1439. [PubMed: 8936909]

Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.

Boyle M, Pickles A. Influence of maternal depressive symptoms on ratings of childhood behavior. Journal of Abnormal Child Psychology 1997;25:399–412. [PubMed: 9421748]

*Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, Roth C, et al. A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Archives of General Psychiatry 1997;54:877–885. [PubMed: 9294380]

Breslau N, Davis G, Prabucki K. Depressed mothers as informants in family research—Are they accurate? Psychiatry Research 1988;24:345–359. [PubMed: 3406244]

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

Capaldi DM, Stoolmiller M. Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: III. Prediction to young-adult adjustment. Development & Psychopathology 1999;11:59–84. [PubMed: 10208356]

Clark LA, Watson D, Mineka S. Temperament, personality, and the mood and anxiety disorders. Journal of Abnormal Psychology 1994;103:103–116. [PubMed: 8040472]

Clarke, GN.; DeBar, LL.; Lewinsohn, PM. Cognitive–behavioral group treatment for adolescent depression. In: Kazdin, AE., editor. Evidenced-based psychotherapies for children and adolescents. New York: Guilford Press; 2003. p. 120-134.

*Clarke GN, Hawkins W, Murphy M, Sheeber LB, Lewinsohn PM, Seeley JR. Targeted prevention of unipolar depressive disorder in an at-risk sample of high school adolescents: A randomized trial of a group cognitive intervention. Journal of the American Academy of Child & Adolescent Psychiatry 1995;34:312–321. [PubMed: 7896672]

*Clarke GN, Hornbrook M, Lynch F, Polen M, Gale J, Beardslee W, et al. A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. Archives of General Psychiatry 2001;58:1127–1134. [PubMed: 11735841]

*Clarke GN, Hornbrook M, Lynch F, Polen M, Gale J, O'Connor E, et al. Group cognitive–behavioral treatment for depressed adolescent offspring of depressed parents in a health maintenance organization. Journal of the American Academy of Child & Adolescent Psychiatry 2002;41:305–313. [PubMed: 11886025]

Clarke GN, Lewinsohn PM, Hops H, Seeley JR. A self-and parent-report measure of adolescent depression: The Child Behavior Checklist Depression scale (CBCL-D). Behavioral Assessment 1992;14:443–463.

*Clarke GN, Rohde P, Lewinsohn PM, Hops H, Seeley JR. Cognitive–behavioral treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. Journal of the American Academy of Child & Adolescent Psychiatry 1999;38:272–279. [PubMed: 10087688]

Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: Erlbaum; 1988.

Cohen J. Things I have learned (so far). American Psychologist 1990;45:1304–1312.

Cole DA, Peeke L, Martin JM, Truglio R, Seroczynski AD. A longitudinal look at the relation between depression and anxiety in children and adolescents. Journal of Consulting and Clinical Psychology 1998;66:451–460. [PubMed: 9642883]

Committee on Safety of Medicines. Use of selective serotonin reuptake inhibitors in children and adolescents with major depressive disorder. 2004. Retrieved November 18, 2005, from http://www.antidepressantsfacts.com/2003–12-10-CSM-SSRIs-children-UK_bestanden/csmhomemain.htm

Cook D, Guyatt G, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should unpublished data be included in meta-analyses? Journal of the American Medical Association 1993;269:2749–2753. [PubMed: 8492400]

Costello EJ, Pine DS, Hammen C, March JS, Plotsky PM, Weissman MM, et al. Development and natural history of mood disorders. Biological Psychiatry 2002;52:529–542. [PubMed: 12361667]

*Curtis, SE. Cognitive–behavioral treatment of adolescent depression Unpublished doctoral dissertation. Utah State University; Logan: 1992.

*Dana, E. A cognitive-behavioral intervention for conduct disordered and concurrently conduct disordered and depressed children. Unpublished doctoral dissertation. Adelphi University School of Social Work; Garden City, New York: 1998.

*De Cuyper S, Timbremont B, Braet C, De Backer V, Wullaert T. Treating depressive symptoms in school children: A pilot study. Journal of European Child and Adolescent Psychiatry 2004;13:105–114.

*Diamond GS, Reis BF, Diamond GM, Siqueland L, Isaacs L. Attachment-based family therapy for depressed adolescents: A treatment development study. Journal of the American Academy of Child & Adolescent Psychiatry 2002;41:1190–1196. [PubMed: 12364840]

*Ettelson, RG. The treatment of adolescent depression Unpublished doctoral dissertation. Illinois State University; Normal: 2003.

*Fischer, SA. Development and evaluation of group cognitive–behavioral therapy for depressed and suicidal adolescents in juvenile detention. Unpublished doctoral dissertation. University of Alabama; Tuscaloosa: 1995.

Glass RM. Treatment of adolescents with major depression. Journal of the American Medical Association 2004;292:861–863. [PubMed: 15316001]

Gould MS, King R, Greenwald S, Fisher P, Schwab-Stone M, Kramer R, et al. Psychopathology associated with suicidal ideation and behavior among children and adolescents. Journal of the American Academy of Child & Adolescent Psychiatry 1998;37:915–923. [PubMed: 9735611]

Hammen, C.; Rudolph, KD. Childhood depression. In: Mash, EJ.; Barkley, RA., editors. Child psychopathology. New York: Guilford Press; 1996. p. 153-195.

Hammen C, Rudolph K, Weisz JR, Burge D, Rao U. The context of depression in clinic-referred youth: Neglected areas in treatment. Journal of the American Academy of Child & Adolescent Psychiatry 1999;38:64–71. [PubMed: 9893418]

Harrington R, Campbell F, Shoebridge P, Whittaker J. Meta-analysis of CBT for depression in adolescents. Journal of the American Academy of Child & Adolescent Psychiatry 1998;37:1005–1006. [PubMed: 9785707]

Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. San Diego, CA: Academic Press; 1985.

Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychological Methods 1998;3:486–504.

*Hickman, KA. Effects of social skills training on depressed children attending a behavioral day treatment program. Unpublished doctoral dissertation. Hofstra University; Hempstead, New York: 1994.

Hollon SD. Do cognitive change strategies matter in cognitive therapy? Prevention & Treatment 2000;3:1–5.

Horvath AO, Luborsky L. The role of the therapeutic alliance and outcome in psychotherapy: A meta-analysis. Journal of Consulting and Clinical Psychology 1993;61:561–573. [PubMed: 8370852]

Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, et al. A component analysis of cognitive–behavioral treatment for depression. Journal of Consulting and Clinical Psychology 1996;64:295–304. [PubMed: 8871414]

Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: Returning to contextual roots. Clinical Psychology: Science and Practice 2001;8:255–270.

Jensen PS. What is the evidence for evidence-based treatments? Emotional & Behavioral Disorders in Youth 2003;3:37–41.

*Kahn JS, Kehle TJ, Jenson WR, Clark E. Comparison of cognitive–behavioral, relaxation, and self-modeling interventions for depression among middle-school students. School Psychology Review 1990;19:196–211.

*Kahn, RHC. The effect of a group support intervention program on depression, social adjustment, and self-esteem. Unpublished doctoral dissertation. Catholic University of America; Washington, DC: 1989.

Kazdin AE. Developing a research agenda for child and adolescent psychotherapy. Archives of General Psychiatry 2000;57:829–835. [PubMed: 10986544]

Kazdin AE, Bass D, Ayers WA, Rodgers A. Empirical and clinical focus of child and adolescent psychotherapy research. Journal of Consulting and Clinical Psychology 1990;58:729–740. [PubMed: 2292623]

*Kerfoot M, Harrington R, Harrington V, Rogers J, Verduyn C. A step too far? Randomized trial of cognitive–behaviour therapy delivered by social workers to depressed adolescents. European Child and Adolescent Psychiatry 2004;13:92–99. [PubMed: 15103534]

King NJ, Ollendick TH, Gullone E. Negative affectivity in children and adolescents: Relations between anxiety and depression. Clinical Psychology Review 1991;11:441–459.

Kovacs M, Devlin B. Internalizing disorders in children. Journal of Child Psychology and Psychiatry and Allied Disciplines 1998;39:47–63.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174. [PubMed: 843571]

Lewinsohn PM, Clarke GN. Psychosocial treatments for adolescent depression. Clinical Psychology Review 1999;19:329–342. [PubMed: 10097874]

*Lewinsohn PM, Clarke GN, Hops H, Andrews J. Cognitive–behavioral treatment for depressed adolescents. Behavior Therapy 1990;21:385–401.

Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major depression in community adolescents: Age at onset, episode duration, and time to recurrence. Journal of the American Academy of Child & Adolescent Psychiatry 1994;33:809–818. [PubMed: 7598758]

*Liddle B, Spence SH. Cognitive–behaviour therapy with depressed primary school children: A cautionary note. Behavioural Psychotherapy 1990;18:85–102.

Lipsey, MW.; Wilson, DB. Practical meta-analysis. 49. Thousand Oaks, CA: Sage; 2001. Applied Social Research Methods Series

*Marcotte D, Baron P. L'efficacite d'une strategie d'intervention emotivorationnelle aupres d'adolescents depresssifs du milieu scolaire [The efficacy of a school-based rational-emotive intervention strategy with depressive adolescents]. Canadian Journal of Counselling 1993;27:77–92.

McLeod BD, Weisz JR. Using dissertations to examine potential bias in child and adolescent clinical trials. Journal of Consulting and Clinical Psychology 2004;72:235–251. [PubMed: 15065958]

McRoberts C, Burlingame GM, Hoag MJ. Comparative efficacy of individual and group psychotherapy: A meta-analytic perspective. Group Dynamics: Theory, Research, and Practice 1998;2:101–117.

Michael KD, Crowley SL. How effective are treatments for child and adolescent depression? A meta-analytic review. Clinical Psychology Review 2002;22:247–269. [PubMed: 11806021]

Michael KD, Huelsman TJ, Crowley SL. Interventions for child and adolescent depression: Do professional therapists produce better results? Journal of Child and Family Studies 2005;14:223–236.

*Moldenhauer, Z. Adolescent depression: A primary care pilot intervention study. Unpublished doctoral dissertation. University of Rochester; New York: 2004.

*Mufson L, Dorta KP, Wickramaratne P, Nomura Y, Olfson M, Weissman MM. A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. Archives of General Psychiatry 2004;61:577–584. [PubMed: 15184237]

*Mufson L, Weissman MM, Moreau D, Garfinkel R. Efficacy of interpersonal psychotherapy for depressed adolescents. Archives of General Psychiatry 1999;56:573–579. [PubMed: 10359475]

*Reed MK. Social skills training to reduce depression in adolescents. Adolescence 1994;29:293–302. [PubMed: 8085482]

Reinecke MA, Ryan NE, DuBois DL. Cognitive–behavioral therapy of depression and depressive symptoms during adolescence: A review and meta-analysis. Journal of the American Academy of Child & Adolescent Psychiatry 1998a;37:26–34. [PubMed: 9444896]

Reinecke MA, Ryan NE, DuBois DL. Meta-analysis of CBT for depression in adolescents: Dr. Reinecke et al. reply. Journal of the American Academy of Child & Adolescent Psychiatry 1998b;37:1006–1007.

*Reynolds WM, Coats KI. A comparison of cognitive–behavioral therapy and relaxation training for the treatment of depression in adolescents. Journal of Consulting and Clinical Psychology 1986;54:653–660. [PubMed: 3534032]

Richters JE. Depressed mothers as informants about their children: A critical review of the evidence for distortion. Psychological Bulletin 1992;112:485–499. [PubMed: 1438639]

*Roberts C, Kane R, Bishop B, Matthews H, Thomson H. The prevention of depressive symptoms in rural school children: A follow-up study. International Journal of Mental Health Promotion 2004;6:4–16.

*Roberts C, Kane R, Thomson H, Bishop B, Hart B. The prevention of depressive symptoms in rural school children: A randomized controlled trial. Journal of Consulting and Clinical Psychology 2003;71:622–628. [PubMed: 12795585]

*Rohde P, Clarke GN, Mace DE, Jorgensen JS, Seeley JR. An efficacy/effectiveness study of cognitive–behavioral treatment for adolescents with and without comorbid major depression and conduct disorder. Journal of the American Academy of Child & Adolescent Psychiatry 2004;43:660–668. [PubMed: 15167082]

Rohde P, Lewinsohn PM, Seeley JR. Comorbidity of unipolar depression: II. Comorbidity with other mental disorders in adolescent and adults. Journal of Abnormal Psychology 1991;100:314–322.

*Rosello J, Bernal G. The efficacy of cognitive–behavioral and interpersonal treatments for depression in Puerto Rican adolescents. Journal of Consulting and Clinical Psychology 1999;67:734–745. [PubMed: 10535240]

Rosenthal R. How are we doing in soft psychology? American Psychologist 1990;45:775–777.

Safer DJ. Changing patterns of psychotropic medications prescribed by child psychiatrists in the 1990s. Journal of Child and Adolescent Psychopharmacology 1997;7:267–274. [PubMed: 9542697]

Schmidt FL. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. American Psychologist 1992;47:1173–1181.

Smith, ML.; Glass, GV.; Miller, TL. The benefits of psychotherapy. Baltimore: Johns Hopkins University Press; 1980.

Sohn D. Publication bias and the evaluation of psychotherapy efficacy in reviews of the research literature. Clinical Psychology Review 1996;16:147–156.

Southam-Gerow MA, Weisz JR, Kendall PC. Youth with anxiety disorders in research and service clinics: Examining client differences and similarities. Journal of Clinical Child and Adolescent Psychology 2003;32:375–385. [PubMed: 12881026]

*Stark, KD. Childhood depression: School-based intervention. New York: Guilford Press; 1990.

*Stark KD, Reynolds WM, Kaslow NJ. 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 1987;15:91–113. [PubMed: 3571741]

*Treatment for Adolescents with Depression Study (TADS) Team. Fluoxetine, cognitive–behavioral therapy, and their combination for adolescents with depression. Journal of the American Medical Association 2004;292:807–820. [PubMed: 15315995]

U.S. Food and Drug Administration. Worsening depression and suicidality in patients treated with antidepressant medications. 2004. Retrieved December 12, 2004, from http://www.fda.gov/cder/drug/antidepressants/default.html

Vitiello B, Swedo S. Antidepressant medications in children. New England Journal of Medicine 2004;350:1489–1491. [PubMed: 15071123]

*Vostanis P, Feehan C, Grattan E, Bickerton WL. A randomised controlled out-patient trial of cognitive–behavioural treatment for children and adolescents with depression: 9-month follow-up. Journal of Affective Disorders 1996a;40:105–116. [PubMed: 8882920]

*Vostanis P, Feehan C, Grattan E, Bickerton WL. Treatment for children and adolescents with depression: Lessons from a controlled trial. Clinical Child Psychology and Psychiatry 1996b;1:199–212.

Weissman MM. Psychotherapy in the maintenance treatment of depression. British Journal of Psychiatry 1994;165:42–50.

Weissman MM, Wolk S, Goldstein RB, Moreau D, Adams P, Greenwald S, et al. Depressed adolescents grown up. Journal of the American Medical Association 1999;281:1707–1713. [PubMed: 10328070]

Weisz, JR. Psychotherapy for children and adolescents: Evidence-based treatments and case examples. Cambridge, England: Cambridge University Press; 2004.

Weisz JR, Hawley KM, Jensen Doss A. Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. Child and Adolescent Psychiatric Clinics of North America 2004;13:729–816. [PubMed: 15380784]

Weisz JR, Jensen PS. Efficacy and effectiveness of child and adolescent psychotherapy and pharmacotherapy. Mental Health Services Research 1999;1:125–157. [PubMed: 11258738]

*Weisz JR, Thurber CA, Sweeney L, Proffitt VD, LeGagnoux GL. Brief treatment of mild-to-moderate child depression using primary and secondary control enhancement training. Journal of Consulting and Clinical Psychology 1997;65:703–707. [PubMed: 9256573]

Weisz, JR.; Valeri, SM.; McCarty, CA.; Moore, PS. Interventions for child and adolescent depression: Features, effects, and future directions. In: Essau, CA.; Petermann, FP., editors. Depressive disorders in children and adolescents. Northvale, NJ: Jason Aronson; 1999. p. 383-436.

Weisz JR, Weiss B, Alicke MD, Klotz ML. Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. Journal of Consulting and Clinical Psychology 1987;55:542–549. [PubMed: 3624610]

Weisz JR, Weiss B, Han SS, Granger DA, Morton T. Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. Psychological Bulletin 1995;117:450–468. [PubMed: 7777649]

Westen D, Novotny CM, Thompson-Brenner H. Empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. Psychological Bulletin 2004;130:631–663. [PubMed: 15250817]

Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: Systematic review of published versus unpublished data. Lancet 2004 April 2004;363:1341–1345. [PubMed: 15110490]

Wilson, DB. SPSS macros. 2003. Retrieved April 1, 2003, from http://mason.gmu.edu/~dwilsonb/ma.html

*Wood A, Harrington R, Moore A. Controlled trial of a brief cognitive–behavioural intervention in adolescent patients with depressive disorders. Journal of Child Psychology and Psychiatry 1996;37:737–746. [PubMed: 8894955]

NIH-PA Author Manuscript

Table 1

Characteristics of Individual Studies Included in the Meta-Analysis

Study	Type of sample	Sample size[a]	Age group	% boys	Type of control group	Attrition (%)	Nondepression measures assessed	Overall ES	Depression ES
Ackerson et al. (1998)	Diagnosed community sample	22	14–18	36	Waitlist	27	Negative cognitions	1.52 cognitive bibliotherapy	1.63 cognitive bibliotherapy
Asarnow et al. (2002)	Symptomatic school sample	23	Grades 4–6	35	Waitlist	0	Negative cognitions	0.41 CBT with family component	.25 CBT with family component
Brent et al. (1997)	Diagnosed recruited plus referred outpatient sample	99	13–18	24	Nondirective supportive therapy	7	Functional impairment Suicidality	0.32 CBT; 0.15 systemic behavior family therapy	0.36 CBT; 0.18 systemic behavior family therapy
Butler et al. (1980)	Symptomatic school sample	41	Grades 5–6	53	Attention placebo	2	Self-esteem Locus of control Stimulus appraisal	0.52 role-play; 0.40 cognitive restructuring	0.28 role-play; 0.09 cognitive restructuring
Clarke et al. (1999)	Diagnosed recruited sample	96	14–18	29	Waitlist	22	Externalizing behavior Functional impairment	0.32 CBT; 0.15 CBT plus parent group	0.18 CBT; −0.14 CBT plus parent group
Clarke et al. (1995)	Symptomatic high school sample	120	Grades 9–10	30	No treatment	20	Functional impairment	0.30 cognitive restructuring	0.27 cognitive restructuring
Clarke et al. (2001)	Symptomatic at-risk sample	94	13–18	36	HMO usual care	6.3	Externalizing behavior Functional impairment Number of nondepression diagnoses Suicidality	0.11 cognitive restructuring	.09 cognitive restructuring
Clarke et al. (2002)	Diagnosed offspring of depressed parents from HMO	88	13–18	31	HMO usual care	2	Externalizing behavior Functional impairment Number of nondepression diagnoses Suicidality	0.003 CBT	0.02 CBT
Curtis (1992)	Diagnosed high school sample	19	High school	11	Waitlist	17	School performance Problem behaviors	1.03 CBT	2.02 CBT
Dana (1998)	Symptomatic special education youths	19	8–13	87	No treatment	0	Conduct problems Self-worth Social skills	−0.15 CBT	−0.10 CBT
De Cuyper et al. (2004)	Symptomatic school sample	21	9–11	25	Waitlist	9	Self-worth Anxiety	0.40 CBT	0.34 CBT
Diamond et al. (2002)	Diagnosed youths referred by parents or school	32	13–17	22	Waitlist	0	Anxiety Attachment Externalizing behavior Family functioning Hopelessness Suicidality	0.68 attachment-based family treatment	0.72 attachment-based family treatment
Ettelson (2003)	Symptomatic high school sample	25	Grades 9–12	44	Supportive contact waitlist	20	Anxiety Coping Negative cognitions	0.63 CBT	0.67 CBT
Fischer (1995)	Symptomatic youths in detention	16	12–17	88	Attention placebo	71	Coping Suicidality	0.91 CBT	0.07 CBT
Hickman (1994)	Diagnosed youths attending a day treatment program	9	8–11	89	Day treatment (treatment as usual)	0	Social skills	0.60 social skills training	0.43 social skills training
R. H. C. Kahn (1989)	Symptomatic youths abroad	30	15–17	47	No treatment	17	Social adjustment Self-concept	−0.44 group support	−0.46 group support
J. S. Kahn et al. (1990)	Symptomatic school sample	68	10–14	49	Waitlist	0	Self-concept	1.27 CBT; 0.91 relaxation; 0.85 self-modeling	1.34 CBT; 0.95 relaxation; 0.96 self-modeling
Kerfoot et al. (2004)	Symptomatic sample involved in social services	46		54	Treatment as usual	12	Global functioning	0.22 CBT	−0.07 CBT
Lewinsohn et al. (1990)	Diagnosed recruited sample	59	14–18	39	Waitlist	7	Anxiety Externalizing behavior Conflict resolution Negative cognitions	0.53 CBT; 0.73 CBT plus parent	0.68 CBT; 1.31 CBT plus parent

Study	Type of sample	Sample size[a]	Age group	% boys	Type of control group	Attrition (%)	Nondepression measures assessed	Overall ES	Depression ES
Liddle & Spence (1990)	Symptomatic school sample	21	7–11	68	Attention placebo	Not reported	Pleasant events Social skills Interpersonal difficulty	0.21 social competency training	0.67 social competency training
Marcotte & Baron (1993)	Symptomatic high school sample	25	14–17	24	Waitlist	11	Irrational beliefs	0.38 rationale motive	−0.27 rationale motive
Moldenhauer (2004)	Symptomatic primary care sample	19	12–17	27	Healthy lifestyle class	27	Anxiety Conflict resolution Family functioning Negative thoughts Self-esteem	0.06 CBT	−0.003 CBT
Mufson et al. (1999)	Diagnosed outpatient sample	32	12–18	27	Clinical monitoring	33	Social problem-solving Functional impairment	0.61 IPT	0.54 IPT
Mufson et al. (2004)	Diagnosed high school sample	63	12–18	16	Treatment as usual	10.9	Social adjustment Functional impairment	0.45 IPT	0.44 IPT
Reed (1994)	Diagnosed sample	17	14–19	50	Art and imagery group	6	Self-esteem	−0.56 structured learning therapy	−0.66 structured learning therapy
Reynolds & Coats (1986)	Symptomatic high school sample	24	14–17	37	Waitlist	20	Self-concept Anxiety Self-esteem	1.06 CBT; 1.33 relaxation	1.49 CBT; 1.48 relaxation
Roberts et al. (2003)	Symptomatic rural school sample	179	11–13	50	Usual health education class	5	Anxiety Externalizing behavior Negative cognitions Social skills	0.07 CBT	0.05 CBT
Rohde et al. (2004)	Diagnosed sample in juvenile justice system	91	13–17	52	Tutoring and life skills training	2	Social adjustment Externalizing behavior Functional impairment	0.24 CBT	0.38 CBT
Rosello & Bernal (1999)	Diagnosed, school-referred sample	58	13–17	46	Waitlist	18	Self-esteem Social adjustment Family criticism and overinvolvement Problem behavior	.11 CBT; 0.47 IPT	0.37 CBT; 0.72 IPT
Stark (1990) Stark et al. (1987)	Symptomatic school sample Symptomatic school sample	19 28	Grades 4–5 9–12	37 57	Traditional counseling Waitlist	10 3	Automatic thoughts Self-esteem Anxiety	0.61 CBT 0.61 self-control; 0.41 behavioral problem solving	0.65 CBT 0.63 self-control; 0.56 behavioral problem solving
Treatment for Adolescents with Depression Study (TADS) Team (2004)	Diagnosed outpatient sample	439	12–17	46	Medication placebo	10.9	Suicidality	0.09 CBT	−0.07 CBT
Vostanis et al. (1996a, 1996b)	Diagnosed outpatient sample	57	8–17	44	Nonfocused intervention	10	Anxiety Self-esteem Aggression Social adjustment	−0.04 CBT	0.24 CBT
Weisz et al. (1997)	Symptomatic school children	48	Grades 3–6	54	No treatment	None	None	0.45 CBT	0.45 CBT
Wood et al. (1996)	Diagnosed outpatient sample	48	9–17	38	Relaxation training	9	Anxiety Self-esteem Conduct problems Functional impairment	0.44 CBT	0.54 CBT

Note. All effect size (ES) values incorporate Hedges's correction for small sample size. CBT = cognitive–behavioral therapy; IPT = interpersonal psychotherapy.

[a]Sample size reflects the number of participants used to compute ESs at posttreatment. If intent-to-treat analyses were used, then all randomized participants were included in the *N*. When two control groups were present (e.g., Butler et al., 1980; Liddle & Spence, 1990), the attention-placebo group was used instead of the no-treatment control.

Table 2

Characteristics of the Psychotherapy Evidence Base for Treatment of Child and Adolescent Depression: 35 Randomized Trials

Variable	Results
% child (C), adolescent (A), and mixed (CA) Samples	C = 20; A = 60; CA = 20
% using diagnosed sample	43 (15/35)
% using placebo condition	43 (15/35)
% involving clinic-referred sample	17.1 (6/35)
% of treatment conducted in clinical setting	31.4 (11/35)
% of therapists with primary clinical vocation	25.7 (9/35)
No. of therapy hours	
M	13.3
SD	7.3
No. of depression outcome measures	
M	2.0^a
SD	0.9
No. of nondepression outcome measures	
M	$3.0,^{ab}$
SD	2.2
No. of different informants providing outcome information (e.g., child, parent, teacher)	
M	2.1
SD	0.7
Useable follow-up data available to compute ES (%)	54 (19/35)
Lag between posttreatment and follow- up (weeks)	
M	37.5
SD	41.8

Note. Treatment credibility, treatment satisfaction, and homework compliance measures were not considered outcome measures and thus are not included in the table. ES = effect size.

[a] If the same or parallel measure was completed by different informants (e.g., one version completed by child and another completed by parent), the number of outcome measures reflected the number of informants. Although a measure may be counted once in Table 2 (e.g., Child Behavior Checklist), for the ES comparisons, different ESs were computed for each subscale score that was reported (e.g., an ES for internalizing score and an ES for anxious/depressed score). Internalizing problems/scores were considered "other" problems.

[b] If follow-up assessments were conducted at two different time points, the final follow-up assessment was considered when computing the lag between posttreatment and follow-up.