Cognitive-Behavioral Therapy for Adolescent Depression: A Meta-Analytic Investigation of Changes in Effect-Size Estimates

Jesse B. Klein, B.A., **Rachel H. Jacobs, B.A.**, and **Mark A. Reinecke, Ph.D.**
Department of Psychiatry, Feinberg School of Medicine, Northwestern University, Chicago

Abstract

Objective—To explicate differences between early and recent meta-analytic estimates of the effects of cognitive-behavioral therapy (CBT) for adolescent depression.

Method—Meta-analytic procedures were used to investigate whether methodological characteristics moderated mean effect sizes among 11 randomized, controlled trials of CBT focusing on adolescents meeting diagnostic criteria for unipolar depression.

Results—Cumulative meta-analyses indicated that effects of CBT have decreased from large effects in early trials, and confidence intervals have become narrower. Effect sizes were significantly smaller among studies that used intent-to-treat analytic strategies, compared CBT to active treatments, were conducted in clinical settings, and featured greater methodological rigor based on CONSORT (Consolidated Standards of Reporting Trials) criteria. The mean posttreatment effect size of 0.53 was statistically significant.

Conclusions—Differences in estimates of the efficacy of CBT for depressed adolescents may stem from methodological differences between early and more recent investigations. Overall, results support the effectiveness of CBT for the treatment of adolescent depression.

Keywords

depression; cognitive-behavioral therapy; meta-analysis

Estimates of the efficacy of clinical interventions often change with the accumulation of outcome data (Ioannidis and Lau, 2001). A recent example of this can be found in the contrast between meta-analytic findings reported by Weisz et al. (2006) in their evaluation of the effects of psychotherapy for depression among youths (effect size [ES] = 0.34 across 31 randomized, controlled trials [RCTs] of treatments that were primarily cognitive-behavioral in nature) and earlier estimates of the effects of cognitive-behavioral therapy (CBT) for this population (ES = 1.27 [Lewinsohn and Clarke, 1999]; ES = 1.02 [Reinecke et al., 1998]). With the accumulation of outcome data for a particular treatment, mean ESs often approach the population parameter. However, heterogeneity in subject, design, and methodological features of trials may also cause large ES fluctuations. Weisz et al. (2006) investigated this possibility by evaluating the effects of 13 potential moderator variables. They found that treatment effects were moderated only by outcome measure informant, with youths reporting greater clinical improvement than parents.

Although the meta-analysis of Weisz et al. (2006) is notable for its broad scope, it may be necessary to adopt a narrower focus to detect moderator effects and thereby explicate reasons

Correspondence to Jesse Klein, B.A., Department of Psychiatry and Behavioral Sciences, Northwestern University, Feinberg School of Medicine, 710 North Lake Shore Drive, Abbott Hall, Suite 1205, Chicago, IL 60611; e-mail: jesseklein@northwestern.edu.

for the reduced ESs observed. The analyses conducted by Weisz et al. (2006) included data from studies that focused on prevention and treatment of depression among youths presenting with a broad range of symptom severity and impairment. It remains possible, however, that moderation may be apparent in a narrower set of studies of treatment for clinically depressed youths. Accordingly, to explain the observed decline in the effects of CBT, we focus specifically on evaluating treatment effects across studies involving youths meeting diagnostic criteria for unipolar depressive disorders.

Methodologies used in RCTs of CBT for children and adolescents with depressive disorders have recently undergone several changes that may explain the apparent reduction in the effectiveness of treatment. These changes reflect a natural progression in research methodology wherein case studies and uncontrolled open trials are followed by RCTs with increasingly complex patients, more stringent controls, and more sophisticated methodologies. Typically, these trials are followed by effectiveness studies examining the utility of an intervention in community-care settings with clinically heterogeneous samples. By the mid-1990s, accumulating evidence of the efficacy of CBT for depressed youths in controlled research settings led to increased interest in evaluating the effects of CBT among more severely depressed youths in clinical settings. Overall, ESs in this group of studies have been smaller than those observed in early RCTs.

The discrepancy between early estimates of the effects of CBT for depression among youths and the recent findings of Weisz et al. (2006) may be related to methodological differences between early RCTs and more recent trials. This possibility has not been evaluated among trials focusing exclusively on the effects of CBT for treating clinical depression. Accordingly, the purpose of this investigation was to elucidate the relationship between methodology and observed outcomes. Our secondary goals were to quantify the acute and follow-up effectiveness of CBT for youths with major depression and to systematically evaluate the methodological rigor of CBT trials for this population.

Method

Literature Review

The literature search spanned the period January 1980 to September 2006. Studies were located via searches of medical and psychological databases (*PsychINFO*, *Medline*, 1966–2006) and a review of references from identified outcome studies of CBT. Key words used in the searches included "depression," "dysthymia," and "major depression," with searches limited to youth populations. Following inclusion criteria used in several similar meta-analyses (Reinecke et al., 1998; Weisz et al., 1995a), we included only published outcome studies, relying on the peer-review process as one means of ensuring methodological rigor in included trials. Although this approach may maximize the quality of studies reviewed, it remains controversial given potential for publication bias against studies with negative findings (Sohn, 1996). Independent searches were conducted by the authors, each of whom generated a list of studies for potential inclusion, following the criteria detailed below. After pooling lists for a second independent review, there were no disagreements regarding the studies that met the inclusion criteria.

Inclusion Criteria

CBT was defined as interventions that promote emotional and behavioral change by teaching youths to change thoughts, thought processes, and behaviors in an overt, active, and problem-oriented manner. To be included, studies had to have included only participants with diagnoses of depressive disorders based on *DSM-III* (American Psychiatric Association, 1980) or later, Research Diagnostic Criteria (Feighner et al., 1972), or Bellevue Index of Depression criteria (Petti, 1978); used CBT as described above; randomly assigned participants to a CBT group

and a control or alternative psychotherapy group and reported pre- and posttest data for both groups; and been written in English.

Because CBT may be less effective among young children relative to youths functioning at the formal operational level (Durlak et al., 1991), we did not want to combine studies in our analyses that focused on children with those focusing on adolescents. However, we identified only one study focusing on the effects of CBT among children with major depression (Liddle and Spence, 1990). Accordingly, here we focus exclusively on adolescents and did not include the Liddle and Spence (1990) trial. In the present investigation, we defined adolescent samples as those in which the mean participant age was between 12 and 18 years.

Coding Procedures

Due to data limitations inherent in the small number of studies identified for inclusion in our analysis, we assigned dichotomous ratings for each of the coded variables. Cutoff values for all continuous variables were assigned by median split to ensure roughly comparable group sizes for each contrast. Interrater reliability was evaluated via κ values, and disagreements were reconciled following clarification of coding procedures and a joint review of study details. Coded variables included individual versus group treatment (κ = 1.0), whether samples included more or less than 37.3% males (κ = 1.0), whether participants were permitted to receive treatment for depression outside each study (κ = 1.0), and whether studies used intent-to-treat (ITT) analyses (κ = .78). ITT analyses include data obtained from all of the randomized participants regardless of participant adherence to or withdrawal from a given treatment protocol subsequent to randomization. We also coded the following variables:

- Treatment Duration: We coded whether the experimental group received more or less than 867 treatment minutes, assuming 1 hour per session, unless otherwise specified (κ = 1.0).

- Sample: We coded whether samples were composed primarily of students/community participants or outpatients/participants in the juvenile justice system (κ = 1.0).

- Type of Control Group: We coded whether control conditions were active or inactive treatments for depression. Therapies were considered active if they were posited to target symptoms of depression and/or provide empathy and therapeutic support. Given the large placebo effects in antidepressant trials (Walsh et al., 2002), placebo control groups were coded as active control treatments. Control conditions were considered inactive if they involved no treatment or were not posited to target symptoms of depression or provide therapeutic support (e.g., relaxation training, tutoring; κ = .80).

- Setting: We coded whether treatment was primarily conducted in nonclinical settings (e.g., school, university lab) or in clinical settings (κ = 1.0).

- Methodological Rigor: The CONSORT (Consolidated Standards of Reporting Trials) criteria for the reporting of RCTs (Begg et al., 1996) were used as an index of methodological rigor. We coded whether more or less than 17 of the 22 criteria enumerated by the CONSORT committee were fulfilled in each study (κ = 1.0).

- Therapist Vocation: We coded whether the majority of treating clinicians were primarily used as clinicians or as research assistants and/or graduate students (κ = .81).

- Severity: Estimates of participants' baseline symptom severity in each study were computed using z scores comparing participants' mean baseline depression score with the mean scores of clinical participants in normative samples available for the measure (s) administered in each trial. For studies in which we included more than one measure

of depression in our analyses, *z* scores were averaged to yield summary estimates of severity (L.V. Hedges, personal communication, April 2006). We then coded whether severity estimates indicated more or less severe symptomatology than was observed among clinical participants in normative samples ($\kappa = 1.0$).

General Analytic Procedure

To derive inferences about the distribution of effect parameters in a universe of similar, potential studies that use different participants and methods, we used a random-effects analytic approach to estimating mean treatment effects (Hedges and Vevea, 1998). Fixed-effects analysis only accounts for uncertainty due to the particular samples included in a specific set of studies and thus supports only conditional inference about that set of studies. However, because previous meta-analytic investigations of the effects of CBT for the treatment of depression among youths have used fixed-effects analytic approaches, we also estimated the mean effects of treatment with fixed-effects analyses to facilitate comparisons with earlier findings. ESs were calculated by dividing the posttherapy difference between CBT and control group means on measures of depression outcome by their pooled SD (Hedges and Olkin, 1985), applying Hedges' (1982) correction for small-sample bias when necessary. Although other indices of ES may have yielded comparable findings, the Hedges and Olkin (1985) approach was deemed to be most appropriate for our purposes given comparable variance between the treatment and control groups in the included studies at both pre- and posttreatment (Rosenthal, 1994). When studies included a waitlist control, ESs were computed by comparing CBT and waitlist scores. Otherwise, effects of CBT were compared to those of the alternative active-control treatment hypothesized to have the least therapeutic effect (Table 1). ESs were weighted by the inverse of their variances (Hedges and Olkin, 1985) and calculated such that positive scores indicate that CBT participants improved more than controls. When studies included multiple depression measures, we calculated a single, average ES and weight for each trial (Hedges and Olkin, 1985). To maximize the reliability of our findings, we included data obtained with the two measures per study with the greatest demonstrated reliability and validity and included only depression measures for which clinical norms are available.

To determine whether ESs in each study were significantly different from zero, we conducted *z* tests by dividing each ES by its SE (Lipsey and Wilson, 2001). A similar approach was used to determine whether the mean ES across studies differed from zero. We tested the assumption of homogeneity across observed ESs with the *Q* statistic, which is distributed as a χ^2 with $k - 1$ degrees of freedom, where *k* is the number of ESs (Hedges and Olkin, 1985). If ESs are similar across studies, then the χ^2 value of the *Q* statistic is nonsignificant.

Evaluation of Cumulative ESs Over Time

To evaluate changes in the observed efficacy and effectiveness of CBT over time, we conducted a series of cumulative random-effects meta-analyses, with information steps defined by the publication date of each included study. For these analyses, we iteratively calculated the mean cumulative effects of treatment across extant trials at the publication date of each subsequent study.

Evaluation of Moderators

We conducted mixed-effects analogs to the analysis of variance (ANOVA) for each coded variable to determine whether ESs were moderated by these variables (Lipsey and Wilson, 2001). The analog to the ANOVA tests whether some or all of the observed variability in ESs can be modeled based on mutually exclusive study groupings, where groups are defined by assigned codes reflecting study characteristics for the particular moderating variable of interest (e.g., group versus individual treatment). The mixed-effects model assumes that variability in

observed ESs is attributable to systematic between-study variables (moderators), subject-level sampling error, and an additional random-effects variance component.

Following Lipsey and Wilson (2001), the analog to the ANOVA partitions the total homogeneity statistic Q_T into the portion explained by the categorical moderator variable Q-between (Q_B) and the residual within group portion, Q-within (Q_W), pooled across groups. If the between category variance (Q_B) is significant, then mean ESs between groups differ by more than sampling error, and it is likely that the variable in question (or a variable that covaries with it) exerted a significant moderating influence. For the analog to the ANOVA, the Q statistic has $k - j$ degrees of freedom, where k is the number of ESs and j is the number of groups.

Evaluation of Publication Bias

To evaluate the threat of publication bias to the validity of our findings (Sohn, 1996), we estimated the number of unpublished trials with null findings that would be required to reduce the observed mean effects of CBT to nonsignificant levels, following Rosenthal's (1991) and Orwin's (1983) procedures. Under random-effects assumptions, 50 studies with null findings would be required to reduce the observed mean posttreatment effects to nonsignificance following Rosenthal (1991), and 48 studies with no effect would be needed to reduce the observed mean ES to 0.1 following Orwin (1983).

Power Analysis

To determine whether we had sufficient power to detect small, medium, and large mean ES (Cohen, 1988) differences between moderator study groupings, we conducted retrospective power analyses (Hedges and Pigott, 2004). Across moderator contrasts, we had a mean power of 0.11, 0.42, and 0.80 to detect small (0.2), medium (0.5), and large (0.8) mean ES differences, respectively, with an α of .05 (two-tailed).

Evaluation of Methodological Rigor

The methodological rigor of the included trials was evaluated by rating whether the revised criteria enumerated by the CONSORT committee (Begg et al., 1996) were met in each study. The CONSORT statement includes 22 guidelines for the reporting of RCTs. Consensus ratings followed joint review of study details.

Results

From 11 RCTs, 11 posttreatment and 9 follow-up control comparisons incorporated data from 809 and 638 participants, respectively. Populations studied included students, outpatients, children of outpatients, and youths in the juvenile justice system. Across studies, mean ages ranged from 12.7 to 16.2 years, and treatment involved 17.60 hours of therapy on average. Treatment with CBT was compared to a variety of control conditions including waitlists, conditions controlling for nonspecific aspects of treatment, active treatments targeting nondepressive symptoms, active treatment for depression, and a medication placebo. Six trials featured group treatment, and treatment was conducted in clinical settings in five studies. Characteristics of the trials are presented in Table 1.

Posttreatment Efficacy

The mean weighted fixed-effects posttreatment ES of CBT was 0.34 (SD 0.07; 95% confidence interval [CI] 0.20–0.48), significantly different from zero ($z = 4.68$; $p < .01$), and ES homogeneity was not supported ($Q(10) = 35.74$; $p < .01$). Under random-effects assumptions, the mean weighted posttreatment ES was .53 (SD 0.15; 95% CI 0.24–0.82; $Q(10) = 11.10$; p

= .35), again significantly different from zero ($z = 3.58$; $p < .01$). Posttest and follow-up data for each study are presented in Table 2.

Follow-up Effects

The effects of CBT at follow-up were evaluated in each study. However, control group comparisons were not possible in two studies due to rerandomization of participants in one study (Clarke et al., 1999) and unavailability of follow-up data reported by treatment group in another study (Brent et al., 1997). When data from multiple follow-up assessment points were reported, we used data gathered closest to 6-months posttreatment to minimize differences in follow-up duration between trials. The mean duration between termination of treatment and follow-up assessments was 5.58 months. For the nine follow-up comparisons, the mean weighted follow-up ES of CBT was 0.62 (fixed effects; SD 0.08; 95% CI 0.46–0.78; $Q(8) = 55.23$; $p < .01$), significantly different from zero ($z = 7.58$; $p < .01$). Under random-effects assumptions, the mean weighted follow-up ES of CBT was 0.59 (SD 0.23; 95% CI 0.14–1.05; $Q(8) = 6.64$; $p = .58$), also significantly different from zero ($z = 2.56$; $p < .01$).

For three of the nine follow-up comparisons, control participants were offered active treatment following the acute stage, and it was necessary to compare CBT group scores at follow-up to control group data gathered at end of the acute stage. The mean random-effects ES for these comparisons (ES = 1.21; SD 0.12; 95% CI 0.97–1.45; $Q(2) = 0.45$; $p = .80$) was significantly larger ($Q(1) = 28.02$; $p < .01$) than the mean follow-up ES for comparisons between CBT participants and control group participants both assessed concurrently at follow-up (ES = 0.18; SD 0.15; 95% CI −0.11 to 0.48; $Q(5) = 9.19$; $p = .10$).

Cumulative Efficacy of CBT Over Time

As detailed in Figure 1, the mean cumulative ES of CBT for depression among adolescents has decreased steadily from the large effects observed in the earliest trials, and CIs have become progressively narrower. With the accumulation of outcome data for a particular treatment, it could be argued that the cumulative mean ES of treatment should approach the population ES parameter. There are numerous methodological differences between the included studies, however, and it is plausible that the decrease in the effects of CBT is related to these differences.

Moderator Analyses

We found that ESs were larger in studies in which fewer than 17 CONSORT criteria were fulfilled (ES = 0.90, SE 0.21) relative to effects in studies that fulfilled greater than 17 CONSORT criteria (ES = 0.14, SE 0.09; $Q_B = 11.30$; $p < .01$). Most studies, however, were in compliance with the majority of the CONSORT criteria. The number of CONSORT criteria met in each of the studies ranged from 14 (Kahn and Kehle, 1990) to 21 (Treatment for Adolescents With Depression Study [TADS] Team, 2004; mean 17.4). Unfulfilled criteria reflected a lack of attention to discussion of the way in which sample sizes were determined, the methods used to generate and implement random group assignment, the blinding of investigators to group assignment, and the incidence of adverse effects.

In addition, we found that several other methodological variables moderated the observed posttreatment effects of CBT. Effects of treatment were significantly greater in studies that did not use ITT analyses (ES = 0.94, SE 0.28) relative to effects in studies that used ITT (ES = 0.26, SE 0.12; $Q_B = 4.94$; $p < .05$); studies comparing CBT to inactive treatments for depression (ES = 0.72, SE 0.17) relative to effects in studies comparing CBT to other active treatments or to a placebo control (ES = 0.11, SE 0.13; $Q_B = 7.90$; $p < .01$); and studies conducted in nonclinical settings (ES = 0.95, SE 0.27) relative to studies in clinical settings (ES = 0.25, SE 0.12; $Q_B = 5.94$; $p < .05$). Other coded variables did not moderate the effects of CBT. Although the moderator contrasts were sufficiently powered to detect large mean ES differences,

statistical power to detect smaller differences was low. Accordingly, the null results for moderator contrasts should be interpreted as inconclusive rather than as negative findings. Study groupings and results for analyses of moderation are presented in Table 3.

In interpreting these results, it should be noted that due to the small number of included studies, overlap in study groupings for several moderator contrasts was substantial. For example, studies conducted in clinical settings tended to feature greater methodological rigor. As such, it cannot be determined whether larger ESs in studies conducted in nonclinical settings were primarily associated with treatment setting, the methodological rigor of these studies, or to a combination of these characteristics.

Discussion

In the present investigation, we attempted to identify reasons for the reduced effects of CBT for adolescent depression in recent trials relative to large effects in early RCTs. In contrast to recent meta-analyses (Weisz et al., 2006), which included studies of youths with varying degrees of depressive symptomatology, we focused exclusively on RCTs involving adolescents with depressive diagnoses. Iterative cumulative meta-analyses involving 11 RCTs showed that the effects of CBT have declined steadily from large effects in early trials, and CIs have become progressively narrower. Smaller treatment effects were associated with several methodological characteristics that, in general, distinguish recent from early trials. These characteristics include the use of ITT analyses, comparison of effects of CBT to active treatment control conditions, administration of treatment in clinical settings, and the application of greater methodological rigor. Differences in the methodologies applied in recent trials of CBT for adolescent depression relative to early investigations, then, may help to explain the difference between large mean ESs observed in early meta-analyses (Lewinsohn and Clarke, 1999; Reinecke et al., 1998) and the smaller *ES*, as recently reported by Weisz et al. (2006).

The mean random-effects posttreatment ES of CBT was moderate (ES = 0.53) and statistically significant. Our evaluation of the methodological rigor of extant RCTs indicated that the limitations of these trials were minor. As such, this body of work would appear to support confident conclusions regarding the efficacy and effectiveness of CBT. Although methodological rigor emerged as a significant moderator of treatment effects, differences in the number of criteria met in each of the included studies were small (mean 17.36, SD 2.11). To promote increased methodological rigor in future trials, it will be important for investigators to discuss the way in which sample sizes were determined, the methods used to generate and implement random group assignments, the blinding of investigators to group assignment, and the incidence or absence of adverse events.

Limitations

Several considerations should be borne in mind in interpreting our findings regarding variables found to moderate treatment effects. First, it should be noted that because of the small number of studies included in our analyses, overlap between moderator groupings was not random and covariation among methodological variables was common. For example, five of six studies that fulfilled less than the median number of CONSORT criteria were conducted in nonclinical settings. Overlap between study groupings may confound the effects of the variables evaluated as moderators in the present investigation. Second, there are a number of additional factors that may moderate treatment outcome, including degree of comorbidity among participants, socioeconomic and racial composition of samples, and other demographic and clinical characteristics. Due to inconsistent reporting in the literature, we were unable to evaluate the moderating effect of several of these factors. Unexplored variables may account in part for the findings and should be investigated in future research. To facilitate further investigation of such variables, increased attention to more detailed reporting of participant characteristics will

be important subsequent work. Third, because statistical power in our moderator analyses was sufficient only to detect large ES differences, null results for these analyses should be interpreted as inconclusive rather than as negative findings. Fourth, due to data limitations associated with the small number of included studies, we were unable to evaluate potential covariation. As such, it remains possible that covariation may have obscured significant findings. Fifth, because we included only studies of adolescents with clinical depression, the findings may not generalize to the effects of CBT for the treatment of adolescents with subsyndromal depression, bipolar affective disorders, or to the prevention of depressive symptomatology.

With regard to the follow-up findings, it should be noted that use of outside psychiatric services was common across studies subsequent to posttreatment assessment. As a result, follow-up ESs cannot be confidently interpreted as indicators of the maintenance of treatment gains. Although ethical considerations preclude withholding services during follow-up, receipt of additional treatment makes it difficult to evaluate the maintenance of therapeutic gains. Consistent reporting of the type and amount of services received by participants following acute treatment, however, may enable future investigators to statistically control for open treatment.

Although our calculations indicated that it is unlikely that publication bias played a role in our findings, this possibility cannot be ruled out. It is noteworthy, however, that Weisz et al. (2006) did not find evidence of a difference in treatment effects between published and non–peer-reviewed studies included in their analyses.

Research Implications

Although findings to date have been promising, existing treatments for depression are not entirely effective. Many adolescents do not respond to treatment, and among those who do, relapse is common. In addition, many youths who benefit from treatment continue to experience persistent symptoms of depression or functional impairment after treatment has been completed. Accordingly, as the efficacy of CBT for the treatment of depression among adolescents becomes more firmly established, additional research is needed to identify those components of treatment that are most strongly associated with clinical improvement and to identify factors associated with risk of relapse, recurrence, or incomplete treatment response. To further support and clarify cognitive-behavioral models of depression and to refine and improve interventions, research is also needed to evaluate variables purported to mediate change such as attributional styles, environmental reinforcement, problem solving, dysfunctional attitudes, assumptions, social support, ruminative style, cognitive distortions, and amelioration of stress.

Clinical Implications

Extant RCTs support the conclusion that group or individual CBT may be effective for treating depression among adolescents. Findings from the Treatment for Adolescents With Depression Study, however, indicate that at least in the acute stage of treatment, the combination of CBT and medication may provide more rapid and significant clinical improvement (Treatment for Adolescents With Depression Study [TADS] Team, 2004). Because only one study has been completed examining the effectiveness of combined CBT and medication for treating depressed adolescents, further research is warranted.

Although consistent with findings highlighting disparities in treatment outcomes between research and clinical settings (Weisz et al., 1995b), the relatively small effects of CBT delivered in clinical settings remain cause for concern. Weisz et al. (1995b) outlined several potential reasons for the discrepancy between outcomes obtained in research and clinical settings. Some

of these include the referral status of treated individuals (i.e., referred versus recruited), homogeneity of participants in RCTs relative to patients treated in clinic settings, caseload size differences between clinicians in research settings relative to those in clinical settings and differences in the breadth of the focus of treatment between settings. Weisz et al. (1995b) also proposed an agenda for bridging the gap between treatment outcomes in research and clinical settings, which includes three primary elements: enriching the database on outcomes of traditional clinic therapy, identifying factors that may account for the large effects of research therapy relative to clinic therapy, and exporting treatments from laboratories to clinical practice sites. Continued attention to each of these factors in future research will be crucial to improving outcomes in clinical settings.

Although we did not limit studies included in our analysis to those that used manual-based CBT protocols, each of the studies did use a form of manual-based CBT. Treatment fidelity and therapist allegiance to specific models of treatment appear to be associated with more positive outcomes. In clinical settings, however, manuals may not be used and clinicians may adopt more eclectic approaches. To the extent that clinicians diverge from the use of empirically supported forms of CBT, treatment outcomes may be less optimal.

Given previous work demonstrating that depression severity is associated with response to CBT among adolescents (Jayson et al., 1998), we are somewhat surprised not to have found that severity moderated outcomes. It is possible, however, that the null findings may be related to the use of different outcome measures across studies. Specifically, several of these measures were validated among clinically depressed adults and may not be sensitive measures of depression among adolescents. In addition, differences between clinical standardization samples (e.g., inpatient versus outpatient samples) recruited for the validation of these measures may have compromised the validity of our estimates of baseline symptom severity and thereby obscured differences in treatment effects as a function of depression severity. It is also possible that severity did not moderate treatment effects in the present investigation because we restricted inclusion to studies of adolescents with major depressive disorder, thereby limiting the range of symptom severity among participants.

Conclusions

Differences in methodological characteristics between early and recent RCTs may contribute to discrepancies among estimates of the effects of CBT for adolescent depression. These differences appear to reflect both a shift from an initial emphasis on demonstrating the efficacy of treatment in controlled research settings to an emphasis on demonstrating the effectiveness of treatment and the application of increased statistical and methodological rigor over time. Taken together, the results indicate that CBT may be effective for the acute treatment of depression among adolescents, although treatment effects may be more modest in clinical settings than findings from early trials would suggest.

References

Asterisks indicate studies included in the meta-analysis.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 3rd edition (DSM-III). Washington, DC: American Psychiatric Association; 1980.

Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–639. [PubMed: 8773637]

*. Brent DA, Holder D, Kolko D, et al. A clinical psychotherapy trial for adolescent depression comparing cognitive, family and supportive therapy. Arch Gen Psychiatry 1997;54:877–885. [PubMed: 9294380]

*. Clarke GN, Hornbrook M, Lynch F, et al. Group cognitive-behavioral treatment for depressed adolescent offspring of depressed parents in a health maintenance organization. J Am Acad Child Adolesc Psychiatry 2002;41:305–313. [PubMed: 11886025]

*. Clarke GN, Rohde P, Lewinsohn PM, Hops H, Seeley JR. Cognitive-behavioral treatment of adolescent depression: efficacy of acute treatment and booster sessions. J Am Acad Child Adolesc Psychiatry 1999;38:272–279. [PubMed: 10087688]

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. 2nd. Hillsdale, NJ: Erlbaum; 1988.

Durlak J, Fuhrman T, Lampman C. Effectiveness of cognitive-behavior therapy for maladapting children: a meta-analysis. Psychol Bull 1991;110:204–214. [PubMed: 1835106]

Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 1972;26:57–63. [PubMed: 5009428]

Hedges LV. Estimation of effect size from a series of independent experiments. Psychol Bull 1982;92:490–499.

Hedges, LV.; Olkin, I. Statistical Methods for Meta-Analysis. San Diego: Academic; 1985.

Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. Psychol Methods 2004;4:426–445. [PubMed: 15598097]

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

Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative meta-analyses. Proc Natl Acad Sci U S A 2001;98:831–836. [PubMed: 11158556]

Jayson D, Wood A, Kroll L, Fraser J, Harrington R. Which depressed patients respond to cognitive-behavioral treatment? J Am Acad Child Adolesc Psychiatry 1998;37:35–39. [PubMed: 9444897]

*. Kahn JS, Kehle TJ. Comparison of cognitive-behavioral, relaxation, and self-modeling interventions for depression among middle-school students. Sch Psychol Rev 1990;19:196–212.

Lewinsohn PM, Clarke GN. Psychosocial treatments for adolescent depression. Clin Psychol Rev 1999;19:329–342. [PubMed: 10097874]

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

Liddle B, Spence SH. Cognitive-behaviour therapy with depressed primary school children: a cautionary note. Behav Psychother 1990;18:85–102.

Lipsey, MW.; Wilson, DB. Practical Meta-Analysis. Thousand Oaks, CA: Sage; 2001.

Orwin RG. A fail-safe N for effect size in meta-analysis. J Educ Stat 1983;8:157–159.

Petti TA. Depression in hospitalized child psychiatry patients: approaches to measuring depression. J Am Acad Child Adolesc Psychiatry 1978;17:49–59.

Reinecke MA, Ryan NE, Dubois DL. Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: a review and meta-analysis. J Am Acad Child Adolesc Psychiatry 1998;37:26–34. [PubMed: 9444896]

*. Reynolds WM, Coats KI. A comparison of cognitive-behavioral therapy and relaxation training for the treatment of depression in adolescents. J Consult Clin Psychol 1986;54:653–660. [PubMed: 3534032]

*. Rohde P, Clarke GN, Mace DE, Jorgensen JS, Seeley JR. An efficacy/effectiveness study of cognitive-behavioral treatment for adolescents with comorbid major depression and conduct disorder. J Am Acad Child Adolesc Psychiatry 2004;43:660–668. [PubMed: 15167082]

Rosenthal, R. Meta-Analytic Procedures for Social Research. Newbury Park, CA: Sage; 1991.

Rosenthal, R. Parametric measures of effect size. In: Cooper, H.; Hedges, LV., editors. Handbook of Research Synthesis. New York: Russell Sage Foundation; 1994. p. 231-244.

*. Rossello J, Bernal G. The efficacy of cognitive-behavioral and interpersonal treatments for depression in Puerto Rican adolescents. J Consult Clin Psychol 1999;67:734–745. [PubMed: 10535240]

Sohn D. Publication bias and the evaluation of psychotherapy efficacy in reviews of the research literature. Clin Psychol Rev 1996;16:147–156.

*. Treatment for Adolescents With Depression Study (TADS) Team. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression. JAMA 2004;292:807–820. [PubMed: 15315995]

*. Treatment for Adolescents With Depression Study (TADS) Team. The Treatment for Adolescents with Depression Study (TADS): long-term effectiveness and safety outcomes. Arch Gen Psychiatry. in press

*. Vostanis P, Feehan C, Grattan E, Bickerton W. Treatment for children and adolescents with depression: lessons learned from a controlled trial. Clin Child Psychol Psychiatry 1996a;1:199–212.

*. Vostanis P, Feehan C, Grattan E, Bickerton W. A randomized controlled out-patient trial of cognitive-behavioural treatment for children and adolescents with depression: 9-month follow-up. J Affect Disord 1996b;40:105–116. [PubMed: 8882920]

Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002;287:1840–1847. [PubMed: 11939870]

Weisz JR, Donenberg GR, Han SS, Weiss B. Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. J Consult Clin Psychol 1995b;63:688–701. [PubMed: 7593861]

Weisz JR, McCarty CA, Valeri SM. Effects of psychotherapy for depression in children and adolescents: a meta-analysis. Psychol Bull 2006;132:132–149. [PubMed: 16435960]

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. Psychol Bull 1995a;117:450–468. [PubMed: 7777649]

*. Wood A, Harrington R, Moore A. Controlled trial of a brief cognitive-behavioural intervention in adolescent patients with depressive disorders. J Child Psychol Psychiatry 1996;37:737–746. [PubMed: 8894955]
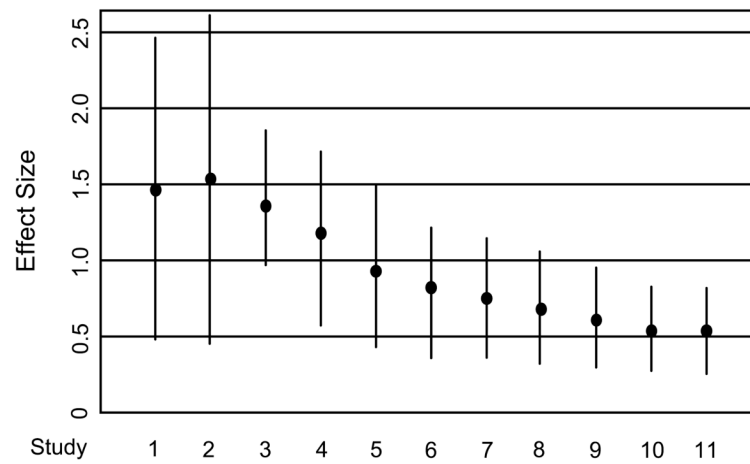
Fig. 1.
Cumulative effect sizes of cognitive-behavioral therapy over time (random effects, acute treatment stage). Information steps defined by subsequent publications. Study: 1 Reynolds and Coats, 1986; 2 Kahn and Kehle, 1990; 3 Lewinsohn et al., 1990; 4 Vostanis et al., 1996a; 5 Wood et al., 1996; 6 Brent et al., 1997; 7 Clarke et al., 1999; 8 Rossello and Bernal, 1999; 9 Clarke et al., 2002; 10 Rohde et al., 2004; 11 Treatment for Adolescents With Depression Study (TADS) Team, 2004.

TABLE 1

Study Characteristics

Study	Sample; Baseline *N* by Gender; Mean Age and/or Age Range	Treatment Arm Descriptions (*N* Posttest)	Therapy Length/Format[a]
Reynolds and Coats, 1986	Students; 11 M, 19 F; 15.7	CBT* (9), relaxation training (11), WL* (10)	Ten 50-min group sessions
Kahn and Kehle, 1990	Students; 32 M, 36 F; 10–14	CBT* (17), relaxation training (17), self-modeling (17), WL* (16)	Twelve 60-min group sessions
Lewinsohn et al., 1990	Students; 22 M, 37 F; 16.2, 14–18	CBT* (19), CBT + parent groups* (21), WL* (19)	Fourteen 2-h group sessions (+ seven 2-h parent groups for CBT + parent group)
Vostanis et al., 1996a,b	Outpatients; 25 M, 32 F; 12.7, 8–17	CBT* (29), nonfocused intervention* (28)	Mean no. of sessions: 6
Wood et al., 1996	Outpatients; 20 M, 33 F; 14.2, 9–17	CBT* (26), relaxation training* (27)	Mean no. of sessions: 5.4
Brent et al., 1997	Outpatients; 26 M, 81 F; 15.7, 13–18	CBT* (37), systemic behavior family therapy (35), nondirective supportive therapy* (35)	Mean no. of sessions: CBT, 12.1; nondirective supportive therapy, 11.2
Clarke et al., 1999	Community youths; 28 M, 68 F; 16.2, 14–18	CBT* (37), CBT + parent groups* (32), WL* (27)	Sixteen 2-h group sessions (+ eight 2-h parent groups for CBT + parent group)
Rossello and Bernal, 1999	Students; 33 M, 38 F; 14.7, 13–17	CBT* (21), IPT (19), WL* (18)	Twelve 1-h sessions
Clarke et al., 2002	Children of depressed parents; 27 M, 61 F; 15.3, 13–18	CBT+HMO care* (41), HMO care* (47)	Sixteen 2-h group sessions
Rohde et al., 2004	Youths in the juvenile justice system; 48 M, 45 F; 15.1, 13–17	CBT* (44), life skills/tutoring* (47)	Sixteen 2-h group sessions
Treatment for Adolescents With Depression Study (TADS) Team, 2004, in press	Outpatients; 200 M, 239 F; 14.6, 12–17	CBT* (111), fluoxetine (109), fluoxetine with CBT (107), placebo* (112)	Fifteen 50- to 60-min sessions

Note: Groups included in our analyses are indicated by asterisks. CBT = cognitive-behavioral therapy; HMO = health maintenance organization; IPT = interpersonal psychotherapy; WL = waitlist.

[a] Individual treatment unless noted.

TABLE 2

Posttreatment and Follow-up Effects

Study	Assessment Point	ES	1/v	Fixed Effects 95% CI	1/v	Random Effects 95% CI
Reynolds and Coats, 1986	Acute	1.39	4.04	0.41–2.36 **	2.43	0.13–2.64 **
	Follow-up (5 wk)	1.04	3.17	−0.06–1.86 *	1.43	−0.59–2.68
Kahn and Kehle, 1990	Acute	1.57	7.46	0.85–2.29 **	3.69	0.55–2.59 **
	Follow-up (1 mo)	1.13	7.73	0.43–1.83 **	1.95	−0.27–2.53
Lewinsohn et al., 1990	Acute	1.28	8.81	0.62–1.94 **	4.00	0.30–2.26 **
	Follow-up (6 mo)[a]	1.38	7.48	0.66–2.09 **	1.93	−0.03–2.79 *
Vostanis et al., 1996a,b	Acute	0.28	13.98	−0.24–0.81	4.80	−0.61–1.18
	Follow-up (9 mo)	−0.13	14.42	−0.65–0.39	2.20	−1.45–1.19
Wood et al., 1996	Acute	0.76	12.58	0.21–1.32 **	4.63	−0.15–1.68
	Follow-up (6 mo)	0.14	11.88	−0.43–0.71	2.14	−0.43–0.71
Brent et al., 1997	Acute	0.41	16.57	−0.07–0.89	5.08	−0.46–1.28
Clarke et al., 1999	Acute	0.38	14.77	−0.13–0.89	4.89	−0.51–1.26
Rossello and Bernal, 1999	Acute	0.35	9.43	−0.29–0.99	4.12	−0.61–1.32
	Follow-up (3 mo)[a]	1.01	6.24	0.38–1.65 **	1.84	−0.43–2.46
Clarke et al., 2002	Acute	0.15	21.72	−0.27–0.57	5.47	−0.69–0.99
	Follow-up (12 mo)	−0.05	21.96	−0.47–0.37	2.33	−1.29–1.28
Rohde et al., 2004	Acute	0.28	22.39	−0.13–0.70	5.52	−0.55–1.12
	Follow-up (6 mo)	0.00	21.46	−0.43–0.42	2.33	−1.29–1.28
Treatment for Adolescents With Depression Study (TADS) Team, 2004, 2007	Acute	−0.06	55.86	−0.32–0.20	6.57	−0.83–0.71
	Follow-up (6 mo)[a]	1.21	53.56	0.94–1.48	2.49	−0.03–2.45 *

Note: ES = effect size; 1/v = inverse variance weight.

[a] Follow-up CBT means compared to posttreatment control means.

*
p < .05.

**
p < .01.

TABLE 3

Study Groupings and Results for Analyses of Variance

		1	2	3	4	5	6	7	8	9	10	11	Mean ES (SE)	Q_B[a]	P	$Q_w(df)$
Treatment modality	Individual				x	x	x		x			x	0.30 (0.15)	2.70	.101	3.19 (4)
	Group	x	x	x				x		x	x		0.76 (0.24)			6.06 (5)
Analytic strategy	ITT				x		x	x		x	x	x	0.26 (0.12)	4.94*	.026	4.61 (5)
	Not ITT	x	x	x		x			x				0.94 (0.28)			3.97 (4)
Treatment duration	<867 min	x	x	x	x	x		x	x				0.60 (0.22)	0.16	.685	6.35 (6)
	>867 min						x			x	x	x	0.48 (0.20)			4.34 (3)
Sample	Student/community	x	x	x		x		x		x	x		0.65 (0.23)	1.00	.318	6.97 (6)
	Outpatient/juvenile justice				x		x		x			x	0.39 (0.10)			2.88 (3)
Type of control group	Inactive	x	x	x	x	x		x		x	x		0.72 (0.17)	7.90**	.005	8.20 (7)
	Active						x		x			x	0.11 (0.13)			2.02 (2)
Setting	Nonclinical	x	x	x		x		x					0.95 (0.27)	5.94*	.015	4.09 (4)
	Clinical				x		x		x	x	x	x	0.25 (0.12)			4.63 (5)
Methodological rigor	<17 CONSORT criteria met	x	x	x	x	x		x					0.90 (0.12)	11.30**	.001	5.39 (5)
	>17 CONSORT criteria met						x		x	x	x	x	0.14 (0.09)			4.00 (4)
Gender representation	<37.3% male	x				x	x	x	x			x	0.64 (0.18)	0.60	.437	6.06 (5)
	>37.3% male		x	x	x					x	x		0.42 (0.23)			5.31 (4)
Outside services	Permitted						x			x	x	x	0.27 (0.11)	3.09	.079	0.66 (3)
	Not permitted	x	x	x	x	x		x	x				0.76 (0.25)			5.37 (6)
Therapist vocation	Clinicians				x		x	x	x	x	x	x	0.41 (0.17)	2.22	.136	7.58 (6)
	Research assistants/graduate students	x	x	x		x							0.86 (0.25)			4.03 (3)
Severity	Less depressed	x	x	x		x							0.56 (0.16)	0.20	.887	7.66 (3)
	More depressed						x		x	x	x		0.51 (0.31)			3.64 (3)

Note: Study: 1 Reynolds and Coats, 1986; 2 Kahn and Kehle, 1990; 3 Lewinsohn et al., 1990; 4 Vostanis et al., 1996a; 5 Wood et al., 1996; 6 Brent et al., 1997; 7 Clarke et al., 1999; 8 Rossello and Bernal, 1999; 9 Clarke et al., 2002; 10 Rohde et al., 2004; 11 Treatment for Adolescents With Depression Study (TADS) Team, 2004. ES = effect size; ITT = intent-to-treat analysis; CONSORT = Consolidated Standards of Reporting Trials.

[a] *df* for $Q_B = 1$.

* $p < .05$.

** $p < .01$.