Original Investigation | META-ANALYSIS

Performance of Evidence-Based Youth Psychotherapies Compared With Usual Clinical Care
A Multilevel Meta-analysis

John R. Weisz, PhD; Sofie Kuppens, PhD; Dikla Eckshtain, PhD; Ana M. Ugueto, PhD; Kristin M. Hawley, PhD; Amanda Jensen-Doss, PhD

IMPORTANCE Research across more than 4 decades has produced numerous empirically tested evidence-based psychotherapies (EBPs) for psychopathology in children and adolescents. The EBPs were developed to improve on usual clinical interventions. Advocates argue that the EBPs should replace usual care, but this assumes that EBPs produce better outcomes than usual care.

OBJECTIVE To determine whether EBPs do in fact produce better outcomes than usual care in youth psychotherapy. We performed a meta-analysis of 52 randomized trials directly comparing EBPs with usual care. Analyses assessed the overall effect of EBPs vs usual care and candidate moderators; we used multilevel analysis to address the dependency among effect sizes (ES) that is common but typically unaddressed in psychotherapy syntheses.

DATA SOURCES We searched the PubMed, PsychINFO, and Dissertation Abstracts International databases for studies from January 1, 1960, through December 31, 2010.

STUDY SELECTION We identified 507 randomized youth psychotherapy trials. Of these, the 52 studies that compared EBPs with usual care were included in the meta-analysis.

DATA EXTRACTION AND SYNTHESIS Sixteen variables (participant, treatment, outcome, and study characteristics) were extracted from studies, and ESs were calculated for all comparisons of EBP vs usual care. We used an extension of the commonly used random-effects meta-analytic model to obtain an overall estimate of the difference between EBP and usual care while accounting for the dependency among ESs. We then fitted a 3-level mixed-effects model to identify moderators that might explain variation in ESs within and between studies by adding study or ES characteristics as fixed predictors.

MAIN OUTCOMES AND MEASURES Primary outcomes of our meta-analysis were mean ES estimates across all studies and for levels of candidate moderators. These ES values were based on measures of symptoms, functioning, and other outcomes assessed within the 52 randomized trials.

RESULTS Evidence-based psychotherapies outperformed usual care. Mean ES was 0.29; the probability was 58% that a randomly selected youth would have a better outcome after EBP than a randomly selected youth after receiving usual care. The following 3 variables moderated treatment benefit: ESs decreased for studies conducted outside North America, for studies in which all participants were impaired enough to qualify for diagnoses, and for outcomes reported by informants other than the youths and parents in therapy. For certain key groups (eg, studies of clinically referred samples and youths with diagnoses), significant EBP effects were not demonstrated.

CONCLUSIONS AND RELEVANCE Evidence-based psychotherapies outperform usual care, but the EBP advantage is modest and moderated by youth, location, and assessment characteristics. The EBPs have room for improvement in the magnitude and range of their benefit relative to usual clinical care.

JAMA Psychiatry. 2013;70(7):750-761. doi:10.1001/jamapsychiatry.2013.1176
Published online May 29, 2013.

Editorial page 666

CME Quiz at jamanetworkcme.com and CME Questions page 768

Author Affiliations: Department of Psychology, Harvard University, Cambridge, Massachusetts (Weisz); Judge Baker Children's Center, Harvard Medical School, Boston, Massachusetts (Weisz, Eckshtain, Ugueto); Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium (Kuppens); EMGO Institute for Health and Care Research, VU University Amsterdam, Amsterdam, the Netherlands (Kuppens); Department of Psychological Sciences, University of Missouri, Columbia (Hawley); Department of Psychology, University of Miami, Miami, Florida (Jensen-Doss).

Corresponding Author: John R. Weisz, PhD, Department of Psychology, Harvard University, 33 Kirkland St, Cambridge, MA 02138 (jrw@wjh.harvard.edu).

A half-century of treatment development research has produced an array of evidence-based psychotherapies (EBPs) for children and adolescents (hereinafter referred to as *youths*). These EBPs—ie, treatments meeting multiple scientific criteria, including replicated support in randomized clinical trials (RCTs)—have been featured in numerous scholarly publications[1-3] and governmental and professional association and academy websites.[4,5] Many researchers argue that EBPs should replace the usual treatments used in everyday clinical care.[6-8] Critics disagree,[9-13] arguing that EBPs (1) have been tested mainly in youths with subclinical problems and may not work well in those with more serious, complex, diagnosed disorders treated in real-world intervention settings; (2) are too rigidly manualized to permit the personalizing of treatment that professionals attempt in usual care; and (3) are mainly products of North American Western culture that may not travel well across ethnic, cultural, or national boundaries. Clearly, whether youth EBPs are superior or inferior to usual clinical care is subject to debate.

This debate highlights a critical empirical question: When youth EBPs and usual care are compared directly, does one form of treatment produce superior outcomes? The question is important scientifically, but also practically and clinically. Given the substantial cost of implementing most EBPs—with proprietary manuals and measures and lengthy training and supervision often required—potential users may reasonably ask whether EBPs reliably outperform usual care, and if so to what extent. Most RCTs cannot answer this question because they have compared EBPs with waiting-list or no-treatment (passage of time) conditions, with attention-only control groups, or with psychological or medication placebo control groups.[2] Those comparison conditions are all designed specifically to be weaker than the active treatment, controlling only for the passage of time, attention paid to the patient, or patient expectancies, and are explicitly not designed to have beneficial therapeutic effects. By contrast, usual care is typically a stronger comparison condition because it entails an array of active interventions designed to produce genuine benefit to the patient.

Thus, comparisons of EBPs with usual care are not only important scientifically and clinically, they also generally represent a stronger standard for testing EBPs than other control groups do. To apply this strong standard, we identified 52 RCTs in which youths were randomly assigned to EBPs or usual clinical care. This study collection is larger and meets more rigorous inclusion standards than any previous work on the topic.[14,15] We conducted a meta-analysis of these 52 studies, assessing the effect of EBPs relative to usual care and testing candidate moderators of treatment benefit. To strengthen the analyses, we used a recently developed multilevel approach to research synthesis that has not previously been applied to psychotherapy research. This approach allowed us to model the dependency among effect sizes (ESs) that is common, but typically unaddressed, in psychotherapy meta-analysis.

Methods

Data Sources, Study Selection, Inclusion criteria

We searched for RCTs of youth psychotherapy that encompassed internalizing (eg, anxiety, depression) and externalizing (eg, misconduct, attention-deficit/hyperactivity disorder) dysfunction.[16,17] Our first search used PsycINFO and PubMed from January 1, 1960, through December 31, 2010. For PsycINFO, we used 21 psychotherapy-related key terms (eg, *psychother-*, *counseling*) from previous youth psychotherapy meta-analyses.[18,19] The PubMed-controlled indexing system of Medical Subject Headings searches publishers who may use different keywords for the same concepts; we used *mental disorders* with the search limits *clinical trial*, *child (3-18 years)*, *published in English*, and *human subjects*. Next, we searched reviews and meta-analyses of youth psychotherapy, followed reference trails, and obtained studies suggested by investigators in the field. Standard guidelines for performing meta-analyses[20-22] recommend addressing publication bias partly by including unpublished studies of acceptable methodological quality. Dissertations are particularly appropriate because they are (1) free of publication bias; (2) reliably identifiable through a systematic search of the Dissertation Abstracts International database; and (3) strong in methodological quality even compared with published studies (perhaps partly because dissertations require faculty committee supervision).[19] Therefore, we searched Dissertation Abstracts International using the same search terms as for the published literature search.

From the studies retrieved, we identified those that compared an EBP with a usual care intervention. Evidence-based psychotherapies were defined as treatments listed in at least 1 of the published reviews systematically identifying EBPs for youths based on the level of empirical support.[1,2,6,23-28] Usual care was defined as psychotherapy, counseling, or other nonmedication interventions provided through outpatient clinics, public programs and agencies (eg, child welfare, probation), or residential facilities (eg, inpatient, group home, detention) for youths. Usual care in which participants sought their own outside services were only included if the authors facilitated service use (eg, arranged intake appointments) or documented that equivalent percentages of participants in usual care and EBP groups (ie, not differing by more than 10%) received services. Other inclusion criteria were (1) participant psychopathology (mental disorder or elevated behavioral/emotional symptoms) documented through pretreatment and posttreatment assessment; (2) random assignment to treatment conditions; and (3) a mean age of 3 to 18 years. We defined psychopathology as meeting criteria for a *DSM* disorder (study years spanned *DSM-II*, *DSM-III*, and *DSM-IV*) or showing elevated behavioral/emotional symptoms because diagnostic and symptom approaches to operationally defining psychopathology are common in the youth treatment outcome literature. Youths who have elevated behavioral/emotional symptoms experience serious impairment[1,2,29,30] and are often referred to and treated in mental health

clinics.[3,31] Including both kinds of studies allowed us to test whether requiring vs not requiring a diagnosis was a moderator of treatment effects.

Data Extraction

Studies were coded for study and sample characteristics, treatment procedures, and multiple candidate moderators of treatment outcome. To assess intercoder agreement, 30 randomly selected studies were independently coded by 4 project coders (D.E., A.M.U., K.M.H., and A.J.D.). Agreement was good for both categorical codes (κ values, 0.71-0.91) and continuous codes (intraclass correlation coefficients, 0.94-0.99).

Data Synthesis: ES Calculation

Effect sizes were represented as Cohen *d* values,[32] reflecting the standardized mean difference between EBP and usual care. Most ES calculations were based on raw data reported in the studies or obtained by contacting study authors; we calculated the difference between the EBP and usual care group means divided by the pooled SD. A positive ES implied superiority of EBP compared with usual care. For studies reporting results using other metrics (eg, frequencies, significance test results), we transformed data to d values using Lipsey-Wilson procedures.[22] Studies reporting only *P* values or significant effects (assumed to reflect $P < .05$ if not otherwise stated) were assigned the minimum d value that would achieve that significance level given the sample size. Studies merely reporting a nonsignificant effect were assigned a d value of 0. Effect size values were adjusted using the Hedges small sample correction.[33]

Data Synthesis: Rationale for and Description of the Multilevel Approach

Because most studies (46 studies [88%]) reported on multiple outcome measures and/or multiple time points, generating multiple ESs per study, the assumption of independence that underlies traditional meta-analytic approaches was violated.[22] Common strategies to deal with dependent ESs have included averaging the ESs within studies, selecting only 1 ES from each study, ignoring the dependency, or applying a "shifting unit of analysis" approach. These approaches ignore or avoid dependency and can distort meta-analytic results.[34] In contrast, multilevel models can more appropriately address multiple ESs within the same study.[35,36] Although multilevel models largely parallel traditional random-effects models,[37] the former do not require independence of ESs; rather, dependence among multiple ESs within studies is modeled by adding an intermediate level. We used a 3-level model including the sampling variation for each ES (level 1), variation across ESs within a study (level 2), and variation across studies (level 3). The basic model consists of the following 3 regression equations referring to each of these levels:

$$d_{jk} = \beta_{0jk} + r_{jk} \text{ with } r_{jk} \sim N(0, \sigma^2_{r_{jk}})$$
$$\beta_{0jk} = \theta_{00k} + u_{0jk} \text{ with } u_{0jk} \sim N(0, \sigma^2_{u})$$
$$\theta_{00k} = \gamma_{000} + v_{00k} \text{ with } v_{00k} \sim N(0, \sigma^2_{v})$$

The first-level equation (equation 1) indicates that the *j*th observed ES from study *k* equals its population value, plus a random deviation, which is assumed to be normally distributed. In a meta-analysis, this residual variance is estimated before performing the meta-analysis. The mean observed sampling variance of standardized mean difference was used in this study; it equaled 0.105. The second-level equation (equation 2) states that the population values comprise a study mean and random deviation from this mean, which is again assumed to be normally distributed. At the third level (equation 3), study mean effects are assumed to vary randomly around an overall mean.

We used this extension of the commonly used random-effects meta-analytic model to obtain an overall estimate of the difference between EBP and usual care. Similarly to traditional mixed-effects models, we subsequently fitted a 3-level mixed-effects model to identify moderators that might explain variation in ESs within and between studies by adding study (level 3) or ES (level 2) characteristics as fixed predictors. Moderator analyses were only conducted if each category contained at least 3 studies. Because including multiple moderators with multiple categories may inflate type II error rates,[38] separate 3-level mixed models were fitted for each moderator variable. Afterward, we fitted a 3-level mixed-effects model that included moderators found to be significant in the separate models, to address possible confounding among moderators.

Parameters estimated in a multilevel meta-analysis are the regression coefficients of the highest-level equations and the variances at the second and third level. Fixed-model parameters are tested using a Wald test, which compares the difference in parameter estimate and the hypothesized population value divided by the standard error with a *t* distribution. For categorical variables with more than 2 categories, the omnibus test of the null hypothesis that the group mean ESs are equal follows an *F* distribution. Likelihood ratio tests comparing the deviance scores of the full model and models excluding variance parameters were used to test variance components. Parameters were estimated using the restricted maximum likelihood procedure implemented in SAS PROC MIXED.[39] Observed ESs were weighted by the inverse of the sampling variance, with a general Satterthwaite approximation used for the denominator degrees of freedom for tests of the regression coefficients.

Publication Bias

We addressed risk of publication bias[22,40,41] in 4 ways. First, we included unpublished dissertations, as discussed above. Second, we compared the mean ES for published studies vs dissertations; the difference was not significant ($t_{53.9} = -0.70$; $P = .49$). Third, we created a funnel plot[42]; standard error was plotted on the vertical axis as a function of ES on the horizontal axis. The plot should resemble an inverted funnel with studies distributed symmetrically around the mean ES if publication bias is absent. With publication bias, the funnel plot should look asymmetrical.[40] Our plot, tested using the weighted regression test of Egger et al,[43] was not asymmetrical ($t_{50} = 0.76$; $P = .45$). Fourth, we computed a classic fail-safe N value,[41] which showed that 565 studies with a mean ES of 0 would need to be added to yield a nonsignificant summary effect. This re-

sult exceeded Rosenthal's[41] benchmark of 80 (5*n* + 10), suggesting that our findings are robust to the threat that excluded studies might have yielded a nonsignificant effect.

Methodological Rigor

Methodological rigor was assessed using the following risk of bias criteria suggested by the Cochrane Collaboration[21]: (1) random sequence generation, (2) blinding of participants, and (3) completeness of outcome data (ie, attrition rate). Because less rigorous studies have been found to yield overestimates of ES,[44] we tested whether ESs differed according to the separate criteria. All studies passed the random sequence generation criterion, and we found no significant differences in mean ES on the blinding criterion ($t_{148} = -1.19$; $P = .24$) or the completeness criterion (ie, attrition rate <40% [$t_{97} = -0.64$; $P = .52$]).

Results

Study Pool

Our search yielded 52 RCTs (45 published trials and 7 dissertations) that met the inclusion criteria (**Figure**). These RCTs included 341 dependent ESs comparing EBPs with usual care.[45-111] The studies, spanning 1973 through 2010, included 5101 participants at the first available measurement point after treatment; mean group size was 46.4 (SD, 67.0); mean age, 12.63 (SD, 2.84) years; and mean sex distribution, 62.67% male (SD, 29.67%). The types of EBP and usual care interventions are described within **Table 1**. Most studies (n = 49) assessed outcomes after therapy; 22 studies included follow-up assessment, ranging from 8 to 76 weeks after the end of treatment (mean [SD], 30.92 [18.74] weeks); 3 studies included only a follow-up assessment. Of those studies reporting race/ethnicity, white youths were the majority in 22 and ethnic minorities in 15. More studies focused on adolescents (n = 37) than children (n = 15). Table 1 provides the other study characteristics.

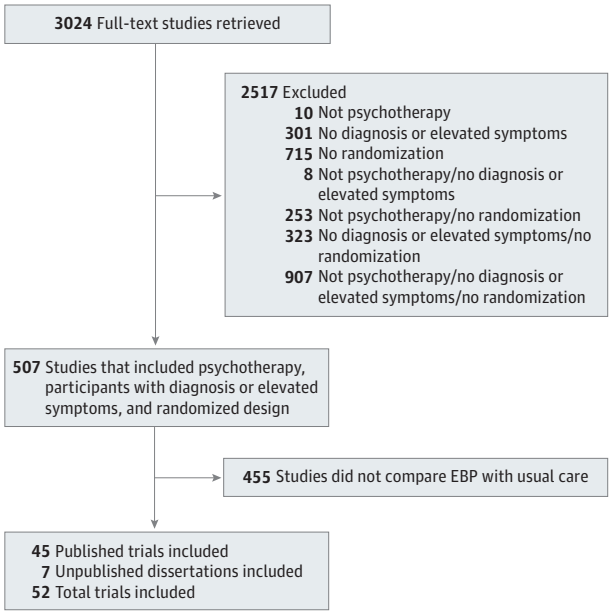
Power

Given the novelty and complexity of the applied 3-level meta-analytic approach, a priori power calculation remains an understudied area. Therefore, we used the procedures of Borenstein et al[20] for standard meta-analysis for an approximate a priori estimate of power. Assuming a high level of between-study variance, a statistical power of 0.80, and an α value of .05, at least 32 studies with a mean sample size of 25 participants would be needed to detect a small overall ES ($d = 0.20$).

Difference Between EBP and Usual Care

Our 3-level model without moderators focused on the overall difference between EBP and usual care across the 341 dependent ESs retrieved from the 52 studies. The mean ES (*d* value) was 0.29 (95% CI, 0.19-0.38; $t_{47.7} = 5.95$; $P < .001$). Effect sizes differed significantly between studies ($\sigma_v^2 = 0.096$; $\chi^2_1 = 112.2$; $P < .001$); differences between dependent ESs within studies were marginally significant ($\sigma_u^2 = 0.011$; $\chi^2_1 = 3.5$; $P = .06$). About 45% of the total ES variance was attributable to differences between studies and about 5% to differences within studies. To assess the effect of larger, more recent trials on the overall mean ES, we calculated the mean of the ES values for the 10 studies in the most recent decade with samples larger than 100; taking into account the multilevel structure of the data, their mean ES was 0.14 (95% CI, 0.02-0.26). This result did not suggest that including more of the larger modern trials would have increased the overall mean ES. Table 1 shows the mean ES for each of the 52 studies.

Figure. Flowchart



Flowchart for the search and identification of randomized clinical trials comparing evidence-based psychotherapy (EBP) with usual clinical care.

Moderator Analyses

Given the heterogeneity of ESs, moderator analyses were first conducted for each moderator separately to identify characteristics that might explain these differences; moderators found to be significant ($P < .05$) were then examined simultaneously to address confounding. Results of the first step, presented in **Table 2**, are summarized herein.

Assessment Timing

Testing whether ES is smaller at follow-up than in the posttreatment period can shed light on the holding power of treatment effects. We found almost identical mean ESs for immediate posttherapy assessments and follow-up assessments a mean of 30.92 (SD, 18.74) weeks later. The number of weeks between the posttherapy assessment and follow-up was also not significantly associated with ES. In the 19 studies that included posttherapy and follow-up assessments, we found no significant effect of assessment time ($t_{51.8} = 0.20$; $P = .84$) or the number of weeks since the end of therapy ($t_{67.4} = -0.19$; $P = .85$). In summary, we found no evidence that effects were significantly weakened over time after treatment.

Table 1. Characteristics of the 52 Randomized Clinical Trials of Evidence-Based Psychotherapies vs Usual Care Included in the Meta-analysis

Source	Target Problem	Sample Size[a]	Mean Age, y	Male Sex, %	Type of EBP	Type of Usual Care[b]	Mean ES[c]
Alexander and Parsons,[45] 1973; Parsons and Alexander,[87] 1973; Klein et al,[77] 1977	Delinquency	29	14.5	44.2	Behavioral Family Systems Therapy (later renamed Functional Family Therapy)	Usual outpatient services (client-centered family groups or psychodynamic family therapy)	0.24
Asarnow et al,[46] 2005	Depression	344	17.2	22	CBT (quality improvement intervention)	Usual outpatient services	0.18
Bank et al,[47] 1991	Delinquency	54	14	100	BPT (Oregon Parent Management Training)	Usual outpatient services	0.07
Barrington et al,[48] 2005	Anxiety	29	9.99	35.19	CBT (for youths, parents, and family)	Usual outpatient services	0.06
Borduin et al,[51] 2009	Delinquency: sexual offenses	46	14	95.8	Multisystemic therapy	Usual outpatient services	0.80
Borduin et al,[49] 1990	Delinquency: sexual offenses	16	14	100	Multisystemic therapy	Usual outpatient services	0.71
Chamberlain and Reid,[54] 1998; Eddy and Chamberlain,[61] 2000; Eddy et al,[62] 2004	Delinquency	79	14.9	100	Multidimensional Treatment Foster Care	Usual residential services	0.46
Davidson,[55] 1976[d]	Delinquency	24	14.5	91.7	Behavioral contracting and usual care	Usual system/agency services	0.40
Deblinger et al,[57] 1996; Deblinger et al,[58] 1999	Anxiety: PTSD	90	9.8	17	CBT for youths; parent training in youth CBT and youth management skills; combination of CBT for youth and parent training	Usual system/agency services	0.53
Diamond et al,[59] 2010	Depression	60	15.1	16.66	Attachment-based family therapy	Usual outpatient services	0.40
Dirks-Linhorst,[60] 2004[d]	Delinquency	141	14.38	63.63	Multisystemic therapy	Usual system/agency services	−0.07
Emshoff and Blakely,[63] 1983; Davidson et al,[56] 1987	Delinquency	136	14.2	83	Behavioral contracting and advocacy	Usual system/agency services	0.14
Fleischman,[64] 1982	Conduct problems	64	7.5	Not provided	BPT (Oregon Parent Management Training)	Usual outpatient services	0.00
Garber et al,[110] 2009	Depression	301	14.8	41.5	CBT (Coping With Depression Course-Adolescents)	Usual outpatient services	0.27
Gillham et al,[65] 2006	Depression	215	11.5	46.86	CBT (Penn Resiliency Program)	Usual outpatient services	0.17
Glisson et al,[66] 2010	Multiple problems	285	14.9	69.1	Multisystemic therapy	Usual outpatient and residential services	0.03
Grant,[67] 1988[d]	Delinquency	26	15.8	100	CBT (problem solving training and usual care)	Usual residential services	−0.25
Hawkins et al,[68] 1991	Delinquency	141	15.5	73	CBT (CBT Skills Training and usual care)	Usual residential services	0.96
Henggeler et al,[69] 1991; Henggeler et al,[70] 1992; Henggeler et al,[71] 1993	Delinquency	56	51.5	77	Multisystemic therapy	Usual system/agency services	0.68
Henggeler et al,[73] 1996; Brown et al,[52] 1999; Henggeler et al,[72] 1999	Delinquency + substance abuse	140	15.7	79	Multisystemic therapy	Usual system/agency services	0.27
Huey et al,[74] 2004	Depression	110	12.9	65	Multisystemic therapy	Usual residential services	0.08
Jarden,[75] 1995[d]	Conduct problems	50	13.5	100	Problem solving skills training and usual care; problem solving skills training, generalization component, and usual care	Usual residential services	0.27
Leve et al,[79] 2005; Chamberlain et al,[53] 2007; Kerr et al,[76] 2009	Delinquency	81	15.3	0	Multidimensional Treatment Foster Care	Usual residential services	0.34
Leve and Chamberlain,[78] 2007; Kerr et al,[76] 2009	Delinquency	83	15.3	0	Multidimensional Treatment Foster Care	Usual residential services	0.43
Luk et al,[80] 1998; Luk et al,[81] 2001	Conduct problems	30	8.6	62.5	CBT (parent-youth modification), Behavioral Family Systems Therapy	Usual outpatient services	−0.39
Mann et al,[82] 1990; Borduin et al,[50] 1995	Delinquency	176	14.8	67.5	Multisystemic therapy	Usual outpatient services	0.48
McCabe and Yeh,[111] 2009	Significant behavioral problems	58	4.4	70.69	BPT (Parent-Child Interaction Therapy-standard and -culturally modified)	Usual outpatient services	0.62
McLaughlin,[83] 2011[d]	Depression	22	11.82	59	CBT (Coping With Depression Course-Adolescents)	Usual outpatient services	0.25

(continued)

Table 1. Characteristics of the 52 Randomized Clinical Trials of Evidence-Based Psychotherapies vs Usual Care Included in the Meta-analysis (continued)

Source	Target Problem	Sample Size[a]	Mean Age, y	Male Sex, %	Type of EBP	Type of Usual Care[b]	Mean ES[c]
Morris,[84] 1981[d]	Delinquency	20	14.75	100	Anger control program and usual care	Usual residential services	0.26
Ogden and Hagen,[85] 2008	Conduct problems	112	8.44	80.4	BPT (Oregon Parent Management Training)	Usual outpatient services	0.15
Ogden and Halliday-Boykins,[86] 2004	Antisocial behaviors	96	14.95	63	Multisystemic therapy	Usual system/agency services and usual residential services	0.23
Patterson et al,[88] 1982	Conduct problems	19	6.80	69	BPT (Oregon Parent Management Training)	Usual outpatient therapy	0.46
Rohde et al,[89] 2004	Conduct problems	64	16.3	100	CBT (Coping With Depression Course-Adolescents)	Usual residential services	0.05
Rowland et al,[90] 2005	Serious emotional disturbance	31	14.5	58	Multisystemic therapy	Usual outpatient services	0.06
Scahill et al,[91] 2006	Disruptive behavior	24	8.9	75	BPT (defiant children)	Usual outpatient services	0.24
Scherer et al,[92] 1994	Delinquency	55	15.1	81.8	Multisystemic therapy (family preservation version)	Usual system/agency services	0.13
Sexton and Turner,[93] 2010	Delinquency	916	15.75	79	Functional family therapy	Usual system/agency services	0.00
Southam-Gerow et al,[94] 2010	Anxiety	36	10.9	43.8	CBT (Coping Cat)	Usual outpatient services	−0.33
Spence and Marzillier,[95] 1981	Delinquency with deficits in interpersonal skills	49	13	100	Social skills training and usual care	Usual residential services	−0.27
Stevens and Pjihl,[96] 1982	Anxiety, low self-esteem, at risk for failure	32	12.5	64.6	CBT	Usual outpatient	0.00
Sukhodolsky et al,[97] 2009	Disruptive/oppositional behavior	26	12.7	92.31	Anger control training	Usual outpatient services	0.80
Sundell et al,[98] 2008	Conduct problems	156	15	61	Multisystemic therapy	Usual outpatient services	−0.10
Szigethy et al,[99] 2007	Depression	38	14.99	49	CBT (PASCET)	Usual outpatient services	0.53
Tang et al,[100] 2009	Depression	73	15.25	34.25	IPT-A-IN	Usual outpatient services	0.71
Taylor et al,[101] 1998	Conduct problems	32	5.6	74.1	BPT	Usual outpatient services	0.50
Timmons-Mitchell et al,[102] 2006	Delinquency: juvenile justice youth	93	15.1	78	Multisystemic therapy	Usual system/agency services	1.30
Van de Weil et al,[103] 2003	Conduct problems	68	10.5	Not reported	Utrecht Coping Power Program	Usual outpatient services	0.00
van den Hoofdakker et al,[105] 2007; van den Hoofdakker et al,[104] 2010	ADHD	94	7.4	80.9	BPT (defiant children, and helping the noncompliant child)	Usual outpatient services	0.17
Weisz et al,[106] 2009	Depression	45	11.77	44	CBT (PASCET)	Usual outpatient services	0.13
Whittington,[107] 1983[d]	Delinquency	44	16	100	Assertiveness training and usual care	Usual residential services	0.27
Young et al,[109] 2010	Depression	52	14.51	40.3	IPT-adolescent skills training	Usual outpatient services	0.30
Young et al,[108] 2006	Depression	40	13.4	14.6	IPT-adolescent skills training	Usual outpatient services	1.23

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; BPT, behavioral parent training; CBT, cognitive behavioral therapy; EBP, evidence-based psychotherapy; ES, effect size; IPT, interpersonal psychotherapy; IPT-A-IN, IPT for depressed adolescents with suicidal risk; PASCET, Primary and Secondary Control Enhancement Training; PTSD, posttraumatic stress disorder.

[a] Sample size reflects the actual number of subjects used to compute ES at the first available measurement point after treatment.

[b] Usual outpatient services included various individual, group, and family-focused interventions in outpatient clinical programs. Usual residential services included various individual and group-focused interventions in youth inpatient, detention, group home, and other residential facilities. Usual system/agency services included various individual, group, and family-focused interventions arranged through probation and child welfare agencies.

[c] Indicates model-based mean ES estimates.

[d] Indicates dissertation.

Study Timing

Effect size was not related to study year ($P = .61$), and we did not find significant interactions of study year with the target problem ($P = .67$), type of EBP ($P = .65$), or developmental period ($P = .51$). The effect of study year was also not significant within any specific category of these moderators (eg, externalizing target problems; $P > .30$ for all).

Study Geographic Location

We tested whether the mean ES differed according to the region in which studies were conducted. Leading EBP researchers[6] have argued that EBPs are evidence based for particular groups and settings, not universally. Because most EBPs were originally developed and tested in North America, they may not fare as well when moved to other locations. Nine

Table 2. Results of Moderator Analyses Based on 3-Level Mixed-Effects Models With 341 Dependent ESs From 52 Studies

Moderator	No. of Studies[a]	No. of ESs	Estimate (95% CI)	Test Statistic	*P* Value
Assessment					
Posttreatment	49	241	0.28 (0.19 to 0.38)	$t_{109} = 0.10$	.92
Follow-up	22	100	0.29 (0.18 to 0.40)		
Posttreatment lag time, wk	39	257	−0.00 (−0.00 to 0.00)	$t_{83.7} = -0.32$	.75
Study year	52	341	0.00 (−0.01 to 0.01)	$t_{51.5} = 0.51$	.61
Location					
North America	42	288	0.33 (0.23 to 0.43)	$t_{44.9} = -2.23$	.03
Outside North America	9	49	0.06 (−0.15 to 0.27)		
Participant recruitment					
Recruited	10	77	0.41 (0.20 to 0.62)	$F_{2,44.9} = 1.85$	.17
Referred	19	140	0.17 (−0.02 to 0.32)		
Nonvoluntary	22	119	0.31 (0.17 to 0.45)		
Same vs different treatment setting					
EBP same as usual care	32	207	0.25 (0.13 to 0.36)	$t_{34.9} = 0.67$	.51
EBP different from usual care	2	14	0.43 (−0.08 to 0.93)		
Sample ethnicity/race majority reported					
White race	22	134	0.42 (0.28 to 0.57)	$t_{31.1} = -1.38$	.18
Ethnic minority	15	116	0.27 (0.10 to 0.43)		
Male sex, %	50	326	−0.00 (−0.01 to 0.00)	$t_{44.8} = -0.46$	.65
Developmental period					
Childhood	15	123	0.16 (−0.01 to 0.33)	$t_{46.6} = 1.73$	.09
Adolescence	37	218	0.34 (0.23 to 0.45)		
Target problem					
Externalizing	34	202	0.31 (0.20 to 0.43)	$F_{2,47} = 1.86$	.17
Internalizing	14	123	0.30 (0.13 to 0.48)		
Mixed	4	16	−0.05 (−0.39 to 0.30)		
Diagnosis given to participants					
All	10	78	0.09 (−0.08 to 0.27)	$t_{14.2} = 2.69$	.02
Some or none	9	82	0.45 (0.26 to 0.65)		
Informant					
Youth	31	117	0.30 (0.19 to 0.40)	$F_{3,228} = 4.18$	.007
Parent	22	79	0.24 (0.12 to 0.36)		
Teacher	9	21	0.10 (−0.10 to 0.29)		
Therapist	3	15	−0.12 (−0.37 to 0.12)		
EBP type					
Youth focused, learning based	21	127	0.31 (0.16 to 0.44)	$F_{3,96.5} = 1.10$	.35
Parent or family focused	13	81	0.16 (−0.01 to 0.33)		
Multisystem approaches	16	99	0.35 (0.19 to 0.52)		
Combinations	4	34	0.29 (0.06 to 0.52)		
Usual care treatment, services					
Outpatient	30	189	0.28 (0.15 to 0.40)	$F_{2,43.2} = 0.31$	.73
Residential	11	68	0.26 (0.04 to 0.48)		
System/agency	9	79	0.37 (0.15 to 0.59)		
Treatment dosage, EBP vs usual care					
More EBP than usual care	11	94	0.45 (0.23 to 0.67)	$F_{2,24.5} = -3.29$	.054
Equal	4	15	0.22 (−0.18 to 0.62)		
Less EBP than usual care	8	51	0.05 (−0.21 to 0.30)		
Investigator allegiance to EBP					
Yes	35	240	0.32 (0.21 to 0.43)	$t_{93.9} = -1.28$	.20
No	19	101	0.21 (0.07 to 0.36)		

Abbreviations: EBP, evidence-based psychotherapy; ES, effect size.

[a] Indicates the number of studies for which information needed for the moderator test was provided.

studies were conducted outside North America (6 in Europe, 2 in Australia, and 1 in Asia). Location showed a significant moderating effect, with lower ES for studies outside North America. Adding this moderator explained 10% of the between-study variance. Two possible explanations for this moderator effect might have been that the efficacy of EBP alone, or usual care alone, differed across countries. However, follow-up logistic regression models based on a logit link function showed no location effect on pretherapy-to-posttherapy gain (0 indicates no gain; 1, gain) for usual care ($t_{145} = -0.10$; $P = .92$) or EBP ($t_{145} = -0.05$; $P = .96$).

Sample Recruitment/Referral

We compared the mean ES for studies involving participants who were recruited (eg, through advertisements), clinically referred, and incarcerated. The groups did not differ significantly in mean ES. However, the mean ES for referred youths was modest ($d = 0.17$) and not statistically significant.

Other Study Variables

Sample Characteristics | We found no significant mean ES difference between studies in which EBP and usual care took place in the same vs different settings. Given that the EBPs were generally not originally designed for minority youths, we investigated whether the difference compared with usual care was smaller in ethnic minority samples than in white samples.[10] The mean ES was somewhat lower for minority than majority samples, but not significantly so. To explore whether sex composition might moderate treatment effects, we tested whether the mean ES was significantly associated with the percentage of boys in the study samples. It was not. We also tested whether EBPs might be more effective with adolescents than children, as suggested by others.[112] The mean ES was more than twice as large for studies with adolescents (mean sample age, ≥12 years; $d = 0.34$) than studies with children (mean sample age, <12 years; $d = 0.16$), but we found no significant moderator effect. Notably, the mean ES for children was not statistically significant. We tested whether ES differed according to the form of youth mental health impairment (ie, internalizing, externalizing, or mixed). Results of the omnibus test were not significant.

Diagnosis | Leaders in the field[113] have suggested that EBP effects may be diminished in samples with more severe psychopathology. Indeed, the mean ES for studies that included only youths with psychopathology severe enough to meet *DSM* criteria was significantly lower than the mean ES for studies not requiring a diagnosis, and the mean ES for diagnosed samples was nonsignificant. Adding this moderator explained 30% of the between-study variance.

Informant | Some researchers have found that youths, parents, and other informants differ in their reports of youth improvement after treatment.[114,115] In our omnibus test, the mean ES differed significantly by informant. Follow-up contrasts revealed a larger mean ES for youth report than teacher report ($t_{228} = 2.00$; $P = .047$) and therapist report ($t_{228} = 3.46$; $P = .001$). The mean ES was also larger for parent report than therapist report ($t_{228} = 2.88$; $P = .004$). Adding the informant moderator explained 27% of the between-study variance and 100% of the within-study variance.

Treatment Variables | The mean ES for parent/family-based treatments was somewhat lower than the mean ES for youth-focused, learning-based, multisystem, or combined treatments, but the difference was not significant. The mean ES was somewhat higher for usual system/agency services than for usual outpatient services and usual residential services; however, the difference among these usual care treatments was not significant. The mean ES was highest ($d = 0.45$) when treatment dose was higher for the EBP than the usual care condition, dropped markedly when dose was the same ($d = 0.22$), and dropped further still when dose was lower for EBP ($d = 0.05$). The mean ES was not significant in the latter 2 conditions. The pattern suggested that EBP superiority might be partially an artifact of larger treatment dose, but the omnibus test was only marginally significant. The dose × type of EBP interaction was also not significant ($P = .27$). The dose was not consistently reported and could be coded in only 23 of the 52 studies.

Investigator Allegiance | Following the example of several researchers,[15] we coded whether study authors had a likely allegiance to the EBP being tested based on whether or not the EBP developer was an author of the article or a committee member for the dissertation. Although the mean ES appeared somewhat larger when investigator allegiance was evident ($d = 0.32$ vs $d = 0.21$; both means were significant), the difference between them was not significant.

Addressing Confounding Among Moderators

Although moderators are the keys to explaining ES differences, moderators may not only be associated with ES but also with each other, complicating the interpretation of single-moderator effects. To address this issue, we simultaneously included all 3 moderators that had shown significant effects within a 3-level mixed-effects model to test the effect of each moderator holding the others constant. We also used a parsimonious modeling approach to test for interactions between moderators, adding possible interactions one at a time. Because results of the moderator analysis for the informant variable revealed similar mean ESs for youth and parent reports and for teacher and therapist reports, these pairs of categories were collapsed into youth or parent reports vs teacher or therapist reports to increase power. Missingness was also coded to reduce loss of information when modeling multiple moderators.

The mean ES for the base category—EBP vs usual care comparisons reported by youths or parents from studies conducted in North America not requiring a diagnosis—was calculated as $d = 0.43$ (95% CI, 0.21-0.66; $t_{43.2} = 3.71$; $P < .001$). The mean ESs decreased significantly when teachers or therapists were the informants ($d = 0.22$; $t_{331} = -2.29$; $P = .02$) and nonsignificantly when studies were conducted outside North America ($d = 0.25$; $t_{44.6} = -1.42$; $P = .16$) and when all participants received a formal diagnosis ($d = 0.17$; $t_{42.7} = -1.60$; $P = .12$). We also found a significant study location × informant inter-

action ($F_{2,232}$ = 5.63; P = .004); in North American studies, EBPs outperformed usual care for youth or parent reports (d = 0.30), but not for teacher or therapist reports (d = −0.11). For studies outside North America the opposite held, with EBPs outperforming usual care on teacher or therapist reports (d = 0.17), but not on youth or parent reports (d = −0.19). The study samples outside North America all met formal diagnostic criteria, which might partially explain their lower mean ESs, but the study location × diagnosis interaction was not significant ($t_{42.3}$ = 0.09; P = .93).

Discussion

Our findings support the perspectives of both EBP proponents and critics. In support of the proponents who argue that EBPs should replace usual care, we found that EBPs produced better outcomes than usual care. The mean standardized difference of 0.29 was not only significant but rather durable as well. Effects at follow-up assessments a mean of 31 weeks after treatment ended were very similar to effects in the immediate posttreatment period, suggesting that the benefit of EBPs relative to usual care may last well beyond the end of treatment.

That said, the mean ES of d = 0.29 was modest, somewhat above the Cohen threshold[32] for a small effect and reflecting a probability of only 58% that a randomly selected youth receiving EBP would be better off after treatment than a randomly selected youth receiving usual care.[116] These findings suggest that (1) the youth EBPs that have been tested to date may be less potent than some have assumed, when pitted against active usual care treatments, and (2) some forms of usual care may be more potent than some have assumed. Indeed, a review of Table 1 reveals several instances in which certain forms of usual care outperformed EBPs. Moreover, the effects of EBPs varied widely, even the effects of the same EBP when tested in relation to different forms of usual care (eg, the variation for multisystemic therapy in Table 1). These variations in ES may also relate to trial design. Studies using tightly controlled efficacy designs might be expected to produce somewhat larger effects than studies using effectiveness designs in which EBPs are evaluated under more usual clinical practice conditions.

Our findings appear to support some of the concerns raised by critics of EBPs[9-13] and noted in the introduction. The concern that EBPs have been tested mostly among youths with subclinical psychopathology and might not fare well among youths with the more serious, complex, diagnosed disorders seen in real-world treatment settings was supported by the low and nonsignificant ES values we found for studies using exclusively diagnosed samples (d = 0.09) and studies focused on clinically referred youths (d = 0.17). In addition, more severe cases may need medication, alone or in combination with psychotherapy. The concern that EBPs may not generalize well beyond their culture of origin was supported by our finding that EBPs, which looked relatively strong within studies in North America, where most EBPs were developed (d = 0.33), showed a much-diminished and nonsignificant effect in studies from other countries (d = 0.06). This finding suggests the potential value of cultural adaptation of treatments.[117] A third concern noted in the introduction—that EBPs are too rigidly manualized to permit the personalization that professionals can attempt in usual care—could not be tested directly in this meta-analysis, but the recent success of modular strategies for personalizing EBPs (eg, trial by Weisz and colleagues[118]) suggests that this possibility bears study in the future. One further concern was raised by our finding that EBP effects that were significant for outcomes reported by the youths (d = 0.30) and parents (d = 0.24) who participated in therapy became nonsignificant for outcomes reported by teachers (d = 0.10), who were more likely to be blinded to treatment condition. These caveats may warrant attention by those considering the costs of implementing EBPs (described in the introduction) relative to the benefits.

Limitations of this meta-analysis suggest future directions. First, usual care interventions were not described in detail in most of the studies, making it difficult to characterize them precisely. The fact that some studies showed usual care matching or outperforming EBPs suggests that those usual care interventions may deserve further study in their own right. Second, additional research in the future will generate more EBP vs usual care comparisons, increasing power to detect additional moderators and interactions among them (eg, a properly powered test of whether the informant effect differs by target problem). Third, an interesting feature in research of this type is that EBP vs usual care studies tend to be carried out in programs, settings, and contexts where research is valued, or at least allowed. This preference might affect the meaning of findings in ways that are understood poorly at present, and findings might be different in clinical settings where research has low priority. Fourth, a growing body of research focuses on pharmacotherapy and its impact in relation to and in combination with youth psychotherapy; that research, not included here, could be a useful topic in its own right for future meta-analyses. Finally, usual care varies across studies and settings and in some instances could include some elements of empirically tested treatments, thus reducing the difference between EBPs and usual care in studies like those reviewed here. This variability further highlights the need for investigators to document thoroughly the contents of the usual care interventions they study.

Our findings show a modest advantage afforded by current EBPs and the limits of that advantage (eg, for youths with diagnosed disorders and those outside North America), which could be seen as a reality check for clinical scientists who develop EBPs for youths. The findings suggest a need in the years ahead to strengthen and broaden the benefit afforded by these treatments for youths and families who seek help. At a more fine-grained level, the accumulation of research in the future should make identification of specific EBPs that do and do not reliably outperform common forms of usual care increasingly possible. Findings at this level of specificity may be valuable to clinicians, clinical directors, and policy makers, helping to inform their decisions as to which EBPs offer sufficiently robust gains over usual care to justify the effort and expense of implementing them in practice.

ARTICLE INFORMATION

Submitted for Publication: February 3, 2012; final revision received November 19, 2012; accepted November 21, 2012.

Published Online: May 29, 2013. doi:10.1001/jamapsychiatry.2013.1176.

Author Contributions: Drs Weisz, Kuppens, and Eckshtain had full access to all the data in the meta-analysis and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by grant 255.313704 from the Norlien Foundation; by grant MH085963 from the National Institute of Mental Health (Dr Weisz); by grant 209 0037 from the Annie E. Casey Foundation (Dr Weisz); and by the Research Foundation, Flanders, Belgium (Dr Kuppens).

Additional Contributions: Jessica Alfano, BA, Natalia Gil, BA, and Louisa Michl, BA, helped with study retrieval and data preparation for this meta-analysis. David Langer, PhD, made skilled contributions to preparation and management of the dataset.

REFERENCES

1. Silverman W, Hinshaw SP, eds. Evidence-based treatments for child and adolescent disorders. *J Clin Child Adolesc*. 2008;37(special issue):1-301.

2. Weisz JR. *Psychotherapy for Children and Adolescents: Evidence-based Treatments and Case Examples.* New York, NY: Cambridge University Press; 2004.

3. Weisz JR, Kazdin AE. *Evidence-based Psychotherapies for Children and Adolescents.* 2nd ed. New York, NY: Guilford Press; 2010.

4. American Academy of Child and Adolescent Psychiatry. Practice parameters. 2011. http://www.aacap.org/cs/root/member_information/practice_information/practice_parameters/practice_parameters. Accessed 2012.

5. Substance Abuse and Mental Health Service Administration. National Register of Health Service Providers. http://nrepp.samhsa.gov/. Accessed 2012.

6. Chambless DL, Hollon SD. Defining empirically supported therapies. *J Consult Clin Psychol*. 1998;66(1):7-18.

7. General S. *Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda.* Rockville, MD: US Department of Health and Human Services; 2004.

8. National Institutes of Health. Blue-print for change: research on child and adolescent mental health. In: *National Advisory Mental Health Council Workgroup on Child and Adolescent Mental Health Intervention and Deployment*. Rockville, MD: US Dept of Health and Human Services, Public Health Service; 2001.

9. Addis ME, Waltz J. Implicit and untested assumptions about the role of psychotherapy treatment manuals in evidence-based mental health practice [commentary]. *Clin Psychol Sci Pract*. 2002;9(4):421-424.

10. Bernal G, Scharrón-del-Río MR. Are empirically supported treatments valid for ethnic minorities? toward an alternative approach for treatment research. *Cultur Divers Ethnic Minor Psychol*. 2001;7(4):328-342.

11. Garfield SL. Some problems associated with "validated" forms of psychotherapy. *Clin Psychol Sci Pract*. 1996;3(3):218-229.

12. Havik OE, Van den Bos GR. Limitations of manualized psychotherapy for everyday practice. *Clin Psychol Sci Pract*. 1996;3(3):264-267.

13. Westen D, Novotny CM, Thompson-Brenner H. The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychol Bull*. 2004;130(4):631-663.

14. Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: a meta-analysis of direct comparisons. *Am Psychol*. 2006;61(7):671-689.

15. Spielmans GI, Gatlin ET, McFall JP. The efficacy of evidence-based psychotherapies versus usual care for youths: controlling confounds in a meta-reanalysis. *Psychother Res*. 2010;20(2):234-246.

16. Mash EJ, Wolfe DA. Disorders of childhood and adolescence. In: Stricker G, Widiger TA, Weiner IB, eds. *Handbook of Psychology: Clinical Psychology.* Hoboken, NJ: John Wiley & Sons Inc; 2003;8:27-63.

17. Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry*. 2010;49(10):980-989.

18. Weisz JR, Weiss B, Alicke MD, Klotz ML. Effectiveness of psychotherapy with children and adolescents: a meta-analysis for clinicians. *J Consult Clin Psychol*. 1987;55(4):542-549.

19. Weisz JR, Weiss B, Han SS, Granger DA, Morton T. Effects of psychotherapy with children and adolescents revisited: a meta-analysis of treatment outcome studies. *Psychol Bull*. 1995;117(3):450-468.

20. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Front Matter: Introduction to Meta-Analysis.* Chichester, England: John Wiley & Sons Ltd; 2009.

21. Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Collaboration; 2011. http://www.cochrane-handbook.org. Accessed 2012.

22. Lipsey M, Wilson D. *Practical Meta-Analysis (Applied Social Research Methods).* Thousand Oaks, CA: Sage Publications, Inc; 2000.

23. Kazdin AE, Weisz JR. Identifying and developing empirically supported child and adolescent treatments. *J Consult Clin Psychol*. 1998;66(1):19-36.

24. Lonigan CJ, Elbert JC, eds. Empirically supported psychosocial interventions for children. *J Clin Child Psychol*. 1998;27(special issue):138-231.

25. Nathan PE, Gorman JM. *A Guide to Treatments That Work.* New York, NY: Oxford University Press; 1998.

26. Nathan PE, Gorman JM. *A Guide to Treatments That Work.* 2nd ed. New York, NY: Oxford University Press; 2002.

27. Roth A, Fonagy P. *What Works For Whom? A Critical Review of Psychotherapy Research.* New York, NY: Guilford Press; 1996.

28. Roth A, Fonagy P. *What Works For Whom? A Critical Review of Psychotherapy Research.* 2nd ed. New York, NY: Guilford Press; 2004.

29. Angold A, Costello EJ, Farmer EMZ, Burns BJ, Erkanli A. Impaired but undiagnosed. *J Am Acad Child Adolesc Psychiatry*. 1999;38(2):129-137.

30. Costello EJ, Angold A, Keeler GP. Adolescent outcomes of childhood disorders: the consequences of severity and impairment. *J Am Acad Child Adolesc Psychiatry*. 1999;38(2):121-128.

31. Jensen AL, Weisz JR. Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *J Consult Clin Psychol*. 2002;70(1):158-168.

32. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum; 1988.

33. Hedges L, Olkin I. *Statistical Methods for Meta-Analysis.* New York, NY: Academic Press; 1985.

34. Cheung SF, Chan DKS. Dependent effect sizes in meta-analysis: incorporating the degree of interdependence. *J Appl Psychol*. 2004;89(5):780-791.

35. Geeraert L, Van den Noortgate W, Grietens H, Onghena P. The effects of early prevention programs for families with young children at risk for physical child abuse and neglect: a meta-analysis. *Child Maltreat*. 2004;9(3):277-291.

36. Marsh HW, Bornmann L, Mutz R, Daniel HD, O'Mara A. Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches. *Rev Educ Res*. 2009;79(3):1290-1326.

37. Van den Noortgate W, Onghena P. Multilevel meta-analysis: a comparison with traditional meta-analytical procedures. *Educ Psychol Meas*. 2003;63(5):765-790.

38. Raudenbush S, Bryk A. *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences).* Thousand Oaks, CA: Sage Publications, Inc; 2002.

39. Littell R, Milliken G, Stroup W, Wolfinger R, Schabenberger O. *SAS for Mixed Models.* Cary, NC: SAS Press; 2006.

40. Begg CB. Publication bias. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis.* New York, NY: Russell Sage Foundation; 1994:399-409.

41. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull*. 1979;86(3):638-641.

42. Torgerson CJ. Publication bias: the Achilles' heel of systematic reviews? *Br J Educ Stud*. 2006;54(1):89-102.

43. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634.

44. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352(9128):609-613.

45. Alexander JF, Parsons BV. Short-term behavioral intervention with delinquent families: impact on family process and recidivism. *J Abnorm Psychol*. 1973;81(3):219-225.

46. Asarnow JR, Jaycox LH, Duan N, et al. Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics: a randomized controlled trial. *JAMA*. 2005;293(3):311-319.

47. Bank L, Marlowe JH, Reid JB, Patterson GR, Weinrott MR. A comparative evaluation of parent-training interventions for families of chronic delinquents. *J Abnorm Child Psychol*. 1991;19(1):15-33.

48. Barrington J, Prior M, Richardson M, Allen K. Effectiveness of CBT vs standard treatment for childhood anxiety disorders in a community clinic setting. *Behav Change*. 2005;22(1):29-43.

49. Borduin CM, Henggeler SW, Blaske DM, Stein RJ. Multisystemic treatment of adolescent sexual offenders. *Int J Offender Ther*. 1990;34(2):105-113.

50. Borduin CM, Mann BJ, Cone LT, et al. Multisystemic treatment of serious juvenile offenders: long-term prevention of criminality and violence. *J Consult Clin Psychol*. 1995;63(4):569-578.

51. Borduin CM, Schaeffer CM, Heiblum N. A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: effects on youth social ecology and criminal activity. *J Consult Clin Psychol*. 2009;77(1):26-37.

52. Brown TL, Henggeler SW, Schoenwald SK, Brondino MJ, Pickrel SG. Multisystemic treatment of substance abusing and dependent juvenile delinquents: effects on school attendance at posttreatment and 6-month follow-up. *Child Serv (Mahwah NJ)*. 1999;2(2):81-93.

53. Chamberlain P, Leve LD, Degarmo DS. Multidimensional treatment foster care for girls in the juvenile justice system: 2-year follow-up of a randomized clinical trial. *J Consult Clin Psychol*. 2007;75(1):187-193.

54. Chamberlain P, Reid JB. Comparison of two community alternatives to incarceration for chronic juvenile offenders. *J Consult Clin Psychol*. 1998;66(4):624-633.

55. Davidson WS. *The Diversion of Juvenile Delinquents: An Examination of the Processes and Relative Efficacy of Child Advocacy and Behavioral Contracting [dissertation]. Urbana: University of Illinois; 1976.*

56. Davidson WS II, Redner R, Blakely CH, Mitchell CM, Emshoff JG. Diversion of juvenile offenders: an experimental comparison. *J Consult Clin Psychol*. 1987;55(1):68-75.

57. Deblinger E, Lippman J, Steer R. Sexually abused children suffering posttraumatic stress symptoms: initial treatment outcome findings. *Child Maltreat*. 1996;1(4):310-321.

58. Deblinger E, Steer RA, Lippmann J. Two-year follow-up study of cognitive behavioral therapy for sexually abused children suffering post-traumatic stress symptoms. *Child Abuse Negl*. 1999;23(12):1371-1378.

59. Diamond GS, Wintersteen MB, Brown GK, et al. Attachment-based family therapy for adolescents with suicidal ideation: a randomized controlled trial. *J Am Acad Child Adolesc Psychiatry*. 2010;49(2):122-131.

60. Dirks-Linhorst PA. *An Evaluation of a Family Court Diversion Program for Delinquent Youth With Chronic Mental Health Needs [dissertation]. St Louis: University of Missouri; 2004.*

61. Eddy JM, Chamberlain P. Family management and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. *J Consult Clin Psychol*. 2000;68(5):857-863.

62. Eddy JM, Whaley RB, Chamberlain P. The prevention of violent behavior by chronic and serious male juvenile offenders: a 2-year follow-up of a randomized clinical trial. *J Emot Behav Disord*. 2004;12(1):2-8.

63. Emshoff JG, Blakely CH. The diversion of delinquent youth: family focused intervention. *Child Youth Serv Rev*. 1983;5(4):343-356.

64. Fleischman MJ. Social learning interventions for aggressive children: from the laboratory to the real world. *Behav Ther*. 1982;5(2):55-58.

65. Gillham JE, Hamilton J, Freres DR, Patton K, Gallop R. Preventing depression among early adolescents in the primary care setting: a randomized controlled study of the Penn Resiliency Program. *J Abnorm Child Psychol*. 2006;34(2):203-219.

66. Glisson C, Schoenwald SK, Hemmelgarn A, et al. Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *J Consult Clin Psychol*. 2010;78(4):537-550.

67. Grant JE. *A Problem Solving Intervention for Aggressive Adolescent Males: A Preliminary Investigation [dissertation]. Syracuse, NY: Syracuse University; 1988.*

68. Hawkins JD, Jenson JM, Catalano RF, Wells EA. Effects of a skills training intervention with juvenile delinquents. *Res Soc Work Pract*. 1991;1(2):107-121.

69. Henggeler SW, Borduin CM, Melton GB, Mann BJ. Effects of multisystemic therapy on drug use and abuse in serious juvenile offenders: a progress report from two outcome studies. *Fam Dynamics Addict Q*. 1991;1(3):40-51.

70. Henggeler SW, Melton GB, Smith LA. Family preservation using multisystemic therapy: an effective alternative to incarcerating serious juvenile offenders. *J Consult Clin Psychol*. 1992;60(6):953-961.

71. Henggeler SW, Melton GB, Smith LA, Schoenwald SK. Family preservation using multisystemic treatment: long-term follow-up to a clinical trial with serious juvenile offenders. *J Child Fam Stud*. 1993;2(4):283-293.

72. Henggeler SW, Pickrel SG, Brondino MJ. Multisystemic treatment of substance-abusing and dependent delinquents: outcomes, treatment fidelity, and transportability. *Ment Health Serv Res*. 1999;1(3):171-184.

73. Henggeler SW, Pickrel SG, Brondino MJ, Crouch JL. Eliminating (almost) treatment dropout of substance abusing or dependent delinquents through home-based multisystemic therapy. *Am J Psychiatry*. 1996;153(3):427-428.

74. Huey SJ Jr, Henggeler SW, Rowland MD, et al. Multisystemic therapy effects on attempted suicide by youths presenting psychiatric emergencies. *J Am Acad Child Adolesc Psychiatry*. 2004;43(2):183-190.

75. Jarden HW. *A Comparison of Problem-Solving Interventions on the Functioning of Youth With Disruptive Behavior Disorders [dissertation]. Bethlehem, PA: Lehigh University; 1995.*

76. Kerr DCR, Leve LD, Chamberlain P. Pregnancy rates among juvenile justice girls in two randomized controlled trials of multidimensional treatment foster care. *J Consult Clin Psychol*. 2009;77(3):588-593.

77. Klein NC, Alexander JF, Parsons BV. Impact of family systems intervention on recidivism and sibling delinquency: a model of primary prevention and program evaluation. *J Consult Clin Psychol*. 1977;45(3):469-474.

78. Leve LD, Chamberlain P. A randomized evaluation of multidimensional treatment foster care: effects on school attendance and homework completion in juvenile justice girls. *Res Soc Work Pract*. 2007;17(6):657-663.

79. Leve LD, Chamberlain P, Reid JB. Intervention outcomes for girls referred from juvenile justice: effects on delinquency. *J Consult Clin Psychol*. 2005;73(6):1181-1185.

80. Luk ESL, Staiger P, Mathai J, Field D, Adler R. Comparison of treatments of persistent conduct problems in primary school children: a preliminary evaluation of a modified cognitive-behavioural approach. *Aust N Z J Psychiatry*. 1998;32(3):379-386.

81. Luk ESL, Staiger PK, Mathai J, Wong L, Birleson P, Adler R. Children with persistent conduct problems who drop out of treatment. *Eur Child Adolesc Psychiatry*. 2001;10(1):28-36.

82. Mann BJ, Borduin CM, Henggeler SW, Blaske DM. An investigation of systemic conceptualizations of parent-child coalitions and symptom change. *J Consult Clin Psychol*. 1990;58(3):336-344.

83. McLaughlin CL. *Evaluating the Effect of an Empirically-supported Group Intervention for Students at-Risk for Depression in a Rural School District [dissertation]. Kent, OH: Kent State University; 2011.*

84. Morris JP. *The Effectiveness of Anger-Control Training With Institutionalized Juvenile Offenders: The "Keep Cool" Program [dissertation]. Richmond: Virginia Commonwealth University; 1981.*

85. Ogden T, Hagen KA. Treatment effectiveness of parent management training in Norway: a randomized controlled trial of children with conduct problems. *J Consult Clin Psychol*. 2008;76(4):607-621.

86. Ogden T, Halliday-Boykins CA. Multisystemic treatment of antisocial adolescents in Norway: replication of clinical outcomes outside of the US. *Child Adolesc Ment Health*. 2004;9(2):77-83.

87. Parsons BV Jr, Alexander JF. Short-term family intervention: a therapy outcome study. *J Consult Clin Psychol*. 1973;41(2):195-201.

88. Patterson GR, Chamberlain P, Reid JB. A comparative evaluation of a parent-training program. *Behav Ther*. 1982;13(5):638-650.

89. Rohde P, Jorgensen JS, Seeley JR, Mace DE. Pilot evaluation of the Coping Course: a

cognitive-behavioral intervention to enhance coping skills in incarcerated youth. *J Am Acad Child Adolesc Psychiatry*. 2004;43(6):669-676.

90. Rowland MD, Halliday-Boykins CA, Henggeler SW, et al. A randomized trial of multisystemic therapy with Hawaii's Felix class youths. *J Emot Behav Disord*. 2005;13(1):13-23.

91. Scahill L, Sukhodolsky DG, Bearss K, et al. Randomized trial of parent management training in children with tic disorders and disruptive behavior. *J Child Neurol*. 2006;21(8):650-656.

92. Scherer DG, Brondino MJ, Henggeler SW, Melton GB, Hanley JH. Multisystemic family preservation therapy: preliminary findings from a study of rural and minority serious adolescent offenders. *J Emot Behav Disord*. 1994;2(4):198-206. Special Series: Center for Mental Health Services Research Projects.

93. Sexton T, Turner CW. The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. *J Fam Psychol*. 2010;24(3):339-348. doi:10.1037/a0019406.

94. Southam-Gerow MA, Weisz JR, Chu BC, McLeod BD, Gordis EB, Connor-Smith JK. Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? a initial effectiveness test. *J Am Acad Child Adolesc Psychiatry*. 2010;49(10):1043-1052.

95. Spence SH, Marzillier JS. Social skills training with adolescent male offenders, II: short-term, long-term and generalized effects. *Behav Res Ther*. 1981;19(4):349-368.

96. Stevens R, Pihl RO. The remediation of the student at-risk for failure. *J Clin Psychol*. 1982;38(2):298-301.

97. Sukhodolsky DG, Vitulano LA, Carroll DH, McGuire J, Leckman JF, Scahill L. Randomized trial of anger control training for adolescents with Tourette's syndrome and disruptive behavior. *J Am Acad Child Adolesc Psychiatry*. 2009;48(4):413-421.

98. Sundell K, Hansson K, Löfholm CA, Olsson T, Gustle L-H, Kadesjö C. The transportability of multisystemic therapy to Sweden: short-term results from a randomized trial of conduct-disordered youths. *J Fam Psychol*. 2008;22(4):550-560.

99. Szigethy E, Kenney E, Carpenter J, et al. Cognitive-behavioral therapy for adolescents with inflammatory bowel disease and subsyndromal depression. *J Am Acad Child Adolesc Psychiatry*. 2007;46(10):1290-1298.

100. Tang TC, Jou SH, Ko CH, Huang SY, Yen CF. Randomized study of school-based intensive interpersonal psychotherapy for depressed adolescents with suicidal risk and parasuicide behaviors. *Psychiatry Clin Neurosci*. 2009;63(4):463-470.

101. Taylor TK, Schmidt F, Pepler D, Hodgins C. A comparison of eclectic treatment with Webster-Stratton's Parents and Children Series in a children's mental health center: a randomized controlled trial. *Behav Ther*. 1998;29(2):221-240.

102. Timmons-Mitchell J, Bender MB, Kishna MA, Mitchell CC. An independent effectiveness trial of multisystemic therapy with juvenile justice youth. *J Clin Child Adolesc Psychol*. 2006;35(2):227-236.

103. Van De Wiel NMH, Matthys W, Cohen-Kettenis P, Van Engeland H. 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. *Behav Ther*. 2003;34(4):421-436.

104. van den Hoofdakker BJ, Nauta MH, van der Veen-Mulders L, et al. Behavioral parent training as an adjunct to routine care in children with attention-deficit/hyperactivity disorder: moderators of treatment response. *J Pediatr Psychol*. 2010;35(3):317-326.

105. van den Hoofdakker BJ, van der Veen-Mulders L, Sytema S, Emmelkamp PMG, Minderaa RB, Nauta MH. Effectiveness of behavioral parent training for children with ADHD in routine clinical practice: a randomized controlled study. *J Am Acad Child Adolesc Psychiatry*. 2007;46(10):1263-1271.

106. Weisz JR, Southam-Gerow MA, Gordis EB, et al. Cognitive-behavioral therapy versus usual clinical care for youth depression: an initial test of transportability to community clinics and clinicians. *J Consult Clin Psychol*. 2009;77(3):383-396.

107. Whittington CK. *Assertion Training as a Short Term Treatment Method With Long Term Incarcerated Juvenile Delinquents [dissertation]. Tucson: University of Arizona; 1983.*

108. Young JF, Mufson L, Davies M. Efficacy of interpersonal psychotherapy-adolescent skills training: an indicated preventive intervention for depression. *J Child Psychol Psychiatry*. 2006;47(12):1254-1262.

109. Young JF, Mufson L, Gallop R. Preventing depression: a randomized trial of interpersonal psychotherapy-adolescent skills training. *Depress Anxiety*. 2010;27(5):426-433.

110. Garber J, Clarke GN, Weersing VR, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA*. 2009;301(21):2215-2224.

111. McCabe K, Yeh M. Parent-child interaction therapy for Mexican Americans: a randomized clinical trial. *J Clin Child Adolesc Psychol*. 2009;38(5):753-759.

112. Dishion TJ, Patterson GR. Age effects in parent training outcome. *Behav Ther*. 1992;23(4):719-729.

113. March J, Silva S, Petrycki S, et al; Treatment for Adolescents With Depression Study (TADS) Team. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) Randomized Controlled Trial. *JAMA*. 2004;292(7):807-820.

114. De Los Reyes A. Introduction to the special section: more than measurement error: discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *J Clin Child Adolesc Psychol*. 2011;40(1):1-9.

115. Weisz JR, Chorpita BF, Frye A, et al; Research Network on Youth Mental Health. Youth top problems: using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *J Consult Clin Psychol*. 2011;79(3):369-380.

116. McGraw KO, Wong SP. A common language effect size statistic. *Psychol Bull*. 1992;111(2):361-365.

117. Benish SG, Quintana S, Wampold BE. Culturally adapted psychotherapy and the legitimacy of myth: a direct-comparison meta-analysis. *J Couns Psychol*. 2011;58(3):279-289.

118. Weisz JR, Chorpita BF, Palinkas LA, et al; Research Network on Youth Mental Health. Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: a randomized effectiveness trial. *Arch Gen Psychiatry*. 2012;69(3):274-282.