Evidence-Based Youth Psychotherapies Versus Usual Clinical Care

A Meta-Analysis of Direct Comparisons

John R. Weisz — *Judge Baker Children's Center, Harvard University*
Amanda Jensen-Doss — *Texas A&M University*
Kristin M. Hawley — *University of Missouri—Columbia*

In the debate over evidence-based treatments (EBTs) for youth, one question is central: Do EBTs produce better outcomes than the usual interventions employed in clinical care? The authors addressed this question through a meta-analysis of 32 randomized trials that directly compared EBTs with usual care. EBTs outperformed usual care. Effects fell within the small to medium range at posttreatment, increasing somewhat at follow-up. EBT superiority was not reduced by high levels of youth severity or by inclusion of minority youths. The findings underscore a need for improved study designs and detailed treatment descriptions. In the future, the EBT versus usual care genre can inform the search for the most effective interventions and guide treatment selection in clinical care.

Keywords: psychotherapy, children and adolescents, evidence-based treatments, usual clinical care, meta-analysis

In contemporary psychology, few topics have generated more intense discussion and debate than the pros and cons of evidence-based treatments (EBTs)—that is, treatments supported by empirical evidence. Proponents of EBTs argue that interventions showing beneficial effects in outcome research should be taught and used in preference to interventions that have not been tested and shown to be effective (see, e.g., Calhoun, Moras, Pilkonis, & Rehm, 1998; Chambless et al., 1998; Chambless & Hollon, 1998; Chorpita, 2003; Lonigan & Elbert, 1998). Similar recommendations have come from the National Institute of Mental Health (National Advisory Mental Health Council Workgroup on Child and Adolescent Mental Health Intervention and Deployment, 2001), the Office of the Surgeon General (1999, 2004), and the President's New Freedom Commission on Mental Health (2003). In principle, it seems reasonable to favor interventions that have been tested empirically and shown to work. However, concerns have been raised about these interventions and the evidence base on which they rest.

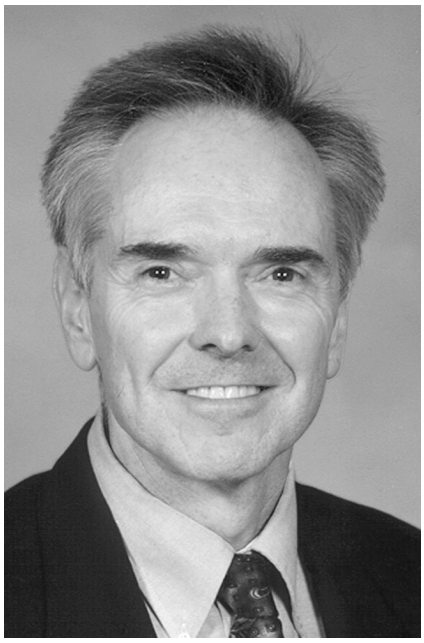
Opponents of the growing emphasis on EBTs argue that these structured, manual-guided treatments have significant limitations that undermine their effectiveness in usual clinical practice contexts. Various writers have suggested that EBTs (a) have been developed and tested with relatively simple, often subclinical, cases and thus may not work well with the complex and severe cases seen in usual clinical care; (b) have been designed for single problems or diagnoses and thus may not work well with comorbidity, which is common in usual clinical care; (c) are so protocol driven that they make it hard to individualize treatment to meet distinctive client needs; (d) are so formulaic that they constrain therapist creativity in addressing unusual or unexpected events in clients' lives; and (e) are so lacking in spontaneity and flexibility that they interfere with rapport building and the development of a good therapeutic relationship. Several of the concerns reflect the view that EBTs may not be well-suited to the challenge of treating clinically referred individuals in the context of usual clinical care (see examples of such concerns discussed in Addis & Krasnow, 2000; Addis & Waltz, 2002; Garfield, 1996; Havik & VandenBos, 1996; Strupp & Anderson, 1997; Westen, Novotny, & Thompson-Brenner, 2004a, 2004b). Concerns have also been raised by experts on culture and ethnicity that EBTs may not have been adequately adapted for and may not work well with members of

John R. Weisz, Judge Baker Children's Center, Harvard University; Amanda Jensen-Doss, Departments of Educational Psychology and Psychology, Texas A&M University; Kristin M. Hawley, Department of Psychological Sciences, University of Missouri—Columbia.

This project was supported in part through the Research Network on Youth Mental Health (Bruce F. Chorpita, Robert Gibbons, Charles Glisson, Evelyn Polk Green, Kimberly Hoagwood, Peter S. Jensen, Kelly Kelleher, John Landsverk, Stephen Mayberg, Jeanne Miranda, Lawrence A. Palinkas, Sonja Schoenwald, and John Weisz), funded by the John D. and Catherine T. MacArthur Foundation. We were supported by Grants R01 MH 547347 and R01 MH 068806 to John R. Weisz, Grant NRSA F31 MH65811 to Amanda Jensen-Doss, and Grant NRSA F31 MH12853 to Kristin M. Hawley from the National Institute of Mental Health, and by San Diego State University Foundation's Oscar Kaplan Postdoctoral Fellowship to Kristin M. Hawley. We are grateful to Brian Baucom, Amie Bettencourt, Vickie Chang, Brian Chu, Erin Dabbs, Jen Durham, Julie Elins, Samantha Fordwood, Sarah Francis, Dan Fulford, Alanna Gelbwasser, Adam Hafdahl, Eunie Jung, Bryce McLeod, Melissa Magaro, Antonio Polo, Kathya Vega, and Robin Weersing for the important roles they played in this project.

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

Vol. 61, No. 7, 671–689 DOI: 10.1037/0003-066X.61.7.671



John R. Weisz

ethnic minority groups (see, e.g., Bernal & Scharrón-Del-Río, 2001; Gray-Little & Kaplan, 2000; Hall, 2001; Sue, 2003).

The debate over the strengths and limitations of EBTs has genuine value for the field—indeed, for all who have a stake in good mental health care. It is clearly important to understand both the ways in which EBTs may offer advantages in clinical care and the ways in which the characteristics of EBTs and their evidence base may fail to reflect realities of clinical care or fall short in other ways (see, e.g., Weisz, Doss, & Hawley, 2005). However, it is also important to note that much of the debate over EBTs is focused on a central question that has not, to our knowledge, been directly addressed in any systematic review: When EBTs and usual care (UC) are directly compared with one another, does one form of treatment produce superior outcomes? In the present article, we address this question, which lies at the heart of the debate over the clinical utility and effectiveness of EBTs. We reasoned that if direct, randomized comparisons show that EBTs produce effects inferior to those of usual clinical care, such a finding would support concerns raised by critics and would undermine the argument that EBTs should be taught and used instead of UC. Indeed, proponents of evidence could hardly argue that EBTs should be favored over UC if the evidence from direct comparisons showed that UC produced superior outcomes. However, if EBTs, despite the limitations that have been attributed to them in previous literature, are found to produce clinical outcomes superior to those of UC, such a finding would support those who propose increased training in and clinical use of EBTs.

Comparing outcomes of EBTs and UC is relevant to therapy for clients across a broad age range, encompassing children and adolescents (herein referred to as *youths*) as well as adults. However, combining studies of youths and adults within the same review may not be ideal, given the many differences in clinical care between youths and adults—differences that might influence the effectiveness of EBTs and the outcomes of EBT-versus-UC comparisons. To illustrate, in contrast to adults, boys and girls rarely refer themselves for treatment. Instead, most youth referrals are made by parents and other adults (see, e.g., Yeh & Weisz, 2001); such referrals, in contrast to adult self-referrals, tend to be prompted by unusually high levels of youth dysfunction, often with high levels of comorbidity (see, e.g., Angold, Costello, & Erkanli, 1999) and high levels of parent and family stress (see, e.g., Hammen, Rudolph, Weisz, Burge, & Rao, 1999), which lead to unpredictable turns of events across successive therapy sessions (see, e.g., Weisz, Southam-Gerow, Gordis, & Connor-Smith, 2003). No-show and dropout risks are significant (see, e.g., Kazdin, 1996), compounded by the fact that both child and parent must be engaged to sustain treatment attendance, and evidence indicates that outcomes in youth psychotherapy, unlike those in adult therapy, are influenced by both youth and parent alliance with the therapist (see, e.g., Hawley & Weisz, 2005). Thus, several of the reality factors said to put EBTs at significant risk of failure in usual clinical care—high severity, comorbidity, stressful life circumstances, unexpected changes in the client or situation across successive therapy sessions, and the need for a strong working alliance to maintain attendance and foster good outcomes—are especially pronounced and have a distinctive character in youth therapy. If these factors do, in fact, undermine the effectiveness of EBTs, as the literature cited above suggests, then EBTs may be particularly ineffective in the context of youth intervention. Thus, youth intervention appears to be an appropriate context for an initial direct comparison of EBTs versus UC.

Another reason to focus initially on youth therapy is that the effects of usual clinical care appear to have been reviewed more systematically for youth therapy than for adult therapy. A narrative review of research on UC for youth (Weisz, Weiss, & Donenberg, 1992) and meta-analytic reviews reporting effect size (ES) statistics for UC versus control groups (e.g., Weisz, 2004; Weisz & Jensen, 2001) have not produced a very positive picture. Although the number of comparisons available in the literature is modest ($N = 14$ in the most recent analysis), the average effect in those comparisons has hovered near zero, suggesting no benefit of UC, on average. By contrast, the mean ES in hundreds of studies of structured, mostly manual-guided treatments for youth tested in randomized trials (reviewed in Weisz et al., 2005) have hovered within the range of 0.50 to 0.80, commonly used cutpoints for medium and large effects, respectively (Cohen, 1988). In related work, Weersing and Weisz (2002) used a benchmarking strategy to compare outcomes for depressed youths treated in community clinic UC with outcomes for depressed youths in clinical trials of cognitive–behavioral therapy (CBT); in general, the UC outcomes more closely approximated those of control groups than treatment groups in the CBT trials.

Amanda Jensen-Doss

At first blush, the contrast between the near-null effects found for UC studies and the medium to large effects found in studies of structured, protocol-guided treatments for youth may appear to suggest that the more structured, manual-guided treatments are more effective than usual clinical care. However, there are good reasons to be cautious about such a conclusion. First, as noted above and in previous reports, the conditions in which UC is usually carried out and tested differ markedly from the conditions of treatment in conventional randomized trials, along multiple dimensions, and in ways that could make treatment effects more difficult to demonstrate in UC than in randomized trials (see, e.g., Costello, Angold, & Burns, 1996; Garland et al., 2001; Hammen et al., 1999; Jensen & Weisz, 2002; Southam-Gerow, Weisz, & Kendall, 2003; Weisz, 2004; Weisz et al., 1992, 2005; Yeh et al., 2002). For example, the participants in UC are typically clinically referred cases, often quite complex and with substantial comorbidity, whereas the participants in the experimental trials have often been recruited to fit a specific profile and are likely to be less complex and show lower levels of comorbidity, family stress, and financial disadvantage. Clients in randomized trials, unlike clients in UC, often receive financial and other incentives for participation and attendance and often for homework completion. Therapists in randomized trials typically learn only one specific treatment protocol and treat only one specific type of client, whereas therapists in UC typically treat a highly diverse clientele (e.g., 5–8 youths with different diagnoses in a single workday) and use a wide range of interventions. In general, treatment in randomized trials has tended to occur under more ideal experimenter-controlled *efficacy trial* conditions, whereas treatment in UC studies has tended to take place under the more challenging conditions of everyday care, within which it might be more difficult to demonstrate beneficial effects.

Given the substantial differences between randomized efficacy trial conditions and UC conditions, comparing the findings of randomized efficacy trials with the findings of completely separate studies of UC cannot provide the most compelling evidence on the relative effects of the two sets of treatments with youth. This is certainly true for the question of whether EBTs and UC differ in their effects. The most appropriate way to address that question is through studies in which youths in a common pool of participants are randomly assigned to receive either UC or an EBT intervention—that is, an intervention that has shown evidence of beneficial effects in prior testing. In the present article, we report the results of a meta-analysis focused exclusively on studies that compared outcomes achieved by groups randomly assigned to UC or EBT conditions. We initially defined both UC and EBT as broadly as was feasible, to ensure the most complete and representative sample of studies possible. After an initial comparison using the most inclusive sample of studies, we carried out a series of analyses using increasingly restrictive inclusion criteria, and we explored various candidate explanations for the findings, focusing on characteristics of the youths treated, the treatments provided, the therapists who delivered the treatments, and characteristics of the intervention context within which UC and the EBTs were carried out.

Method

Literature Review

The search process for this project was carried out as part of a larger review of the youth treatment outcome literature. To identify relevant studies, we first searched standard computerized databases for outcome studies that were reported between 1965 and December 2004. Our start date may warrant explanation: Although EBTs were not formally identified as such until many years after 1965, outcome studies were conducted in the mid-1960s for treatments that would later be designated EBTs, and our search did, in fact, identify trials in the early 1970s that were specifically relevant to the present meta-analysis (see Table 1). We used PsycINFO, applying 21 psychotherapy-related key terms (e.g., *psychother-*, *counseling*, *treatment*) used in our previous youth psychotherapy meta-analyses (Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995), and we used MEDLINE via PubMed, the principal bibliographic database of the National Library of Medicine. PubMed uses a controlled vocabulary indexing system (Medical Subject Headings, also called MeSH) that provides a consistent way to retrieve citations from publishers who may use different keywords for the same concepts; we used *mental disorders*, with the following search limits: *clinical trial*, *child (3–18 years)*, *published in English*, and *human subjects*. In addition to these database searches, we surveyed published reviews and meta-analyses of the youth psychotherapy literature to identify studies not found in PsycINFO or

Kristin M. Hawley

MEDLINE (e.g., Compton, Burns, Egger, & Robertson, 2002; Durlak, Fuhrman, & Lampman, 1991; Farmer, Compton, Burns, & Robertson, 2002). We also followed reference trails of reviewed studies, and we screened studies suggested by investigators in the field. Finally, to identify studies of acceptable methodological quality that had not undergone publication review (see McLeod & Weisz, 2004), we searched Dissertation Abstracts and PsycINFO for dissertations using the same search terms we used for the published literature search.

Criteria for Study Inclusion and Resulting Pool of Studies

We examined published youth psychotherapy outcome studies and unpublished dissertations identified by the procedures noted above to identify all studies that involved a comparison of an EBT to a UC condition. EBTs were defined (broadly, for our initial set of analyses) as treatments that had been included in at least one list of treatments showing beneficial effects (Lonigan & Elbert, 1998; Nathan & Gorman, 1998, 2002; Roth & Fonagy, 1996, 2004; Silverman & Hinshaw, in press; Weisz, Hawley, & Doss, 2004). UC was defined as psychotherapy, counseling, or case management provided as part of the regular services of providers, agencies, organizations, programs, or facilities for youth. Studies in which the authors permitted control participants to seek outside services or provided them with lists of community providers were only included if the authors either facilitated participants' use of these services (e.g., arranged intake appointments) or established that an equivalent proportion of UC and EBT participants (i.e., not differing by more than 10%) received services.

In addition, the studies were required to meet the following criteria: (a) selection of participants with psychological problems or maladaptive behavior, (b) random assignment of participants to treatment conditions, (c) mean participant age of 3–18 years, and (d) posttreatment assessment of the psychological problem(s) or maladaptive behavior for which participants were selected and treated. The random assignment requirement ruled out numerous studies but seemed essential to a fair test. When separate articles were published from the same data set (e.g., separate reports of posttreatment and follow-up findings), these were combined for analysis as a single study. Our final sample consisted of 32 studies (23 published articles and 9 dissertations), with 39 EBT versus UC comparisons (see Table 1).

Coding Procedures and Intercoder Reliability

After acceptable studies had been identified, we coded several study characteristics. To establish interrater reliability for these codes, we had 30 randomly selected studies from the larger outcome literature review independently coded by three project coders: one postdoctoral fellow who served as the master coder and two clinical psychology graduate student coders. Reliability was then computed by comparing the ratings of each of the coders with those of the master coder, with mean kappa statistics computed across coders for categorical codes and mean two-way random-effects intraclass correlation coefficients (ICCs), based on absolute agreement, computed for continuous codes. Quality and clarity of reporting was highly variable across studies, so reliabilities on some variables were less than optimal. For variables with reliabilities falling at or below $\kappa = .60$ or ICC = .60 (see Cicchetti & Sparrow, 1981; Landis & Koch, 1977), two coders completed consensus coding for all studies in the sample.

Study samples, target problem, and comorbidity. We coded the sample size of treatment and control groups (ICC = .99), age at treatment onset (ICC = .99), gender (percentage of boys: mean ICC = .99), ethnicity (ICC = .99), type of target problem identified in the study as what the treatment was intended to ameliorate (e.g., conduct problems, depression; $\kappa = .91$), whether the participants had diagnosed comorbidities ($\kappa = .86$; note that study years spanned different diagnostic systems and even included years prior to widespread use of any diagnostic system), and how participants were brought into the study (recruited, clinically referred, or court referred; $\kappa = .71$). To identify the measures to be used in the analyses, we also coded whether each study measure was a measure of the study target problem ($\kappa = .73$).

Treatment settings, participants, and treatment characteristics. Coders noted whether the treatment took place in a research setting versus a real-world clinical care setting ($\kappa = .53$; because this reliability was somewhat low, two coders consensus coded this item for all studies in the sample), who the primary participants or targets of the intervention were (e.g., youths, parents, families; $\kappa = .60$), and what the theoretical orientation of the treatment was (e.g., operant, respondent or exposure based, cognitive–behavioral, eclectic or multiple interventions, case management; $\kappa = .75$). Each treatment

Table 1
Characteristics of Individual Studies Included in the Meta-Analysis

Study	Target problem	*n*[a]	Mean age (years)	% male	Type of EBT	Type of UC	Mean ES
Alexander & Parsons (1973); Klein, Alexander, & Parsons (1977); Parsons & Alexander (1973)	Delinquency	59	14.5	44.2	Behavioral family systems therapy (later renamed functional family therapy)[A,B,C,E]	Usual outpatient therapy (client-centered family groups)	0.17
						Usual outpatient therapy (psychodynamic family therapy)	0.36
Aubrey (1998)[b]	Substance abuse	39	17	78	Motivational interviewing[c] feedback[D] and UC	Usual outpatient therapy	0.31
Bank, Marlowe, Reid, Patterson, & Weinrott (1991)	Delinquency	54	14	100	Oregon Parent Management Training[A,C,D,E]	Usual outpatient therapy	0.11
Borduin, Henggeler, Blaske, & Stein (1990)	Delinquency: sexual offenses	16	14	100	Multisystemic therapy (MST)[A,B,C,E]	Usual outpatient therapy	0.83
Chamberlain & Reid (1998); Eddy & Chamberlain (2000); Eddy, Whaley, & Chamberlain (2004)	Delinquency	79	14.9	100	Multidimensional treatment foster care[A,B,C,D,E]	Usual residential care	0.42
Davidson (1976)[b]	Delinquency	24	14.5	91.7	Behavioral contracting[E] and UC	Usual probation services	0.40
Emshoff & Blakely (1983); Davidson, Redner, Blakely, Mitchell, & Emshoff (1987)	Delinquency	136	14.2	83	Behavioral contracting[E] and advocacy	Usual probation services	0.00
Fleischman (1982)	Conduct problems	64	7.5	Not provided	Oregon Parent Management Training[A,C,D,E]	Usual outpatient therapy	0.00
Fleming (1982)[b]	Conduct problems	13	11.75	100	Anger management training[D,E] and UC	Usual inpatient care and attention control	0.48
Grant (1988)[b]	Delinquency	26	15.8	100	Problem-solving training[C,D,E] and UC	Usual residential care	−0.25
Guarton (1992)[b]	Conduct problems	18	10.56	87	Anger control training[A,D,E] and UC	Usual day treatment and attention control	−0.86
					Problem-solving training[C,D,E] and UC		−0.51
Hawkins, Jenson, Catalano, & Wells (1991)	Delinquency	141	15.5	73	Cognitive–behavioral skills training[C,D,E] and UC	Usual correctional facility care	0.72
Henggeler, Borduin, Melton, Mann, Smith, Hall, et al. (1991); Henggeler, Melton, & Smith (1992); Henggeler, Melton, Smith, Schoenwald, & Hanley (1993)	Delinquency	56	15.2	77	MST[A,B,C,E]	Usual probation services	0.52

(table continues)

Table 1 (*continued*)

Study	Target problem	*n*[a]	Mean age (years)	% male	Type of EBT	Type of UC	Mean ES
Henggeler, Melton, Brondino, Scherer, & Hanley (1997)	Delinquency	140	15.2	81.9	MST[A,B,C,E]	Usual probation services	0.35
Henggeler, Pickrel, Brondino, & Crouch (1996); Brown, Henggeler, Schoenwald, Brondino, & Pickrel (1999); Henggeler, Pickrel, & Brondino (1999)	Delinquency and substance abuse	114	15.7	79	MST[A,B,C,E]	Usual probation services	−0.09
Jarden (1995)[b]	Conduct problems	50	13.5	100	Problem-solving skills training[A,B,C,D,E] and UC	Usual residential care	0.12
					Problem-solving skills training[A,B,C,D,E] and generalization component and UC		0.44
Kazdin, Esveldt-Dawson, French, & Unis (1987a)	Conduct problems	34	10.1	77.5	Oregon Parent Management Training[A,C,D,E] and problem-solving skills training[A,B,C,D,E] and UC	Usual inpatient care and attention control	0.99
Kazdin, Esveldt-Dawson, French, & Unis (1987b)	Conduct problems	33	10.9	80.4	Problem-solving skills training[A,B,C,D,E] and UC	Usual inpatient care and attention control	1.12
Kerfoot, Harrington, Rogers, & Verduyn (2004)	Depression	46	13.9	53.8	Cognitive–behavior therapy[E] and UC	Routine social services	−0.07
Luk, Staiger, Mathai, Field, & Adler (1998); Luk, Staiger, Mathai, Wong, Birleson, & Adler (2001)	Conduct problems	30	8.6	62.5	Modified parent–child cognitive–behavior therapy[A,D,E]	Usual outpatient therapy	0.00
					Behavioral family systems therapy[E]		−0.92
Mann, Borduin, Henggeler, & Blaske (1990); Borduin, Mann, Cone, Henggeler, Fucci, Blaske, & Williams (1995)	Delinquency	130	14.8	67.5	MST[A,B,C,E]	Usual outpatient therapy	0.74
Morris (1981)[b]	Delinquency	30	14.75	100	Anger control program[D,E] and UC	Usual correctional facility behavioral treatment	0.27
						Usual correctional facility behavioral treatment and attention control	0.44
Mufson et al. (2004)	Depression	63	15.10	16	Interpersonal therapy for adolescents[A,E]	Usual school-based treatment	0.43
Patterson, Chamberlain, & Reid (1982)	Conduct problems	19	6.80	69	Oregon Parent Management Training[A,C,D,E]	Usual outpatient therapy	0.52

Table 1 (*continued*)

Study	Target problem	*n*[a]	Mean age (years)	% male	Type of EBT	Type of UC	Mean ES
Scherer, Brondino, Henggeler, Melton, & Hanley (1994)	Delinquency	43	15.1	81.8	Multisystemic family preservation (variant of MST)[A,B,C,E]	Usual probation services	0.13
Snyder & White (1979)	Conduct problems	15	15.5	60	Contingency awareness group[E] and UC	Usual residential care and attention control	0.25
					Cognitive self-instruction training[D,E] and UC		1.39
Spence & Marzillier (1981)	Delinquency	74	13	100	Social skills training[E] and UC	Usual residential care	0.33
						Usual residential care and attention control	0.38
Stevens & Pihl (1982)	Anxiety and low self-esteem	32	12.5	64.6	Cognitive–behavioral coping skills training[A,D,E]	Usual school-based treatment. Some UC participants received transactional analysis, others just talked about problems of being new in high school.	0.00
Taylor, Schmidt, Pepler, & Hodgins (1998)	Conduct problems	32	5.6	74.1	Webster-Stratton's (1992) *Parents and Children Videotape Series*[A,E]	Usual outpatient therapy	0.51
Van de Wiel, Matthys, Cohen-Kettenis, & Van Engeland (2003)	Conduct problems	68	10.5	Not reported	Utrecht coping power program[A,C,D,E]	Usual outpatient therapy	0.06
Vinick (1983)[b]	Conduct problems	36	15.47	100	Group assertiveness training[E] and UC	Usual inpatient care and study hall	0.21
Whittington (1982)[b]	Delinquency	44	16	100	Assertiveness training[E] and UC	Usual correctional facility care	0.27

Note. Positive ES values indicate superior effects of the EBT; all ES values incorporate Hedges's correction for small sample size. EBT = evidence-based treatment; UC = usual care; ES = effect size. Footnotes called out with capital letters denote sources used to determine which treatments qualified as EBTs for the meta-analysis: [A]Special issue of *Journal of Clinical Child Psychology* (Lonigan et al., 1998) and special issue of *Journal of Clinical Child and Adolescent Psychology* (Silverman & Hinshaw, in press); [B]*Blueprints for Violence Prevention* (Elliott, 1998); [C] Nathan & Gorman (1998, 2002); [D]Roth & Fonagy (1996, 2004); [E]Weisz, Hawley, & Doss (2004).

[a] Sample size reflects the number of participants included when computing effect sizes at posttreatment. As the number of participants sometimes varied across outcome measures, these sample sizes are averaged across all measures included. [b] Indicates dissertations. [c]Roth & Fonagy (1996) described motivational interviewing as an evidence-based treatment for adult, not adolescent, substance abuse.

condition was also coded as to whether its description indicated that (a) therapists received pretherapy training in the specific therapy techniques that were used in the study ($\kappa = .94$), (b) adherence checks were used (including therapist supervision; $\kappa = .74$), and (c) a treatment manual or equivalent documentation was used ($\kappa = .71$). Treatment conditions were coded for format by noting whether the treatment was delivered via individual, group, or both individual and group sessions ($\kappa = .80$); whether clients received homework ($\kappa = .67$); and which form(s) of treatment contact were involved (i.e., contact between target youth and therapist, youth's parent(s) and therapist, youth's family and therapist, youth's teacher and therapist; $\kappa = .60$). Finally, treatment dose was coded in terms of total number of sessions (ICC = .98), total weeks of treatment (ICC = .97), average length of sessions (ICC = .99), and total hours in treatment (ICC = .99).

Therapists. Therapist degree (high school diploma, bachelor's degree, master's degree, or doctoral degree), professional discipline (paraprofessional, social work, psychology, or psychiatry), and vocation (practicing mental health clinician, graduate student or researcher, or paraprofessional) were coded. Because many studies used therapists from multiple categories, we coded the percentage of therapists falling into each degree and professional discipline category (ICCs = .85 for degree, .71 for professional discipline, and .56 for professional vocation).

Weeks to follow-up. We also coded the number of weeks between the beginning of a study and follow-up assessments (ICC = .94). In the EBT versus UC research genre, treatment duration is typically not equated across groups because UC cannot be constrained and still be usual. Thus, the end-of-treatment time point may represent highly variable amounts of time in treatment. Accordingly, we calculated follow-up time from beginning of treatment, so the calculation would begin at the same time point in treatment for all participants.

ES Calculation Procedures and Reliability

ES values were calculated separately for each target problem outcome measure at posttreatment and then averaged for each study. The ES used was Cohen's (1988) *d*, the posttherapy difference between the UC and EBT group means divided by the pooled standard deviation across both groups. All ESs were calculated such that positive values implied an advantage for the EBT condition over the UC condition. In addition, all ES values were adjusted using Hedges' small sample correction (Hedges & Olkin, 1985), which yields an unbiased estimator of ES.

When means, standard deviations, or other information needed for our calculations was not reported in an article or dissertation, we contacted the author(s) to obtain the information needed for ES estimation directly. The data needed to calculate or estimate ES were reported or obtained for all studies; however, for two of the studies, we were unable to estimate ES for some of the measures, so those measures were not included in the analyses. ESs were independently generated by five graduate students using the ES computer program (Shadish, Robinson, & Lu, 1999). Interrater agreement among the five coders on 30 randomly selected studies, assessed via ICC, was .88. ES values were checked for outliers on the basis of Bollen's (1989) definition of outliers as those ES values lying beyond the first gap of at least one standard deviation between adjacent ES values in a positive or negative direction; there were no such outliers.

Overview of Data-Analytic and Reporting Procedures

Reporting ES means. We pooled ES values up to the most conservative level appropriate for each test. For example, in calculating the overall EBT versus UC ES, we collapsed across treatment groups and averaged across all measures to produce a single ES mean for each study. However, in computing ES means for different treatment characteristics, we retained a single ES mean for each EBT–UC comparison rather than collapsing up to the study level.

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.

Test of homogeneity. A homogeneity analysis (Hedges & Olkin, 1985) was conducted to test the assumption that all the ES values were estimating the same population mean and to inform a decision about whether to use random versus fixed effects analyses (see below). This test was marginally significant, $Q(31) = 43.63$, $p = .07$, suggesting that the youth psychotherapy studies might not estimate common ES parameters (i.e., not all of the EBT–UC comparisons show the same results).

Random-effects versus fixed-effects analyses. As recommended by Hedges and Vevea (1998), the decision to use fixed or random-effects models was undertaken by considering both the types of inference we wished to make and the homogeneity of ES parameters. Random-effects models are appropriate for analyses that involve a heterogeneous set of ESs from which the analyst wishes to make inferences about a population of studies beyond the observed sample of studies. Because we wished to support inferences about the relative efficacy of EBTs and UC in the general population of youth mental health treatment studies and because homogeneity was rejected in our analyses, we used a random-effects approach.

Estimated power and planned analyses. Power was estimated for random-effects tests of mean ESs and mixed-effects tests of moderators. Tests of three different parameters were considered: the mean ES for all studies or a subset of studies at one level of a dichotomous moderator, the difference in mean ES between two such subsets, and the unstandardized regression slope for a continuous moderator. Power estimates were derived following Hedges and Pigott's (2001) Equation 24 and their (2004) Equations 35 and 69, respectively. Namely, each test is a large-sample *z* test of a single parameter, and its power depends on the test's significance level and directionality, the null- (H_0) and alternative-hypothetical (H_1)

values of the focal parameter (i.e., ES mean, mean difference, or slope), and the standard error of the estimated parameter. To most closely approximate the power of each test, we based computations on a standard error derived from the conditional variances and variance component estimate used in that test. For all tests, we specified a two-sided test at $\alpha = .05$ and an H_0 value of 0. Finally, we specified H_1 values of 0.50 for the mean ES and difference in mean ESs; the H_1 slope value was specified as 0.02 for publication year and percentage of participants who are Caucasian and as 0.05 for dose difference and number of weeks until follow-up. We present results for analyses with estimated power between 0.74 and 1.00 to detect the above-noted H_1 values.

Analytic procedures. For comparisons of EBT and UC group characteristics, we used McNemar's (1947) test for categorical variables (e.g., whether a treatment manual was used) and paired *t* tests for continuous variables (e.g., number of sessions). For comparison of ES values to 0, 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 investigate potential moderators of ES, we used a *Q* statistic analog to analysis of variance (ANOVA) for categorical variables and a *Q* statistic modified weighted regression approach for continuous variables (Lipsey & Wilson, 2001). All analyses were conducted using maximum likelihood, random-effects models weighted by the inverse of the variance. These moderator analyses were again done at the most conservative level appropriate to each test, with study-level moderators tested using one ES per study and group-level moderators tested using EBT versus UC comparisons. Finally, we used a paired *t* test for the within-study comparison of posttreatment and follow-up ESs.

Results

Study Characteristics

Table 1 shows sample characteristics and ES values for the full set of 32 studies. Of these, 17 focused on the treatment of delinquency and/or substance abuse, 12 on conduct problems, and 3 on internalizing problems. The mean age of participants across studies was 13.2 years ($SD = 2.9$), and the mean percentage of males was 80.1 ($SD = 20.0$).

Twenty-seven (75.0%) of the 36 EBTs tested were behaviorally oriented (e.g., behavioral contracting, social skills training, CBT, parent management training, behavioral family systems therapy, problem-solving skills training). Seven (19.4%) studies tested multisystem approaches, including multisystemic therapy ($n = 6$, the most frequently tested treatment in our sample) and multidimensional treatment foster care ($n = 1$). The remaining studies included one (2.8%) test of interpersonal therapy and one test of motivational interviewing. The 35 UC conditions tested were usual clinical therapy (outpatient, inpatient, or day treatment; $n = 18$, 51.4%), usual intervention in youth correctional or detention facilities or residential care ($n = 10$, 28.6%), or usual case management ($n = 7$, 20.0%; probation or social work case management).

Treatment and Therapist Characteristics in the EBT and UC Conditions

We first examined differences in the treatment characteristics evident in EBT and UC conditions. Table 2 shows average characteristics of the EBTs and UCs; significant EBT versus UC differences are noted with asterisks. Studies were significantly less likely to report the format (e.g., individual vs. group therapy; $p < .001$) and participants (e.g., child sessions, parent sessions, family sessions, teacher sessions; $p < .01$) of UC sessions than EBT sessions. When studies provided format and participant information for both the UC and the EBT groups (18 studies for format, 27 studies for participants), the groups did not differ on those characteristics. Studies were also significantly less likely to report assigning therapy homework ($p < .001$) in the UC groups than in the EBT groups. As might be expected, the EBT groups showed more effort to support treatment integrity; they were more likely to report pretreatment therapist training ($p < .001$), therapist supervision or other adherence checks ($p < .001$), and the use of treatment manuals to guide treatment delivery ($p < .001$; no UC group reported using a manual). Although EBT procedures were generally described in some detail, we found descriptions of the contents of most UC interventions to be either absent or too thin to permit adequate characterization or classification.

Studies were also significantly less likely to report dose information for UC groups than for EBT groups ($p < .01$). When all available information on dose was examined together to determine which group for each comparison received a larger dose of treatment, the EBT group received a greater dose in 25 EBT–UC pairs (including studies in which UC dose was not reported but the authors reported that the EBT was provided in addition to UC for the EBT group), the UC group received a larger dose in 4 pairs, and the EBT and UC groups had an equal dose in 6 pairs.

Table 3 describes therapist characteristics for the EBT and UC groups. Studies were significantly less likely to report the degree ($p < .001$) or professional discipline ($p < .001$) of their UC therapists than their EBT therapists. There were 24 EBT–UC pairs for which therapist vocation was noted for both conditions, 16 pairs with therapist degree noted for both conditions, and 11 pairs with therapist discipline noted for both conditions. For these cases, EBT therapists were significantly more likely to be researchers or graduate trainees than were UC therapists ($p < .05$). There were no significant differences between conditions for therapist degree or discipline.

Mean EBT Versus UC ES Posttreatment and at Follow-up

Averaging across the 32 studies, we found the mean WLS ES for EBT versus UC was 0.30, indicating that the average youth treated with an EBT was better off after treatment than 62% of youths who received UC. This effect was significantly different from 0 ($z = 4.82$, $p < .0001$). This value falls between conventional cutoffs for small (0.20) and medium (0.50) effects (based on Cohen, 1988). Sixteen

Table 2
Treatment Characteristics of EBTs and UC

Characteristic	EBT				UC			
	%	*M*	*SD*	*n*	%	*M*	*SD*	*n*
Format of treatment sessions								
Groups using individual sessions	41.7*				17.1			
Groups using group sessions	13.9				8.6			
Groups using individual and group sessions	30.6				20.0			
Groups not reporting format	13.9				54.3***			
Treatment participants								
Groups involving youths in sessions	80.6				68.6			
Groups involving parents in sessions	41.7				8.6			
Groups involving families in sessions	36.1				17.1			
Groups involving teachers in sessions	22.2				0.0			
Groups not reporting participants	0.0				25.7**			
Groups w/any homework assigned	72.2***				8.6			
Steps to support treatment integrity								
Groups w/pretreatment therapist training	36.1***				0.0			
Groups w/supervision/adherence checks	50.0***				8.6			
Groups using treatment manuals	47.2***				0.0			
Groups reporting structured treatments	16.7				0.0			
Treatment dose								
Groups not reporting dose	8.3				45.7**			
Mean number of sessions		15.9*	8.7	22		12.0	6.3	13
Mean number of weeks		12.6	7.1	28		15.5	10.6	13
Mean length of sessions in minutes		31.2	22.3	19		48.5	20.9	8
Mean total hours of treatment		19.6	13.1	24		18.3	16.2	11

Note. Significant differences between EBT and UC groups are indicated by asterisks next to the group that was significantly higher. EBT = evidence-based treatment; UC = usual care.
* $p < .05$. ** $p < .01$. *** $p < .001$.

studies included follow-up data, collected a mean of 65.4 weeks ($SD = 57.4$ weeks) after the study began. The mean ES value for these studies was 0.38 ($z = 2.85$, $p < .01$). This ES value was actually larger than that obtained for the posttreatment data for those 16 studies (mean ES = 0.32, $z = 3.37$, $p < .001$), but this difference was not statistically significant, $t(15) = 0.93$, $p = .37$. We also tested whether the number of weeks between the beginning of treatment and the follow-up assessment was significantly correlated with magnitude of the follow-up ES. It was not ($B = 0.002$, $z = 1.06$, $p = .29$).

Did the EBT Versus UC Difference Differ Across Study Years?

We assessed whether the magnitude of the EBT–UC difference changed over the years. Superiority of EBTs might be expected to increase over time if, for example, the lessons learned through research have led to increasingly effective EBTs. As an alternative, the superiority of EBTs might fade over time if, for example, years of clinical experience have improved UC or if information from clinical trials research has led to improvements in UC. To test these possibilities, we assessed the association between study year and EBT versus UC ESs. The relationship was not significant ($B = -0.003$, $z = -0.44$, $p = .66$).

Treatment and Therapist Characteristics as Moderators of ES

Next we examined ES as a function of treatment and therapist characteristics. We focused on variables that previous literature (e.g., Kazdin & Weisz, 2003; Weisz, 2004) has suggested might account for superior effects of EBT over UC conditions. One caveat warrants attention at the outset. The grouping of studies for moderator tests cannot be fully sensitive to all the potentially relevant conceptual issues related to specific treatments. As an example, in one test, we assessed whether the size of the EBT versus UC difference was related to the use of homework in EBTs. Most EBTs emphasize homework, and most of the UC conditions did not appear to emphasize homework, as best we could tell from the descriptions provided, but there were likely exceptions in both cases. Thus, a more precise analysis might have included the degree of homework emphasis in EBT and UC as a factor, but we lacked sufficient information from study descriptions to make such an analysis possible.

Was the superiority of EBTs over UC accounted for by a larger dose of treatment in EBTs? To explore whether EBTs obtained superior effects simply by providing more treatment, we tested

Table 3
Therapist Characteristics of EBTs and UC

Characteristic	EBT			UC		
	M	*SD*	%	*M*	*SD*	%
Therapist vocation						
% graduate student or researcher	20.6*	36.0		2.9	11.8	
% clinician	33.0	43.0		45.6	46.0	
% nonclinician or nonresearcher	11.5	28.8		13.6	27.9	
Groups not reporting vocation			19.4			28.6
Therapist degree						
% therapists with high school diploma	6.9	24.4		0.0	0.0	
% therapists with bachelor's degree	7.5	19.7		2.7	9.2	
% therapists with master's degree	40.0	41.0		29.2	38.2	
% therapists with doctoral degree	10.4	25.8		6.9	16.9	
Groups not reporting therapist degrees			22.2			54.3***
Therapist professional discipline						
% paraprofessional	15.5	33.0		4.7	18.3	
% social workers	11.9	24.8		9.5	23.8	
% psychologists	25.9	36.9		11.6	23.7	
% psychiatrists	1.9	8.7		0.9	3.9	
Groups not reporting discipline			33.3			65.7***

Note. Significant differences between EBT and UC groups are indicated by asterisks next to the group that was significantly higher. EBT = evidence-based treatment; UC = usual care.
* $p < .05$. *** $p < .001$.

whether the mean ES for the 25 cases in which the EBT group received a greater dose of treatment than the UC group did was significantly different from the mean ES for the 10 cases in which the UC group received more treatment or the dose was equal. The mean ES for the 25 comparisons in which the EBT group received a higher dose of treatment was 0.29 ($z = 3.94$, $p < .001$), and the mean ES for the 10 comparisons with a higher dose in UC or an equal dose was 0.25, marginally significantly different from 0 ($z = 1.78$, $p = .07$). This difference in ESs was not significant, $Q(1, 34) = 0.08$, $p = .77$. This dichotomous categorical approach to the dose question allowed us to include a particularly large proportion of our pool of studies in the analysis. But, as a complement, we included a more precise continuous variable analysis as well, investigating whether ES values could be predicted from the difference between EBT and UC doses for each study. We used the difference in total hours of treatment for these analyses, as total hours was the most commonly reported indicator of dose across studies. The difference was computed as hours of EBT minus hours of UC, so that a positive relation between this variable and ES would indicate that studies in which EBTs had a higher dose than UC also had higher ES values. Using the difference also allowed for the use of groups that did not specify the amount of UC received but did state that the EBT group received a specific amount of treatment in addition to UC (e.g., if the EBT group was provided with six hours of CBT in addition to UC, the difference in dose was considered six hours). This variable was available for 20 studies. The difference in hours of treatment was not a significant predictor of ES ($B = 0.005$, $z = 0.74$, $p = .46$).

Was the superiority of EBTs over UC due to the use of homework to facilitate treatment generalization? EBT descriptions were significantly more likely than UC descriptions to note inclusion of therapy homework assignments (see Table 2). We computed mean ES separately for the 29 comparisons in which the EBT group was assigned homework and the 10 comparisons in which they were not. The WLS ES mean for comparisons involving homework was 0.33 ($z = 5.07$, $p < .001$); the mean for studies not involving homework was only 0.15, not significantly different from 0 ($z = 1.60$, $p = .19$). However, this ES difference was not statistically significant, $Q(1, 38) = 1.87$, $p = .17$.

Was the superiority of EBT over UC due to efforts to ensure treatment integrity? EBT descriptions were significantly more likely than UC descriptions to note the use of pretherapy training, treatment manuals, and adherence checks (see Table 2). We computed mean ES separately for the 14 comparisons for which pretherapy training was reported for the EBT group and the 25 comparisons in which no such training was reported for the EBT condition. Mean ES was 0.24 ($z = 2.83$, $p < .01$) for comparisons with pretherapy training and 0.33 ($z = 4.28$, $p < .0001$) for comparisons not reporting pretherapy training; this difference was not statistically significant, $Q(1, 38) = 0.57$, $p = .45$. We also computed mean ES for the 19 comparisons in which ongoing supervision or adherence checks were reported

for the EBT condition and the 20 comparisons for which no supervision or adherence checks were reported. ES was 0.26 ($z = 3.33$, $p < .001$) for comparisons for which supervision or adherence checks were reported and 0.32 ($z = 3.74$, $p < .001$) for comparisons without such checks; this ES difference was not significant, $Q(1, 38) = 0.34$, $p = .56$. Finally, we computed separate mean ES values for EBTs that had manuals ($n = 19$) and EBTs for which no manual was reported ($n = 20$). ES was 0.34 ($z = 4.46$, $p < .0001$) for comparisons with EBTs having manuals and 0.21 ($z = 2.36$, $p < .05$) for comparisons with EBTs without manuals; this ES difference was also not significant, $Q(1, 38) = 1.16$, $p = .28$.

Finally, we computed mean ES separately for the 12 comparisons that reported pretherapy training, ongoing adherence checks, and treatment manuals for the EBT group and the 16 comparisons that reported none of these for the EBT group. Mean ES was 0.21 ($z = 2.51$, $p < .05$) for comparisons that reported all three treatment integrity elements and 0.28 ($z = 3.09$, $p < .01$) for comparisons that did not; this difference (in the opposite direction of what might be expected) was not statistically significant, $Q(1, 27) = 0.29$, $p = .59$.

Was the superiority of EBT over UC accounted for by the use of research therapists to deliver the EBTs? To determine whether EBTs outperformed UC because the EBT therapists were research therapists or graduate students (e.g., from within the investigators' research teams), we compared the mean ES of the 10 comparisons that reported that research therapists or graduate students delivered the EBTs with the mean ES of the 20 comparisons that did not (7 comparisons not reporting therapist vocation for EBT therapists and 2 comparisons reporting that graduate student therapists delivered the UC were excluded from the analysis). The mean ES value for comparisons in which research therapists delivered the EBT was 0.43 ($z = 3.80$, $p < .001$), compared with a mean ES of 0.25 ($z = 3.78$, $p < .001$) for comparisons in which EBTs were not delivered by research therapists. This difference, although substantial, was not significant, $Q(1, 29) = 1.86$, $p = .17$.

Sample Characteristics as Moderators of ES

Next, we examined sample characteristics as moderators of the difference between EBT and UC conditions.

Was the superiority of EBT over UC only evident in samples of voluntary treatment seekers? We compared the ES for the 13 studies involving participants for whom participation in mental health services was voluntary (e.g., clinically referred youths from outpatient, inpatient, day treatment, or residential clinics) with the 17 studies in which participation was not a voluntary choice (e.g., youths incarcerated or court ordered to treatment). The mean ES for the studies with voluntary participants was 0.31 ($z = 2.80$, $p < .01$); the mean ES for the studies with nonvoluntary service-seeking participants was 0.33 ($z = 4.15$, $p < .0001$). This ES difference was not significant, $Q(1, 29) = 0.03$, $p = .88$.

Was the superiority of EBT over UC reduced as ethnic minority representation increased? We tested the possibility that EBTs might not be as effective with samples that include ethnic minority youth, in part because the EBTs were not originally designed specifically for such youth (see, e.g., discussion in Bernal & Scharrón-del-Río, 2001). Twenty studies provided information on the ethnicities of their participants. For those studies, we predicted ES from the percentage of participants who were Caucasian. ES did not change significantly with changes in the proportion of Caucasian versus minority youth ($B = 0.004$, $z = 1.26$, $p = .21$).

Was the superiority of EBTs over UC reduced in more severe samples? Next we tested whether the superiority of EBTs over UC was diminished or vacated in samples of participants with severe psychopathology, a possibility discussed in the introduction. For this test, we defined severe samples as those involving youths who were (a) inpatients, (b) incarcerated, (c) previously arrested, or (d) living in residential institutions. The mean ES for the 21 studies with these high-severity samples was 0.32 ($z = 4.09$, $p < .0001$); the mean ES for the 11 studies with low-severity samples was 0.27 ($z = 3.52$, $p < .05$). The difference in ES was not significant, $Q(1, 31) = 0.14$, $p = .71$.

Was the superiority of EBTs over UC reduced in comorbid samples? To test the possibility that the advantage of EBTs over UC would be diminished by the presence of comorbidity, as discussed in the introduction, we compared the mean ES for the 7 studies reporting diagnosed comorbidity in some or all of their sample with the mean ES for the 25 studies not reporting diagnostic comorbidity. The mean ES for studies in the comorbidity group was 0.26 ($z = 1.87$, $p = .06$); the mean ES for studies not reporting diagnosed comorbidity was 0.31 ($z = 4.50$, $p < .0001$). This difference in ES values was not significant, $Q(1, 31) = 0.08$, $p = .77$.

Design Characteristics as Moderators of ES

Next we turned to design factors that might have accounted for the superiority of EBT over UC conditions.

Was the EBT versus UC difference influenced by differences in therapist pools? In some of the EBT versus UC comparisons, therapists for the two conditions were drawn from the same pool of individuals; in other comparisons, therapists were drawn from entirely different pools (e.g., researcher therapists delivering the EBT vs. staff of a residential facility delivering UC). We sought to learn whether the magnitude of the EBT–UC difference was related to whether investigators selected therapists from the same pool (this was done in 6 of 39 comparisons) and, when they did so, whether individuals from that pool were randomly assigned to EBT versus UC conditions. We could address the first question but not the second, because only two studies involved random assignment of therapists from the same pool. Our analyses showed that the mean ES value for the 6 comparisons involving therapists from the same pool was 0.22 ($z = 1.54$, $p = .13$). This ES value, although smaller, did

not differ significantly from the 33 comparisons that did not involve therapists from the same pool (mean ES = 0.30, $z = 4.79$, $p < .0001$), $Q(1, 38) = 0.29$, $p = .59$.

Was the EBT versus UC difference influenced by differences in treatment setting? In some of the EBT versus UC comparisons, interventions for the two conditions were carried out in the same setting; in other comparisons, entirely different settings were used (e.g., therapist's office and home visits vs. probation office). Because setting characteristics may influence intervention effects, we sought to determine whether the magnitude of the EBT–UC difference was related to whether investigators used the same setting ($n = 26$ comparisons) or different settings ($n = 4$ comparisons; the 9 comparisons in which either setting was unknown were excluded from this analysis). The mean ES value for the comparisons taking place in the same setting was 0.28 ($z = 3.64$, $p < .001$); the mean ES for the comparisons taking place in different settings was 0.45 ($z = 2.91$, $p < .01$). This difference in ES values was not significant, $Q(1, 29) = 0.99$, $p = .32$.

Is the superiority of EBT over UC due to the inclusion of studies that did not involve pure comparisons of a psychotherapy EBT to psychotherapy UC? To include the most complete collection of studies relevant to the EBT versus UC question, we considered studies in which participants received medications in addition to therapy, studies in which the EBT was administered in addition to UC, studies in which a psychotherapy placebo was administered in addition to UC, and studies in which UC involved case management services (e.g., probation and referral) that may or may not have included significant doses of psychotherapy. To assess whether including these studies might have led to larger ESs than would have been obtained had we focused exclusively on studies where a pure EBT psychotherapy condition (i.e., without medication or other services) was compared with a pure usual psychotherapy condition (i.e., where subjects clearly received mental health services), we used a sequential approach, dropping studies in stages, moving progressively closer to a pool of pure EBT versus UC comparison studies. This analysis was done at the group level, because groups varied within studies on characteristics used in the sequential approach.

To avoid potential confounds associated with participants taking medications in addition to their psychotherapy, we first dropped comparisons in which the authors indicated that some or all of the participants in the EBT or UC group were taking psychotropic medications during the study. The resulting sample of 30 EBT versus UC comparisons had a mean WLS ES value of 0.33, which was slightly larger than our overall mean WLS ES of 0.30 and still significantly greater than 0 ($z = 5.00$, $p < .0001$).

We then dropped studies in which the EBT was administered in addition to UC, to avoid the possibility that these studies artificially increased the overall ES simply by providing more services to participants. Dropping these studies also removed all studies where a psychotherapy placebo condition was added to the UC regimen, thus changing the nature of the UC services. The resulting sample of 16 comparisons had a mean WLS ES value of 0.26, somewhat smaller than those reported above but still significantly greater than 0 ($z = 3.14$, $p < .01$).

Finally, we dropped studies where the UC condition did not explicitly include formal psychotherapy, to test whether the ES difference between the EBTs and UC could be explained by the fact that many compared a treatment condition with some form of case management services that might not entail formal psychotherapy for all cases assigned to the condition (e.g., usual probation services where many, but not all, are referred for therapy). The resulting sample of 11 pure comparisons had a mean WLS EBT versus UC ES value of 0.25, which was only marginally greater than 0 ($z = 1.87$, $p = .06$), due in part to small sample size. This ES was also marginally lower than the overall mean WLS ES of 0.30 obtained for the 28 comparisons that had been dropped from the analysis, $Q(1, 38) = 0.06$, $p = .81$.

Was the estimation of the EBT versus UC ES affected by including dissertations? It is possible that the ESs in the present analyses—which included dissertations that may have been methodologically weaker than published studies or EBTs that may not have been so skillfully delivered—were smaller than would have been obtained had we relied solely on studies that had passed through the peer review process for publication. To test this hypothesis, we compared ESs from the 23 published studies with those of the 9 unpublished dissertations. The mean ES value for the published studies was 0.32 ($z = 4.69$, $p < .0001$), compared with the dissertation mean ES of 0.20, which did not differ significantly from 0 ($z = 1.38$, $p = .17$). The ES difference between published and unpublished studies was not statistically significant, $Q(1, 31) = 0.59$, $p = .44$.

Was the superiority of EBT over UC greater for official empirically supported treatments (ESTs)? To be as inclusive as possible, we included in the present study EBTs from several reviews in addition to the official lists of ESTs for youths published in the 1998 special issue of the *Journal of Clinical Child Psychology* (Lonigan & Elbert, 1998) and updates in press at the *Journal of Clinical Child and Adolescent Psychology* (Silverman & Hinshaw, in press). It is possible that the ESs in the present analyses, which included EBTs with some empirical support but not the level required to make the official EST lists, were smaller than would have been obtained had we relied solely on official ESTs. To test this hypothesis, we compared ESs from comparisons that involved official ESTs with those that did not. The mean ES for the 19 comparisons involving official ESTs was 0.35 ($z = 4.62$, $p < .0001$); the mean ES for the 20 comparisons not involving official ESTs was 0.21 ($z = 2.50$, $p < .05$). This difference, although substantial, was not statistically significant, $Q(1, 38) = 1.63$, $p = .20$.

Was the EBT versus UC difference larger in studies conducted by the EBT developer? Finally, we compared ESs in studies in which the EBT developer was involved as a study author or, for disserta-

tions, a member of the author's dissertation committee versus ESs from studies in which neither was the case. There are several reasons why some might expect larger EBT versus UC differences in studies where EBT developers have been involved. It is possible, for example, that some treatment developers may be especially vigilant about study design and procedures when their own handiwork—that is, their treatment program—is under close scrutiny in competition with an alternative intervention. It is also possible that the strength of a particular EBT will be maximized in a trial where such critical elements as therapist training and supervision and fidelity monitoring can be overseen by the developer, who knows the treatment program particularly well. Whatever the reason, we did find that the mean ES for the 30 comparisons in which the EBT developer was author or advisor was 0.33 ($z = 5.24$, $p < .0001$) and that the mean ES for the 9 remaining comparisons was 0.09 ($z = 0.61$, $p = .54$). The ES difference was not significant, $Q(1, 38) = 2.53$, $p = .11$.

Discussion

The question of whether evidence-based youth treatments outperform UC is both hotly debated and highly significant in its implications for psychology and other mental health disciplines. Proponents of EBTs have maintained that interventions that have been tested and shown to work are more likely to be beneficial than interventions that have not been tested empirically and/or have not been shown to work. However, many in the field (cited in the introduction) have expressed concern that the evidence supporting EBTs is flawed in important ways and that the EBTs themselves may be so structured, preplanned, inflexible, and hard to individualize that they will be unable to match the potency and effectiveness of real-world clinical care by providers who are not constrained by protocols listed in manuals. Some have argued that these limitations may be especially problematic with severe cases and with minority youth. Previous reviews bearing on this debate have involved indirect comparison of EBT and UC outcomes based on separate collections of EBT studies and UC studies. Although mean effects have been larger in collections of structured treatments described in manuals than in separate collections of UC interventions, numerous differences between the pools of studies have made the findings difficult to interpret with confidence. In the present meta-analysis, we sought to overcome this problem by examining a collection of studies in which EBTs were directly compared with UC, with youths randomly assigned to EBT versus UC conditions.

Our findings support the view that EBTs have generally outperformed UC in direct, randomized comparisons. With our full sample of 32 studies, we found a mean ES of 0.30, falling between conventional benchmarks for small and medium effects (see Cohen, 1988). With more restrictive comparisons, the mean effect was reduced somewhat (e.g., to 0.25 for the most restrictive pure EBT vs. UC studies). The magnitude of the EBT versus UC effect is lower than the average effect of 0.54 found in the most recent broad-based meta-analyses of youth psychotherapy trials (Weisz et al., 1995). However, that larger effect was based on comparisons of active treatments with control conditions, most of which were passive or inert (e.g., no treatment, attention control, or waitlist groups). The present EBT versus UC studies were all comparisons between two active interventions, both intended to have beneficial effects; treatment-versus-treatment comparisons are known to show markedly more modest effects than treatment–control comparisons (see, e.g., Kazdin, Bass, Ayers, & Rodgers, 1990).

Our findings have implications for both clinical science and clinical practice. On the clinical science front, the findings may be relevant to the culture of treatment testing. Treatment developers are under considerable pressure (in the grant review process and in journal publication practices) to not simply replicate findings from one study to the next but rather to make each study innovative and push the boundaries of what has been done. This press for innovation and risk may limit opportunities to simply pile up a series of large ESs by repeating successful study procedures over and over again. As an example of pushing the boundaries to learn something new, the study in our collection by Kerfoot, Harrington, Rogers, and Verduyn (2004) warrants attention. Kerfoot et al. took a chance by placing considerable responsibility for learning and implementing CBT for youth depression in the hands of practitioners within a service system. By not using efficacy trial procedures to ensure CBT fidelity and skill on the part of the practitioners, Kerfoot et al. sought new information but risked a low ES in comparison to UC; they found just that (ES = −0.07; see Table 1). Studies like this and others in a field where innovation and risk are the norm may well lead to a lower mean ES than if study authors were simply repeating their most successful approaches to accumulate a series of large effects. This risk is magnified by the fact that UC is a heterogeneous category, encompassing diverse interventions and providers, some likely to be quite effective. Thus, any specific EBT may or may not fare well in comparison to UC, depending on the potency of the particular variant of UC that prevails in the setting being studied. In sum, the mean ES found in this meta-analysis may be best viewed not as the largest that could possibly be found in comparing EBTs and UC but rather as a picture of what has been found to date given a broad array of UC contrasts and a research climate that discourages safe replication and encourages novelty and risk.

A closely related form of risk arises when treatments are taken out of the safety zone of efficacy trial conditions and tested in *effectiveness trials*—that is, with clinically referred groups in representative clinical practice conditions. Under these circumstances, it has been suggested (e.g., in Weisz, 2004) that simple replication of EBT procedures exactly as used in efficacy trials may not work well, given the numerous differences that exist between the conditions of efficacy trials and the conditions of clinical practice. To address these differences, we have proposed a *deployment-focused model* of treatment development and testing to guide the process of adapting EBTs and fitting them to clinical practice contexts (see Weisz, 2004). Al-

though relatively few of the studies in the present meta-analysis would qualify as full-fledged effectiveness trials, the studies did involve some effectiveness elements, including efforts to bring tested treatments to the youths, clinicians, and/or settings of everyday intervention, albeit often with little adjustment or adaptation of the treatments noted. It is possible that the most effective EBTs ultimately will be those that have undergone an iterative process involving stages of adaptation and testing to maximize their impact within the most relevant UC contexts. Because this process is likely to be an extended one, the most successful EBTs relative to UC may well be many years in the making and may differ in substantial ways from the current generation of treatments.

As for the current evidence base, a good deal may be learned from the variability of effects across studies. Four studies in our collection showed large effects (by Cohen's 1988 standards) favoring EBTs, suggesting the possibility that certain treatments (e.g., multisystemic therapy, parent management training, problem-solving skills training, and cognitive self-instruction training) may have unusual potential to improve on everyday care in certain intervention contexts. Five additional studies showed medium to large effects favoring EBTs. At the other extreme, five studies found UC to be superior to EBTs (although not all differences were significant; see Table 1). It certainly seems likely that some forms of UC, administered by skilled practitioners, will have beneficial effects. It also seems likely that some ordinarily beneficial EBTs could be delivered by some therapists in unskilled ways that do not help their young clients. Understanding such individual variations and the factors responsible for them could be enormously helpful to the field. More broadly, the genre of research comparing specific structured treatments with usual clinical practices may hold considerable promise for the field as the array of EBTs expands, particularly if complemented by an increasingly well-specified taxonomy of usual clinical practices for various types of referred individuals. Under these conditions, it should be possible to identify with increasing precision those situations in which introducing evidence-based, manual-guided treatments is—and is not—likely to improve outcomes and thus warrant the effort required to alter current practice.

As for current practice, those who plan and deliver clinical care are apt to find the mean effect across all of our EBT versus UC comparisons less useful than evidence from the specific individual comparisons. The fact that some EBTs looked particularly strong in comparison to UC whereas others did not highlights the need for careful selection of treatments. Further complicating the picture is the heterogeneity of UC and the likelihood that some forms of UC may work better than others and may outperform EBTs for certain target problems. Taken together, these facts suggest that those who are selecting treatments for youths in clinical care cannot safely assume that any EBT they choose will improve on any form of UC. Instead, our findings argue for a more precise strategy: detailed examination of the evidence base to identify specific treatments that appear strong for the target problems of interest in the setting of interest combined with frequent assessment of youth treatment responses to gauge whether the interventions selected are, in fact, having the desired effect.

Discussion of clinical practice applications brings us back to the questions many have raised about the real-world utility of EBTs, as discussed in the introduction. Concerns have been raised that EBTs may not fare well with complex, severe, or comorbid cases and that the preplanned and structured nature of many EBTs may hamper therapists' ability to individualize treatment, build a strong therapeutic relationship, respond to unexpected events in the course of therapy, or fit treatments to the needs of minority youths and families (see discussions in Addis & Krasnow, 2000; Addis & Waltz, 2002; Bernal & Scharrón-Del-Río, 2001; Garfield, 1996; Hall, 2001; Havik & VandenBos, 1996; Strupp & Anderson, 1997; Westen et al., 2004a, 2004b). Although we grouped studies in a variety of ways for different EBT versus UC contrasts, the groupings rather consistently showed EBTs to be superior to UC comparison conditions. This suggests that whatever challenges EBTs may confront in treating youths in real-world care situations, those challenges have not prevented EBTs from demonstrating generally superior effects. Significantly, we found that EBTs fared about equally well with more and less severe groups (e.g., inpatient and incarcerated youth vs. outpatients) and that ESs were not significantly reduced as the percentage of minority youths in study samples increased.

However, our findings do not necessarily invalidate all the concerns raised by critics of EBTs. The relatively modest magnitude of the EBT versus UC ESs obtained in our analyses, notwithstanding the previous discussion, do suggest that considerable room for improvement remains. It is quite possible that EBTs could be strengthened by being altered to address concerns raised by critics—for instance, by finding better ways to address comorbidity; introducing greater flexibility in therapist use of EBT components; progressively adapting EBTs to fit the conditions of usual clinical care; and, more broadly, conducting intervention research in collaboration with mental health providers (see, e.g., Chorpita, Delaiden, & Weisz, in press; Garland, Hurlburt, & Hawley, 2006; Weisz, 2004).

It is intriguing to consider our findings in light of the literature on the *dodo bird verdict* that "everybody has won, and all must have prizes" (see, e.g., Luborsky et al., 2002; Rosenzweig, 1936). This is the notion that different therapies will have similar effects, in part because common factors (e.g., therapeutic alliance) drive treatment outcome more than do specific features of specific interventions. From one perspective, our findings are inconsistent with the dodo bird verdict: Across several ways of structuring EBT versus UC comparisons, our results rather consistently showed effects favoring EBT over UC. This suggests that there may be specific factors inherent in EBTs that provide an advantage over UC. However, from another perspective, our findings are not totally inconsistent with the dodo bird notion, in that the overall magnitude of effects was relatively modest. Moreover, some evidence does link therapeutic alliance to treatment benefit in youth

mental health care across significant variations in the specific therapeutic procedures used (see, e.g., Hawley & Weisz, 2005; McLeod & Weisz, 2005; Shirk & Karver, 2003). Understanding the extent to which treatment effects derive from specific intervention procedures, such as those of EBTs, versus more general common factors, such as therapeutic alliance, remains a key task for youth treatment researchers, notwithstanding the evidence presented here on the relative advantage of EBTs over UC.

A meta-analysis of studies comparing EBTs to UC provides a close look at the quality of research on the topic and can thus generate ideas for how the research can be improved. We recognize that many of the studies in our collection were originally designed mainly for purposes other than EBT versus UC comparison and that the methodology in some of these studies may have fit their primary goals well. However, for those planning future studies to directly compare EBTs with UC, we stress that the most interpretable findings are generated by studies in which the EBT and UC conditions are well-matched for factors that might influence treatment outcome. In our collection of studies, relatively few comparisons showed extensive matching in this regard. Intervention dose, for example, was explicitly equated across conditions in only 6 of the 37 studies. We had hoped, at one point, to create a well-matched set of studies in which the EBT and UC conditions had (a) drawn youths from exactly the same pool and randomized them to EBT and UC, (b) drawn therapists from the same pool and randomized them to EBT and UC, (c) arranged for EBT and UC interventions to be delivered in the same setting, and (d) provided equal or at least very similar doses of treatment in the EBT and UC conditions. Although we required all of our studies to meet the first criterion, the number of studies in our collection of 32 that met all four criteria for the most balanced comparison was 0. Our experience left us convinced that there is significant room for improvement in research design in comparisons of EBT and UC.

There is also significant room for improvement in coverage of treated conditions and in study-reporting procedures. On the coverage front, note that most of the studies to date have addressed delinquency and conduct problems. With only two studies on depression and one on anxiety, we can say little thus far about the relative strength of EBTs and UC in the treatment of internalizing problems and disorders. On the study-reporting front, we see a number of problems. One reason for our inability to create a pool of well-matched studies, as noted in the previous paragraph, was that so many of the studies failed to provide necessary information: for example, on therapists (who they were, what their qualifications and characteristics were, how they were selected, whether they were trained, whether EBT and UC therapists came from the same pool, and whether they were randomly assigned to condition), on the interventions (details of the EBT intervention procedures; the UC intervention procedures; and the dose, amount, and duration of each), and on the location(s) (where the intervention occurred and what the setting was like). Even addressing the timely and important question of ethnicity as a moderator of treatment impact was made difficult by the fact that only 20 of the 32 studies provided ethnicity information on their samples. More ample reporting on participants, therapists, interventions, and settings would make studies more valuable and increase the potential for informative reviews and meta-analyses.

Although all of the reporting problems identified here were significant, the most striking was the extreme rarity of reporting on what actually happened (i.e., what procedures were used) in the UC intervention conditions. Moreover, authors reporting UC versus EBTs were significantly less likely to report treatment format, participants, dose, therapist degree, and even professional discipline. If psychotherapy research is to take seriously the notion that effective treatments may be found within current clinical practices (see, e.g., Garland et al., 2006; Weisz, 2004; Westen et al., 2004a, 2004b), it should be clear that an essential starting point is a clear description of what those current practices are and who is delivering them to whom and in what contexts. More information on UC in studies comparing such care with EBTs could help generate an increasingly complete account of the kinds of mental health care that are and are not outperformed by EBTs. Careful documentation could help investigators identify specific kinds of treatments in clinical care that show beneficial effects, warranting further testing and perhaps ultimately dissemination. By contrast, not even those studies in the present meta-analysis that showed UC outperforming EBTs offered much to build on, because the descriptions did not convey in much detail just what the effective UC procedures were, what kinds of therapists provided them, and to exactly what kinds of youths. Much of what we found to be missing (e.g., participant ethnicity, therapist discipline) could be added with only minor effort by investigators and with no new methodology required. But even such complex phenomena as the contents of UC interventions can now be characterized both through therapist report (see Weersing, Weisz, & Donenberg, 2002) and coder review of session recordings (see, e.g., McLeod, 2004). Thus, in principle, it is quite possible to address in future research the reporting problems noted here. The findings of the present meta-analysis, combined with the important kinds of information we needed but could not obtain, highlight the need for much more thorough documentation in studies comparing UC with EBTs.

Although the findings do highlight information gaps, they also provide a rather uniform answer to the main question addressed in this meta-analysis: In our principal analyses, across numerous ways of structuring comparisons and varying the collections of studies being compared, EBTs produced better outcomes than did usual clinical care. Differences across the various analyses involved variations in the magnitude of effects, not in their direction. The evidence synthesized here does indicate that EBTs have moderately outperformed UC in a variety of contexts. Summary evidence of this kind is useful in a number of ways. However, from the perspective of both researchers and practitioners, the most important next steps may well be those taken in relation to specific treatments, provided

by specific clinicians, for specific groups of youths, with specific problems and disorders, in specific contexts where specific forms of UC prevail. In work focused at this level, learning which treatments are most helpful and why will require significant advances in both research design and research reporting in the years ahead.

REFERENCES

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

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68,* 331–339.

Addis, M. E., & Waltz, J. (2002). Implicit and untested assumptions about the role of psychotherapy treatment manuals in evidence-based mental health practice. *Clinical Psychology: Science and Practice, 9,* 421–424.

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

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 57–87.

*Aubrey, L. L. (1998). Motivational interviewing with adolescents presenting for outpatient substance abuse treatment. *Dissertation Abstracts International, 59*(3-B), 1357. (UMI No. 9826598)

*Bank, L., Marlowe, J. H., Reid, J. B., Patterson, G. R., & Weinrott, M. R. (1991). A comparative evaluation of parent-training interventions for families of chronic delinquents. *Journal of Abnormal Child Psychology, 19,* 15–33.

Bernal, G., & Scharrón-Del-Río, M. R. (2001). Are empirically supported treatments valid for ethnic minorities? Toward an alternative approach for treatment research. *Cultural Diversity and Ethnic Minority Psychology, 7,* 328–342.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

*Borduin, C. M., Henggeler, S. W., Blaske, D. M., & Stein, R. J. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 34,* 105–113.

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

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

Calhoun, K. S., Moras, K., Pilkonis, P. A., & Rehm, L. (1998). Empirically supported treatments: Implications for training. *Journal of Consulting and Clinical Psychology, 66,* 151–162.

*Chamberlain, P., & Reid, J. B. (1998). Comparison of two community alternatives to incarceration for chronic juvenile offenders. *Journal of Consulting and Clinical Psychology, 66,* 624–633.

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., et al. (1998, Winter). Update on empirically validated therapies, II. *The Clinical Psychologist, 51,* 3–16.

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

Chorpita, B. (2003). The frontier of evidence-based practice. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 42–59). New York: Guilford Press.

Chorpita, B. F., Delaiden, E., & Weisz, J. R. (in press). Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology.*

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86,* 127–137.

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

Compton, S. N., Burns, B. J., Egger, H. L., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Internalizing disorders. *Journal of Consulting and Clinical Psychology, 70,* 1240–1266.

Costello, E. J., Angold, A., & Burns, B. J. (1996). The Great Smoky Mountains Study of Youth: Goals, designs, methods, and the prevalence of *DSM–III–R* disorders. *Archives of General Psychiatry, 53,* 1129–1136.

*Davidson, W. S., II. (1976). The diversion of juvenile delinquents: An examination of the processes and relative efficacy of child advocacy and behavioral contracting (Doctoral dissertation, University of Illinois, 1976). *Dissertation Abstracts International, 37*(1-B), 456.

*Davidson, W. S., II, Redner, R., Blakely, C. H., Mitchell, C. M., & Emshoff, J. G. (1987). Diversion of juvenile offenders: An experimental comparison. *Journal of Consulting and Clinical Psychology, 55,* 68–75.

Durlak, J. A., Fuhrman, T., & Lampman, C. (1991). Effectiveness of cognitive–behavior therapy for maladapting children: A meta-analysis. *Psychological Bulletin, 110,* 204–214.

*Eddy, J. M., & Chamberlain, P. (2000). Family management and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. *Journal of Consulting and Clinical Psychology, 68,* 857–863.

*Eddy, J. M., Whaley, R. B., & Chamberlain, P. (2004). The prevention of violent behavior by chronic and serious male juvenile offenders: A 2-year follow-up of a randomized clinical trial. *Journal of Emotional and Behavioral Disorders, 12,* 2–8.

Elliott, D. S. (Series Ed.). (1998). *Blueprints for violence prevention.* Boulder: University of Colorado at Boulder, Institute of Behavioral Science.

*Emshoff, J. G., & Blakely, C. H. (1983). The diversion of delinquent youth: Family-focused intervention. *Children and Youth Services Review, 5,* 343–356.

Farmer, E. M. Z., Compton, S. N., Burns, J. B., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Externalizing disorders. *Journal of Consulting and Clinical Psychology, 70,* 1267–1302.

*Fleischman, M. J. (1982, April). Social learning interventions for aggressive children: From the laboratory to the real world. *The Behavior Therapist, 5,* 55–58.

*Fleming, C. C. (1982). Evaluation of an anger management program with aggressive children in residential treatment. *Dissertation Abstracts International, 43*(12-B), 4143. (UMI No. 8310696)

Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy. *Clinical Psychology: Science and Practice, 3,* 218–229.

Garland, A. F., Hough, R. L., McCabe, K. M., Yeh, M., Wood, P. A., & Aarons, G. A. (2001). Prevalence of psychiatric disorders in youths across five sectors of care. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 409–418.

Garland, A. F., Hurlburt, M. S., & Hawley, K. M. (2006). Examining psychotherapy processes in a services research context. *Clinical Psychology: Science and Practice, 13,* 30–46.

*Grant, J. E. (1988). A problem solving intervention for aggressive adolescent males: A preliminary investigation (Doctoral dissertation, Syracuse University, 1988). *Dissertation Abstracts International, 49,* 912.

Gray-Little, B., & Kaplan, D. (2000). Race and ethnicity in psychotherapy research. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes & practices for the 21st century* (pp. 591–613). New York: Wiley.

*Guarton, D. H. (1992). Anger control training and social problem-solving training: Effects on social competence and problem behaviors (Doctoral dissertation, Hofstra University, 1993). *Dissertation Abstracts International, 53,* 3477.

Hall, G. C. N. (2001). Psychotherapy research with ethnic minorities: Empirical, ethical, and conceptual issues. *Journal of Consulting and Clinical Psychology, 69,* 502–510.

Hammen, C., Rudolph, K., Weisz, J. R., Burge, D., & Rao, U. (1999). The context of depression in clinic-referred youth: Neglected areas in treat-

ment. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 64–71.
Havik, O. E., & VandenBos, G. R. (1996). Limitations of manualized psychotherapy for everyday clinical practice. *Clinical Psychology: Science and Practice, 3,* 264–267.
*Hawkins, J. D., Jenson, J. M., Catalano, R. F., & Wells, E. A. (1991). Effects of a skills training intervention with juvenile delinquents. *Research on Social Work Practice, 1,* 107–121.
Hawley, K. M., & Weisz, J. R. (2005). Youth versus parent working alliance in usual clinical care: Distinctive associations with retention, satisfaction, and treatment outcome. *Journal of Clinical Child and Adolescent Psychology, 34,* 117–128.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.
Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6,* 203–217.
Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9,* 426–445.
Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.
*Henggeler, S. W., Borduin, C. M., Melton, G. B., Mann, B. J., Smith, L. A., Hall, J. A., et al. (1991). Effects of multisystemic therapy on drug use and abuse in serious juvenile offenders: A progress report from two outcome studies. *Family Dynamics of Addiction Quarterly, 1,* 40–51.
*Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65,* 821–833.
*Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60,* 953–961.
*Henggeler, S. W., Melton, G. B., Smith, L. A., Schoenwald, S. K., & Hanley, J. (1993). Family preservation using multisystemic treatment: Long-term follow-up to a clinical trial with serious juvenile offenders. *Journal of Child and Family Studies, 2,* 283–293.
*Henggeler, S. W., Pickrel, S., & Brondino, M. (1999). Multisystemic treatment of substance-abusing and dependent delinquents: Outcomes, treatment, fidelity, and transportability. *Mental Health Services Research, 1,* 171–184.
*Henggeler, S. W., Pickrel, S., Brondino, M., & Crouch, J. (1996). Eliminating (almost) treatment dropout of substance abusing or dependent delinquents through home-based multisystemic therapy. *American Journal of Psychiatry, 153,* 427–428.
*Jarden, H. W. (1995). A comparison of problem-solving interventions on the functioning of youth with disruptive behavior disorders. *Dissertation Abstracts International, 55*(7-A), 1828. (UMI No. 9430959)
Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology, 70,* 158–168.
Kazdin, A. E. (1996). Dropping out of child psychotherapy: Issues for research and implications for practice. *Clinical Child Psychology and Psychiatry, 1,* 133–136.
Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58,* 729–740.
*Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987a). Effects of parent management training and problem-solving skills training combined in the treatment of antisocial child behavior. *Journal of the American Academy of Child & Adolescent Psychiatry, 26,* 416–424.
*Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987b). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. *Journal of Consulting and Clinical Psychology, 55,* 76–85.
Kazdin, A. E., & Weisz, J. R. (Eds.). (2003). *Evidence-based psychotherapies for children and adolescents.* New York: Guilford Press.
*Kerfoot, M. H., Harrington, R., Rogers, J., & Verduyn, C. (2004). A step too far? Randomized trial of cognitive–behaviour therapy delivered by social workers to depressed adolescents. *European Child & Adolescent Psychiatry, 13,* 92–99.
*Klein, N. C., Alexander, J. F., & Parsons, B. V. (1977). Impact of family systems intervention on recidivism and sibling delinquency: A model of primary prevention and program evaluation. *Journal of Consulting and Clinical Psychology, 45,* 469–474.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis: Applied Social Research Methods Series* (Vol. 49). Thousand Oaks, CA: Sage.
Lonigan, C. J., & Elbert, J. C., (Eds.). (1998). Empirically supported psychosocial interventions for children [Special issue]. *Journal of Clinical Child Psychology, 27*(2).
Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice, 9,* 2–12.
*Luk, E. S. L., Staiger, P., Mathai, J., Field, D., & Adler, R. (1998). Comparison of treatments of persistent conduct problems in primary school children: A preliminary evaluation of a modified cognitive–behavioural approach. *Australian and New Zealand Journal of Psychiatry, 32,* 379–386.
*Luk, E. S. L., Staiger, P. K., Mathai, J., Wong, L., Birleson, P., & Adler, R. (2001). Children with persistent conduct problems who drop out of treatment. *European Child & Adolescent Psychiatry, 10,* 28–36.
*Mann, B. J., Borduin, C. M., Henggeler, S. W., & Blaske, D. M. (1990). An investigation of systemic conceptualizations of parent–child coalitions and symptom change. *Journal of Consulting and Clinical Psychology, 58,* 336–344.
McLeod, B. D. (2004). *The Therapy Process Observational Coding System for Child Psychotherapy: Development and application.* Unpublished doctoral dissertation, University of California, Los Angeles.
McLeod, B. D., & Weisz, J. R. (2004). Using dissertations to examine potential bias in child and adolescent clinical trials. *Journal of Consulting and Clinical Psychology, 72,* 235–251.
McLeod, B. D., & Weisz, J. R. (2005). The Therapy Process Observational Coding System—Alliance Scale: Measure characteristics and prediction of outcome in usual clinical practice. *Journal of Consulting and Clinical Psychology, 73,* 323–333.
McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12,* 153–157.
*Morris, J. P. (1981). The effectiveness of anger-control training with institutionalized juvenile offenders: The "Keep Cool Program." *Dissertation Abstracts International, 42*(7-B), 799. (UMI No. 8127991)
*Mufson, L., Dorta, K. P., Wickramaratne, P., Nomura, Y., Olfson, M., & Weissman, M. M. (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry, 61,* 577–584.
Nathan, P. E., & Gorman, J. M. (Eds.). (1998). *Treatments that work.* New York: Oxford University Press.
Nathan, P. E., & Gorman, J. M. (Eds). (2002). *Treatments that work* (2nd ed.). New York: Oxford University Press.
National Advisory Mental Health Council Workgroup on Child and Adolescent Mental Health Intervention and Deployment. (2001). *Blueprint for change: Research on child and adolescent mental health.* Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health.
Office of the Surgeon General. (1999). *Mental health: A report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services.
Office of the Surgeon General. (2004). *Report of the Surgeon General's Conference on Children's Mental Health: A national action agenda.* Rockville, MD: U.S. Department of Health and Human Services.
*Parsons, B. V., & Alexander, J. F. (1973). Short-term family intervention: A therapy outcome study. *Journal of Consulting and Clinical Psychology, 41,* 195–201.
*Patterson, G. R., Chamberlain, P., & Reid, J. B. (1982). A comparative evaluation of a parent-training program. *Behavior Therapy, 13,* 638–650.
President's New Freedom Commission on Mental Health. (2003). *Achieving the promise: Transforming mental health care in America: Final*

report (DHHS Publication No. SMA-03–3832). Rockville, MD: Author.
Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6,* 412–415.
Roth, A., & Fonagy, P. (1996). *What works for whom? A critical review of psychotherapy research.* New York: Guilford Press.
Roth, A., & Fonagy, P. (2004). *What works for whom? A critical review of psychotherapy research* (2nd ed.). New York: Guilford Press.
*Scherer, D. G., Brondino, M. J., Henggeler, S. W., Melton, G. B., & Hanley, J. (1994). Multisystemic family preservation therapy: Preliminary findings from a study of rural and minority serious adolescent offenders. *Journal of Emotional and Behavioral Disorders, 2,* 198–206.
Shadish, W. R., Robinson, L., & Lu, C. (1999). *ES: A computer program for effect size calculation* [Computer manual]. Memphis, TN: University of Memphis.
Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71,* 452–464.
Silverman, W. K., & Hinshaw, S. P. (in press). Evidence-based treatments for child and adolescent disorders [Special issue]. *Journal of Clinical Child and Adolescent Psychology*.
*Snyder, J. J., & White, M. J. (1979). The use of cognitive self-instruction in the treatment of behaviorally disturbed adolescents. *Behavior Therapy, 10,* 227–235.
Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2003). Youth with anxiety disorders in research and service clinics: Examining client differences and similarities. *Journal of Clinical Child and Adolescent Psychology, 32,* 375–385.
*Spence, S. H., & Marzillier, J. S. (1981). Social skills training with adolescent male offenders: II. Short-term, long-term and generalized effects. *Behaviour Research and Therapy, 19,* 349–368.
*Stevens, R., & Pihl, R. O. (1982). The remediation of the student at-risk for failure. *Journal of Clinical Psychology, 38,* 298–301.
Strupp, H. H., & Anderson, T. (1997). On the limitations of therapy manuals. *Clinical Psychology: Science and Practice, 4,* 76–82.
Sue, S. (2003). In defense of cultural competency in psychotherapy and treatment. *American Psychologist, 58,* 964–970.
*Taylor, T. K., Schmidt, F., Pepler, D., & Hodgins, C. (1998). A comparison of eclectic treatment with Webster-Stratton's parents and children series in a children's mental health center: A randomized controlled trial. *Behavior Therapy, 29,* 221–240.
*Van De Wiel, N. M. H., Matthys, W., Cohen-Kettenis, P., & Van Engeland, H. (2003). Application of the Utrecht Coping Power Program and care as usual to children with disruptive behavior disorders in outpatient clinics: A comparative study of cost and course of treatment. *Behavior Therapy, 34,* 421–436.
*Vinick, B. A. (1983). The effects of assertiveness training on aggression and self-concept in conduct disordered adolescents (Doctoral dissertation, Memphis State University, 1983). *Dissertation Abstracts International, 44,* 1982.
Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology, 70,* 299–310.
Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology, 31,* 168–180.
Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples.* Cambridge, UK: Cambridge University Press.
Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology, 56,* 337–363.
Weisz, J. R., Hawley, K. M., & Doss, A. J. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics of North America, 13,* 729–816.
Weisz, J. R., & Jensen, A. L. (2001). Child and adolescent psychotherapy in research and practice contexts: Review of the evidence and suggestions for improving the field. *European Journal of Child and Adolescent Psychiatry, 10,* 12–18.
Weisz, J. R., Southam-Gerow, M., Gordis, E. B., & Connor-Smith, J. (2003). Primary and secondary control enhancement training for youth depression: Applying the deployment-focused model of treatment development and testing. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 165–183). New York: Guilford Press.
Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542–549.
Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47,* 1578–1585.
Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117,* 450–468.
Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004a). The empirical status of empirically supported therapies: Assumptions, findings, and reporting in controlled trials. *Psychological Bulletin, 130,* 631–663.
Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004b). The next generation of psychotherapy research: Reply to Ablon and Marci (2004), Goldfried and Eubanks-Carter (2004), and Haaga (2004). *Psychological Bulletin, 130,* 677–683.
*Whittington, C. K. (1982). Assertion training as a short term treatment method with long term incarcerated juvenile delinquents (Doctoral dissertation, University of Arizona, 1983). *Dissertation Abstracts International, 43,* 3552.
Wilson, D. B. (2003). SPSS macros. Retrieved November 1, 2004, from http://mason.gmu.edu/~dwilsonb/ma.html
Yeh, M., McCabe, K., Hurlburt, M., Hough, R., Hazen, A., Culver, S., et al. (2002). Referral sources, diagnoses, and service types of youth in public outpatient mental health care: A focus on ethnic minorities. *Journal of Behavioral Health Services & Research, 29,* 45–60.
Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent–child (dis)agreement on referral problems at treatment entry. *Journal of Consulting and Clinical Psychology, 69,* 1018–1025.